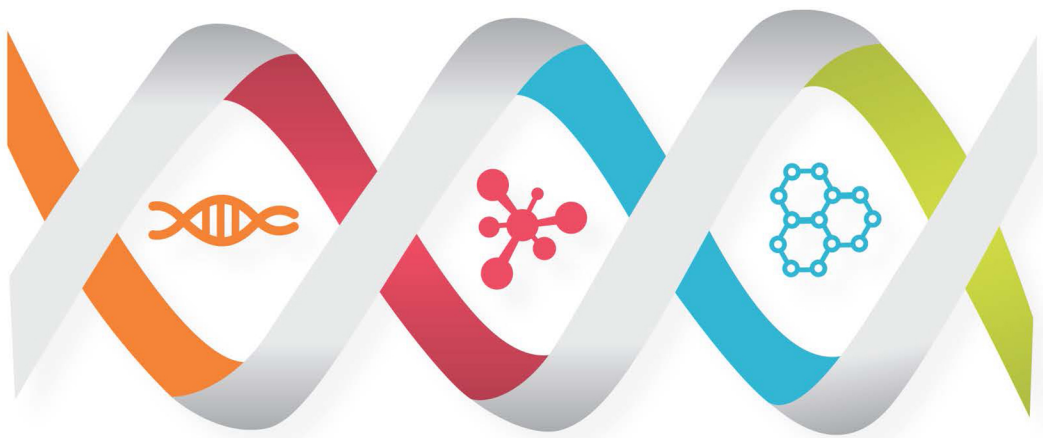


Volume 1

Encyclopedia of Genetics

New Research
(8 Volume Set)



Heidi Carlson
Editor

GENETICS - RESEARCH AND ISSUES

NOVA



Nova
Medicine
&
Health

GENETICS - RESEARCH AND ISSUES

ENCYCLOPEDIA OF GENETICS

NEW RESEARCH

(8 VOLUME SET)

VOLUME

No part of this digital document may be reproduced, stored in a retrieval system or transmitted in any form or by any means. The publisher has taken reasonable care in the preparation of this digital document, but makes no expressed or implied warranty of any kind and assumes no responsibility for any errors or omissions. No liability is assumed for incidental or consequential damages in connection with or arising out of information contained herein. This digital document is sold with the clear understanding that the publisher is not engaged in rendering legal, medical or any other professional services.

GENETICS - RESEARCH AND ISSUES

Additional books and e-books in this series can be found
on Nova's website under the Series tab.

GENETICS - RESEARCH AND ISSUES

ENCYCLOPEDIA OF GENETICS

NEW RESEARCH

(8 VOLUME SET)

VOLUME

HEIDI CARLSON

EDITOR



Copyright © 2019 by Nova Science Publishers, Inc.

All rights reserved. No part of this book may be reproduced, stored in a retrieval system or transmitted in any form or by any means: electronic, electrostatic, magnetic, tape, mechanical photocopying, recording or otherwise without the written permission of the Publisher.

We have partnered with Copyright Clearance Center to make it easy for you to obtain permissions to reuse content from this publication. Simply navigate to this publication's page on Nova's website and locate the "Get Permission" button below the title description. This button is linked directly to the title's permission page on copyright.com. Alternatively, you can visit copyright.com and search by title, ISBN, or ISSN.

For further questions about using the service on copyright.com, please contact:

Copyright Clearance Center

Phone: +1-(978) 750-8400 Fax: +1-(978) 750-4470 E-mail: info@copyright.com.

NOTICE TO THE READER

The Publisher has taken reasonable care in the preparation of this book, but makes no expressed or implied warranty of any kind and assumes no responsibility for any errors or omissions. No liability is assumed for incidental or consequential damages in connection with or arising out of information contained in this book. The Publisher shall not be liable for any special, consequential, or exemplary damages resulting, in whole or in part, from the readers' use of, or reliance upon, this material. Any parts of this book based on government reports are so indicated and copyright is claimed for those parts to the extent applicable to compilations of such works.

Independent verification should be sought for any data, advice or recommendations contained in this book. In addition, no responsibility is assumed by the publisher for any injury and/or damage to persons or property arising from any methods, products, instructions, ideas or otherwise contained in this publication.

This publication is designed to provide accurate and authoritative information with regard to the subject matter covered herein. It is sold with the clear understanding that the Publisher is not engaged in rendering legal or any other professional services. If legal or any other expert assistance is required, the services of a competent person should be sought. FROM A DECLARATION OF PARTICIPANTS JOINTLY ADOPTED BY A COMMITTEE OF THE AMERICAN BAR ASSOCIATION AND A COMMITTEE OF PUBLISHERS.

Additional color graphics may be available in the e-book version of this book.

Library of Congress Cataloging-in-Publication Data

ISBN: 978-1-53614-452-9 (ebook)

Published by Nova Science Publishers, Inc. † New York

CONTENTS

VOLUME 1

Preface		xv
Chapter 1	Speciation in Diatoms: Patterns, Mechanisms, and Environmental Change <i>Joshua T. Cooper and John P. Masly</i>	1
Chapter 2	Sympatric Speciation of Island Plants: The Natural Laboratory of Lord Howe Island <i>Alexander S. T. Papadopoulos, William J. Baker and Vincent Savolainen</i>	37
Chapter 3	Speciation of Arabian Gazelles <i>Hannes Lerp, Torsten Wronski, Thomas M. Butynski and Martin Plath</i>	59
Chapter 4	Chromosome Plasticity, Adaptation and Speciation in Malaria Mosquitoes <i>Igor V. Sharakhov</i>	83
Chapter 5	Transcriptional Differentiation Across the Four Subspecies of <i>Drosophila mojavensis</i> <i>Luciano M. Matzkin and Therese A. Markow</i>	119
Chapter 6	Evolution of Male Courtship Songs in the <i>Drosophila buzzatii</i> Species Cluster <i>Cássia C. Oliveira, Maura H. Manfrin, Fábio de M. Sene and William J. Etges</i>	135
Chapter 7	Prezygotic Isolation in the Parasitoid Wasp Genus <i>Nasonia</i> <i>Maartje C. W. G. Giesbers, Sylvia Gerritsma, Jan Buellesbach, Wenwen Diao, Bart A. Pannebakker, Louis van de Zande, Thomas Schmitt and Leo W. Beukeboom</i>	161
Chapter 8	Where to Look for Speciation Genes When Divergence Is Driven by Postmating, Prezygotic Isolation <i>Jeremy L. Marshall</i>	189

Chapter 9	The Molecular and Evolutionary Basis of Hybrid Sterility: From <i>Odysseus</i> to <i>Overdrive</i> <i>Nitin Phadnis and Harmit S. Malik</i>	203
Chapter 10	The Role of Transposable Elements in Speciation <i>Eric T. Watson and Jeffery P. Demuth</i>	223
Chapter 11	Hybrid Incompatibility: Are Cellular Processes the Battlefields of Genomic Conflict? <i>Norman A. Johnson</i>	247
Chapter 12	Processes in Organic and Cultural Evolution <i>Börje Ekstig</i>	257
Chapter 13	Natural Selection and Diabetes Mellitus <i>Svetlana Shtandel</i>	301
Chapter 14	Genetic Drift among Native People from South American Gran Chaco Region Affects Interleukin 1 Receptor Antagonist Variation <i>Cecilia Inés Catanesi and Laura Angela Glesmann</i>	331

VOLUME 2

Chapter 15	Alternative Splicing and Neurological Disorders: Focus on Parkinson's Disease <i>Velia D'Agata, Valentina La Cognata and Sebastiano Cavallaro</i>	345
Chapter 16	Tau Alternative Splicing in Alzheimer's Disease <i>Gonzalo Emiliano Aranda-Abreu, María Elena Hernández Aguilar, Fausto Rojas Durán, Sonia Lilia Mestizo Gutiérrez and Jorge Manzo Denes</i>	373
Chapter 17	The Role of Splicing Factors in Cancer Prognosis and Treatment <i>Rebeca Martínez-Contreras and Nancy Martínez-Montiel</i>	383
Chapter 18	Nonsense-Mediated Decay and Human Disease <i>Nancy Martínez-Montiel and Rebeca Martinez-Contreras</i>	407
Chapter 19	Alternative Splicing Alterations in Alzheimer's Disease <i>T. A. Ishunina and D. F. Swaab</i>	427
Chapter 20	Plant RNA-Binding Proteins Implicate mRNA Processing in Absciscic Acid (ABA) Responses <i>Raquel F. Carvalho, Vera S. Nunes and Paula Duque</i>	453
Chapter 21	Comprehensive Analyses of Alternative Exons in Neuronally Differentiated P19 cells <i>Hitoshi Suzuki and Toshifumi Tsukahara</i>	475
Chapter 22	Identification of Genuine Alternative Splicing Variants for Rare or Long-Sized Transcripts <i>Seishi Kato</i>	495

Chapter 23	An Epigenetic View on Alternative Splicing <i>Gabriele A. Fontana, Aurora Rigamonti and Silvia M. L. Barabino</i>	515
Chapter 24	Alternative RNA Splicing and Regulation of Nitric Oxide Signaling <i>Iraida G. Sharina</i>	533
Chapter 25	Alternative Splicing by Analyzing a Human mRNA Diversity Using Data of FLJ Human cDNAs <i>Takao Isogai and Ai Wakamatsu</i>	551
Chapter 26	Alternative Splicing in Human Immune System and Autoimmune Diseases <i>Xiang Guo</i>	569
Chapter 27	Poly(ADP-Ribosyl)ation Regulates Alternative Splicing <i>Yingbiao Ji and Alexei V. Tulin</i>	583
Chapter 28	Patent Roadmap for the Biosensor Space <i>Mohamed C. Azeez, Unisha Patel and Dennis Fernandez</i>	595
Chapter 29	Avoidant/Restrictive Food Intake Disorder in a Female Patient Affected by Marfan Syndrome <i>A. P. Verri, A. Serio, M. Grasso, I. Brega and F. Clerici</i>	649
Chapter 30	Optimizing Oil Production in <i>B. napus</i> by Gene Stacking: Transgenic Co-Expression of <i>DGAT1</i> and Partially-Suppressed mtPDCK Cumulatively Enhance Seed Lipid Deposition <i>Elizabeth-France Marillia, Tammy Francis and David C. Taylor</i>	657
Chapter 31	Periodontitis: A Common Oral Disease the Genetic Susceptibility <i>Ying Zheng, You-Qiang Song and W. Keung Leung</i>	689
Chapter 32	Genomic Imprinting and the Brain: Neuron-Specific Switching of Gene Expression at Imprinting Regions <i>Masamitsu Eitoku and Hidenori Kiyosawa</i>	715
Chapter 33	Pharmacogenomics Focusing on Phase Two Metabolizing Enzymes <i>Kung-Hao Liang</i>	739

VOLUME 3

Chapter 34	Sexual Dimorphism in Insect Longevity: Insights from Experimental Evolution <i>Jelica Lazarević, Biljana Stojković and Nikola Tucić</i>	747
Chapter 35	Evidence of Natural and Sexual Selection Shaping the Size of Nuptial Gifts among a Single Bush-cricket Genus (<i>Poecilimon</i> ; Tettigoniidae): An Analysis of Sperm Transfer Patterns <i>J. McCartney, M. A. Potter, A. W. Robertson, K-G. Heller and D. T. Gwynne</i>	777

Chapter 36	Mate Choice Copying in Both Sexes of the Guppy: The Role of Sperm Competition Risk and Sexual Harassment <i>Claudia Zimmer, Alexandros S. Gavalas, Benjamin Kunkel, Janina Hanisch, Sara Martin, Svenja Bischoff, Martin Plath and David Bierbach</i>	795
Chapter 37	Sexual Selection under Parental Choice: Mating Strategies and Reproductive Success <i>Menelaos Apostolou and Marianna Zacharia</i>	813
Chapter 38	Melon Germplasm Characteristics, Diversity, Preservation and Uses <i>C. Mallor and A. Díaz</i>	825
Chapter 39	Preservation and Characterization of Woodland Grape (<i>Vitis vinifera</i> ssp. <i>sylvestris</i> GMEL.) Genotypes of the Szigetköz, Hungary <i>Gizella Jahnke, Zóra Annamária Nagy, Gábor Koltai, Edit Hajdu and János Májer</i>	845
Chapter 40	A Mini Review on Morphological and Genetic Diversity of Sweet (<i>Prunus Avium</i> L.) and Sour Cherry (<i>Prunus Cerasus</i> L.) Cultivars <i>Ioannis Ganopoulos, Panagiotis Madesis, Filippos A. Aravanopoulos, Athanasios Tsiftaris, Thomas Sotiropoulos and Konstantinos Kazantzis</i>	859
Chapter 41	Research on Purple Seed Stain of Soybean: Germplasm Screening and Genetic Resistance <i>Shuxian Li, Pengyin Chen, Bo Zhang and Grover Shannon</i>	871
Chapter 42	Somatic and Gonadal Tissue Cryopreservation: An Alternative Tool for the Germplasm Conservation in Wild Mammals <i>Alexsandra Fernandes Pereira, Alexandre Rodrigues Silva, Gabriela Liberalino Lima and Andréia Maria da Silva</i>	881
Chapter 43	Mitochondrial Gene Diversity of the Mega-Herbivorous Species of the Genus <i>Tapirus</i> (Tapiridae, Perissodactyla) in South America and Some Insights on Their Genetic Conservation, Systematics and the Pleistocene Influence on Their Genetic Characteristics <i>Manuel Ruiz-García, Armando Castellanos, Luz Agueda Bernal, Diego Navas, Myreya Pinedo-Castro and Joseph Mark Shostell</i>	909
Chapter 44	Strategies for Gene Prospecting of Plants in Response to Drought and Salinity <i>Deyvid Novaes Marques, Nicolle Louise Ferreira Barros, Diego Tuloza da Silva, Fabiano Melo de Brito and Claudia Regina Batista de Souza</i>	959

Chapter 45	Clinical Evidence and the Genetic Effects of Traditional Chinese Medicine for the Management of Proteinuria in Patients with Diabetic Nephropathy <i>Xiang Tu, XueFeng Ye, Quan Hu, ChunGuang Xie, ChengShi He, Ming Chen, James B. Jordan and Sen Zhong</i>	993
Chapter 46	Biochemistry and Genetics of Ansamycin Antibiotics <i>Amit Kumar Jha and Jae Kyung Sohng</i>	1017
Chapter 47	BRCA Gene Mutations Mediate Particularly High TNBC Risk by Defective Estrogen Signaling <i>Zsuzsanna Suba</i>	1045
Chapter 48	Genotype-Phenotype Relationships in Language Processes in Rett Syndrome <i>Alessandra Falzone, Antonio Gangemi and Rosa Angela Fabio</i>	1061
Chapter 49	Marfan Syndrome: A Rare and Sometime Very Disabling Genetic Disorder <i>Henry J. Mankin and Keith P. Mankin</i>	1079
Chapter 50	Marfan Syndrome and Periodontitis <i>Jun-ichi Suzuki and Yasunobu Hirata</i>	1093
Chapter 51	Combined Pectus Correction and Aortic Valve Sparing Root Replacement in Marfan Patients <i>Jean-Philippe Verhoye and Jacques Tomasi</i>	1099
Chapter 52	Severe Periodontitis in Marfan Syndrome <i>Naoto Suda</i>	1107
Chapter 53	Preimplantation Genetic Diagnosis for Marfan Syndrome <i>Anver Kuliev and Svetlana Rechitsky</i>	1113

VOLUME 4

Chapter 54	Anuran Cytogenetics: An Overview <i>Cíntia Pelegrineti Targueta, Stenio Eder Vittorazzi, Kaleb Pretto Gatto, Daniel Pacheco Bruschi, Ana Cristina P. Veiga-Menoncello, Shirlei M. Recco-Pimentel and Luciana Bolsoni Lourenço</i>	1117
Chapter 55	Human Cytogenetic Biomonitoring <i>Armanda Teixeira-Gomes, Solange Costa and João Paulo Teixeira</i>	1157
Chapter 56	Complex Karyotypes and Their Detection in Hematologic Malignancies: The Ongoing Role for Classical Cytogenetics <i>Helena Urbankova</i>	1177

Chapter 57	Natural Hybridization between Chromosomal Discrepant Species and the Role of Hybrid Speciation in the Genus <i>Astyanax</i> <i>Francisco de Menezes Cavalcante Sassi, Snaydia Viegas Resende, Rubens Pazza and Karine Frehner Kavalco</i>	1193
Chapter 58	Trichorhinophalangeal Syndrome (TRPS) Revisited <i>N. Selenti, M. Tzetis, E. Tsoutsou and H. Fryssira</i>	1203
Chapter 59	Epigenetic Modifications and Developmental Origin of Health and Diseases (DOHaD) <i>Jia Zheng and Xinhua Xiao</i>	1211
Chapter 60	Effects of Oxidative Stress on Epigenetic Mechanisms <i>Yildiz Dincer and Onur Baykara</i>	1225
Chapter 61	Epigenetic Modifications and Their Potential Role in Tumorigenesis <i>Muthu K. Shanmugam, Frank Arfuso and Gautam Sethi</i>	1241
Chapter 62	Alzheimer's Disease: An Insight from Epigenetic Perspective <i>Yildiz Dincer and Zeynep Sezgin</i>	1291
Chapter 63	Epigenetic Mechanisms in Cardiovascular Diseases <i>Mehmet Güven and Bahadır Batar</i>	1323
Chapter 64	Insights from Epigenomics Analysis of Sperm: Sperm Development and Male Infertility <i>Dheepa Balasubramanian, Eisa Tahmasbpour and Ashok Agarwal</i>	1361
Chapter 65	Epigenetics, Inflammation and Inflammatory Related-Diseases: A General Look <i>Müge Sayitoğlu and Yücel Erbilgin</i>	1383
Chapter 66	A New Molecular Approach to Obesity: Epigenetic Perspective <i>Burcu Bayoglu and Mujgan Cengiz</i>	1419

VOLUME 5

Chapter 67	Genetics and Evolution of Entomopathogenic Fungi <i>A. Garcia, M. Michel, S. Villarreal, F. Castillo, E. Osorio, M. T. García-Ruiz, F. Veana, A. C. Flores and R. Rodríguez</i>	1447
Chapter 68	Mitochondrial Population Genetics Inferences about the Phylogeography and Systematics of the Tayra (<i>Eira barbara</i> , Mustelidae, Carnivora) <i>Manuel Ruiz-García, Nicolás Lichilín-Ortíz, Yolanda Mejia, Juan Manuel Ortega and Joseph Mark Shostell</i>	1491
Chapter 69	The Precision Medicine and Precision Public Health Approaches to Cancer Treatment and Prevention: A Cross-Comparison <i>Stephen M. Modell, Sharon L. R. Kardia and Toby Citrin</i>	1525

Chapter 70	Neuropsychological Profile of People with Williams Syndrome (WS) <i>Anne-Sophie Pezzino, Nathalie Marec-Breton and Agnès Lacroix</i>	1547
Chapter 71	Preimplantation Genetic Diagnosis (PGD) for Chromosome Rearrangements <i>Chun Kyu Lim</i>	1569
Chapter 72	A Family-Based Association Study between the BDNF Gene and Attention Deficit Hyperactivity Disorder in Mexican Children and Adolescents <i>Alan López-López, Miriam Margot Rivera-Angles, Carlos Alfonso Tovilla-Zárate, Roberto Molina-Solís, Araceli Valencia-Hernández, Luis Gómez-Valencia, Thelma Beatriz González-Castro, Isela E Juárez-Rojop, María Lilia López-Narváez, Ana Fresan and Yazmin Hernández-Díaz</i>	1589
Chapter 73	Petroleum Microbiology and Genetics <i>Joana Montezano Marques, Carla Thais Moreira Paixão, Gabriel Villar Monte Palma Pantoja, Artur Luiz da Costa da Silva and Diego Assis das Graça</i>	1601
Chapter 74	Estrogen Activated ERs are the Chief Genetic Safeguards of Somatic and Reproductive Health in Mammals <i>Zsuzsanna Suba</i>	1633
Chapter 75	Motor Abilities are Related to Specific Genotypes in Rett Syndrome <i>R. A. Fabio, T. Caprì, M. Lotan, G. E. Towey and G. Martino</i>	1649
Chapter 76	A Longitudinal Study on the Development of a Boy with Fragile X Syndrome: Developmental Trends and Adaptive Changes in a Single Case Report <i>María Auxiliadora Robles-Bello, Nieves Valencia-Naranjo and David Sánchez-Teruel</i>	1669
Chapter 77	Cornelia De Lange Syndrome: A Review <i>Elizabeth A. M. Frost</i>	1683
Chapter 78	Cognitive-Behavioral Phenotype in Klinefelter Syndrome (KS) and Other Sex Chromosomal Aneuploidies (SCAs): Clinical Variability <i>A. P. Verri, C. D'Angelo, A. Cremante, F. Clerici, A. Mauri and C. Castelletti</i>	1697
Chapter 79	Transcranial Direct Current Stimulation (tDCS) and Cognitive Empowerment for the Functional Recovery of Diseases with Chronic Impairment and Genetic Etiopathogenesis <i>Antonio Gangemi, Tindara Caprì, Rosa Angela Fabio, Paola Puggioni, Alessandra Maria Falzone and Gabriella Martino</i>	1715

VOLUME 6

Chapter 80	Control of Force and Timing during Unimanual and Bimanual Tapping Movements of Adolescents with Down Syndrome <i>Nobuyuki Inui and Junya Masumoto</i>	1727
Chapter 81	Management of Executive Function Following Assisted Cycling Therapy in Adolescents with Down Syndrome <i>Shannon D. R. Ringenbach, Simon D. Holzapfel, Madeline Richter and Jay L. Alberts</i>	1741
Chapter 82	Endocrine Features and Management of Endocrine Problems in Down Syndrome <i>Stefano Stagi, Elena Sandini, Elisabetta Lapi, Perla Scalini and Maurizio de Martino</i>	1757
Chapter 83	Autosomal Dominant Polycystic Kidney Disease: Pathophysiology and Treatment <i>Ashraf M. Mohieldin, Viralkumar S. Upadhyay, Albert C. M. Ong and Surya M. Nauli</i>	1783
Chapter 84	Hereditary Haemorrhagic Telangiectasia or Rendu-Osler-Weber Syndrome <i>Roberto Zarrabeitia, Cristina Amado, Virginia Albiñana and Luisa-María Botella</i>	1807
Chapter 85	Osteogenesis Imperfecta <i>Colin R. Paterson</i>	1831
Chapter 86	Autosomal Dominant Disorders Associated to Breast Cancer <i>Nelly Margarita Macías-Gómez</i>	1849
Chapter 87	Fragile X Syndrome <i>M. Milà</i>	1859
Chapter 88	Fragile X-Associated Primary Ovarian Insufficiency (FXPOI) <i>Maria-Isabel Tejada</i>	1873
Chapter 89	Fragile X-Associated Tremor/Ataxia Syndrome <i>L. Rodriguez-Revenge</i>	1883
Chapter 90	Psychiatric Aspects in Fragile X Syndrome and Related Phenotypes <i>E. Mur and M. Milà</i>	1895
Chapter 91	Clinical Features Associated with <i>FMR1</i> Premutation Carriers <i>M. I. Alvarez-Mora and L. Rodriguez-Revenge</i>	1911
Chapter 92	Genetic Counseling of <i>FMR1</i> <i>I. Madrigal</i>	1927
Chapter 93	Treatment of Fragile X Spectrum: FXS, FXTAS and FXPOI <i>F. J. Ramos and M. P. Ribate</i>	1941

VOLUME 7

Chapter 94	Privacy and Progress in Whole Genome Sequencing <i>Presidential Commission for the Study of Bioethical Issues</i>	1955
Chapter 95	Genetic Testing: Scientific Background for Policymakers <i>Amanda K. Sarata</i>	2045
Chapter 96	Evolution of Human Genome Analysis: Impact on Diseases Diagnosis and Molecular Diagnostic Labs <i>Julie Gauthier, Isabelle Thiffault, Virginie Dormoy-Raclet and Guy A. Rouleau</i>	2057
Chapter 97	Improvements in HSV-1- Derived Amplicon Vectors for Gene Transfer <i>Matias E. Melendez, Alejandra I. Aguirre, Maria V. Baez, Carlos A. Bueno, Anna Salvetti, Diana A. Jerusalinsky and Alberto L. Epstein</i>	2079
Chapter 98	Advances in Viral Genome Research of Papillomaviruses <i>A. C. Freitas, A. L. S. Jesus, A. P. D. Gurgel and M. A. R. Silva</i>	2115
Chapter 99	Synthetic Synthesis of Viral Genomes <i>Monique R. Eller, Roberto S. Dias, Rafael L. Salgado, Cynthia C. da Silva and Sérgio O. De Paula</i>	2131
Chapter 100	Novel Bioinformatics Method to Analyze More Than 10,000 Influenza Virus Strains Easily at Once: Batch-Learning Self Organizing Map (BLSOM) <i>Yuki Iwasaki, Toshimichi Ikemura, Kennosuke Wada, Yoshiko Wada and Takashi Abe</i>	2149
Chapter 101	Oncogenes: Classification, Mechanisms of Activation, and Roles in Cancer Development Role of Oncogenes in Gynecological Pathology <i>Ciro Comparetto and Franco Borruto</i>	2163
Chapter 102	The Classification, Mechanisms of Activation and Roles in Cancer Development of Oncogenes <i>Patricio Barros-Núñez, Mónica Alejandra Rosales-Reynoso and Clara Ibet Juárez-Vázquez</i>	2193

VOLUME 8

Chapter 103	Dicer and MicroRNAs: Oncomirs Are the Next Frontier of Oncogenes Affecting Cancer Etiology and Tumor Progression <i>Ha T. Nguyen and Mandi M. Murph</i>	2227
Chapter 104	The Role of the Epidermal Growth Factor Receptor as a Therapeutic Target in Glioblastoma and Other Malignancies <i>Andrej Pala, Christian Rainer Wirtz and Marc-Eric Halatsch</i>	2255

Chapter 105	Tumor Suppressors Involved in DNA Repair and Carcinoprevention <i>Yasuko Kitagishi and Satoru Matsuda</i>	2267
Chapter 106	Molecular Pathogenesis of Neurofibromatosis Type 1 <i>Ming-Jen Lee, Alton Etheridge, David J. Galas and Kai Wang</i>	2279
Chapter 107	Imaging of Plexiform Neurofibromas in Neurofibromatosis Type 1 <i>Eva Dombi, Nicholas Patronas and Brigitte Widemann</i>	2299
Chapter 108	Surgical Treatment of Giant Neurofibromas <i>Stamatis Sapountzis, Ji Hoon Kim, Abid Rashid and Hung-Chi Chen</i>	2317
Chapter 109	Neurofibromatosis Type 1-Associated Nervous System Tumors and Current Treatment Strategies <i>Nilika Shah Singhal and Sabine Mueller</i>	2329
Chapter 110	Outcome Measures for Optic Pathway Gliomas <i>Peter M. K. de Blank, Grant T. Liu and Michael J. Fisher</i>	2349
Chapter 111	Neurofibromatosis Type 1 Bone Disease: Diagnosis and Management <i>Elizabeth K. Schorry and Alvin H. Crawford</i>	2375
Chapter 112	Cognition and Behaviour in Neurofibromatosis Type 1: Pathogenesis and Emerging Therapies <i>Jonathan M. Payne, Natalie A. Pride and Kathryn N. North</i>	2391
Chapter 113	Communication Disorders Associated with Neurofibromatosis Type 1 <i>Heather L. Thompson, David A. Stevenson, Sean M. Redmond, Bruce L. Smith and David H. Viskochil</i>	2421
Chapter 114	Children with Neurofibromatosis Type 1: Functioning in the Classroom <i>Yafit Gilboa, Sara Rosenblum, Aviva Fattal-Valveski and Naomi Josman</i>	2441
Index		2461
Related Nova Publications		2477

PREFACE

This 8 volume encyclopedia set presents important research on genetics. Some of the topics discussed herein include the speciation of Arabian gazelles, tau alternative splicing in Alzheimer's disease, Cornelia de Lange syndrome and autosomal dominant polycystic kidney disease.

Chapter 1 - Diatoms represent one of the most speciose groups of organisms within the microeukaryota— they occupy almost every aquatic niche worldwide, provide important functions within ecosystems as a major contributor of carbon dioxide fixation from the atmosphere, and also serve as a major contributor to the base of the aquatic food web. Despite this striking example of biodiversity and their ecological importance, the author know relatively little about the speciation processes at work in this group of organisms. In this chapter, the author review the patterns of speciation in marine and freshwater diatoms with respect to their paleontological, environmental, and genetic evolutionary histories. The author summarize the evidence for sex and hybridization among diatom species, and discuss recent work on understanding the mechanisms of reproductive isolation that limit gene flow in the microeukaryota. The author also discuss aspects of diatom genome evolution that provide clues to the genetic basis of speciation in this group. Finally, the author end with a discussion of a natural model system that presents a unique opportunity to address some of the outstanding questions in diatom speciation.

Chapter 2 - The evolution of new species without geographic isolation has received a great deal of attention over the past decade. The mechanisms that can lead to speciation of vascular plants in biogeographic sympatry fall broadly into three categories; speciation with ongoing gene flow (which is often, but not always, synonymous with ecological speciation), hybrid speciation and polyploid speciation. Here, with a particular focus on plant species, the author briefly review the concepts and research on sympatric speciation and its causes. Hybrid speciation and polyploid speciation have occasionally been dismissed as special (“simple”) cases of sympatric speciation and the author consider that, although they are distinctive, there are significant obstacles to overcome before new species can be generated by these mechanisms. As such they are deserving of as much research attention as is currently afforded to speciation with gene flow. The author argue that it remains important to continue identifying cases of speciation in sympatry, in order to gain a better understanding of how all of these mechanisms can drive speciation in restricted areas (rather than simply to quantify sympatric speciation per say) and the author elaborate on the usefulness of Lord Howe Island as a model system for speciation research with this in mind.

Chapter 3 - Gazelles are distributed across Africa and Asia and are adapted to arid and semi-arid environments. In this chapter, the author discusses potential factors promoting the divergence of lineages within this group (*i.e.*, speciation events). The most recent common ancestor of gazelles is thought to have emerged during the Miocene (12-14 Ma) and to have split into the extant genera *Nanger* and *Eudorcas* (both endemic to Africa), *Antilope* (endemic to Asia), and *Gazella* (present in Africa, the Middle East and Asia). Within *Gazella*, two major clades are thought to have evolved allopatrically: (1) a predominantly Asian Clade (*G. bennettii*, *G. subgutturosa*, *G. marica*, *G. leptoceros*, and *G. cuvieri*) and (2) a predominantly African Clade (*G. dorcas*/ *G. saudiya*, *G. spekei*, *G. gazella*, and *G. arabica*). At present, both clades meet in North Africa and, especially, in Arabia. Other splits in this group are better explained by adaptive speciation in response to divergent ecological selection. In both clades, parallel evolution of sister species pairs (a desert-adapted form and a humid mountain-adapted form) can be inferred; desert-dwelling *G. dorcas* in Africa and *G. saudiya* in Arabia have a sister group relationship with mountain-dwelling *G. gazella* in the Levant and *G. arabica* in Arabia. This relationship exists within Africa between the desert-dwelling slender-horned gazelle (*G. leptoceros*) and the mountain-dwelling Cuvier's gazelle (*G. cuvieri*) of the Atlas Mountains. A third species pair occurs in Asia; desert-dwelling goitred gazelle (*G. subgutturosa*) and mountain-dwelling chinkara (*G. bennettii*). These (ecological) speciation events correlate with ecology and behavior: the mountain forms being browsers, sedentary, territorial, and living in small groups, while the desert forms are grazers, migratory/ nomadic, non-territorial, and living in herds. Furthermore, cryptic sister species (*G. gazella*, *G. arabica*), with strikingly similar phenotypes, exist within presumed '*G. gazella*', alluding to a possible allopatric origin of this divergence following an isolation of humid mountain regions during hyper-arid phases. On the other hand, phenotypes within *G. arabica* tend to be variable, but are difficult or impossible to distinguish genetically.

Chapter 4 - Nonrandom distribution of rearrangements is a common feature of eukaryotic chromosomes that is not well understood in terms of genome organization and evolution. In malaria mosquitoes, chromosomal inversions—genome rearrangements that flip chromosomal segments by 180°—are often highly nonuniformly distributed among five chromosomal arms. These rearrangements are associated with epidemiologically important adaptations and, possibly, speciation in mosquitoes. A fundamental question is whether the genomic content of the chromosomal arms is associated with inversion polymorphism and fixation rates. This chapter highlights important differences in evolutionary dynamics of the sex chromosome and autosomes and reviews data about association between characteristics of the genome landscape and rates of chromosomal evolution. Recent studies suggest that a unique combination of various classes of genes and repetitive DNA in each arm, rather than a single type of repetitive element, is likely responsible for arm-specific rates of rearrangements. Additional factors, such as spatial constraints imposed by the nuclear architecture, may be responsible for the nonuniform distribution of rearrangements. Another important question is whether polymorphic inversions on homologous chromosomal arms of distantly related mosquito species nonrandomly share similar sets of genes. The available data indicate that natural selection favors specific gene combinations within polymorphic inversions when distant species are exposed to similar environmental pressures. This knowledge could be useful for the discovery of genes responsible for an association of inversion polymorphisms with phenotypic variations in multiple species. In this chapter, I also review the literature about a possible role of heterochromatin in speciation of malaria mosquitoes. The existing data demonstrate the

elevated evolutionary plasticity of the heterochromatic portion of the mosquito genome. Finally, I discuss the importance of high-quality genome assemblies for reconstructing a gene order-based phylogeny and studying mosquito evolution.

Chapter 5 - Local adaptation can play a fundamental role in the isolation of populations. While less well-studied than differentiation in sequence variation, changes in transcriptional variation during speciation also are fundamental to the evolutionary process. *Drosophila mojavensis* offers an unprecedented opportunity to examine the role of transcriptional differentiation in local adaptation. *Drosophila mojavensis* is a cactophilic fly composed of four ecologically distinct subspecies that inhabit the deserts of western North America. Each of the four subspecies utilizes necrotic tissue of different cactus host species characterized by distinct chemical profiles. The subspecies in Baja California, Mexico uses *Stenocereus gummosus* (Agria), in mainland Sonora it uses *S. thurberi* (Organ Pipe), in the Mojave Desert the host is *Ferocactus cylindraceus* (Red Barrel) and in Santa Catalina Island, USA, *Opuntia littoralis* (Prickly Pear) is the host. In this chapter the author examines how the adaptation to the different environmental conditions across the four subspecies have shaped their transcriptional profiles. Using complete *D. mojavensis* genome microarrays the author examined the transcriptome of third instar larvae from all four subspecies reared in standard laboratory media free of necrotic cactus-derived compounds. This experimental strategy focused on differences between constitutively expressed genes and not genes induced by necrotic cactus-derived compounds. The subspecies exhibited significant differential expression of genes that likely underlie the adaptation to different cactus hosts, such as detoxification genes (Glutathione S-transferases, Cytochrome P450s and UDP-Glycosyltransferases) and chemosensory genes (Odorant Receptors, Gustatory Receptors and Odorant Binding Proteins).

Chapter 6 - Acoustic signals produced to attract mates before, during, and after courtship are frequently involved with sexual selection, sexual isolation, and reproductive isolation in *Drosophila* spp. and other animals, yet few studies have revealed how courtship songs evolve in a larger phylogenetic context. Therefore, the author mapped different acoustic components of courtship songs in the monophyletic *Drosophila buzzatii* species cluster onto an independently derived *period* (*per*) gene + chromosome inversion phylogeny to assess the concordance of courtship song evolution with species divergence. These cactophilic flies are distributed throughout several biomes in southern South America and include the sibling species *D. buzzatii*, *D. koepferae*, *D. serido*, *D. borborema*, *D. seriema*, *D. antonietae*, and *D. gouveai*. All seven species produced two song types; primary and secondary pulse songs, except for *D. borborema* and *D. gouveai* that produced no secondary songs. Courtship songs were characterized by analyzing six commonly studied acoustic components including burst duration (BD), carrier frequency (CF), pulse length (PL), pulse number (PN), inter-burst interval (IBI), and inter-pulse interval (IPI). Significant intra- and inter-specific song variation was observed for BD, PN, and IBI, while CF, PL, and IPI varied in a more species-specific manner, albeit with some overlap. Thus, some song components may be better species recognition signals than others. Multivariate clustering analyses resolved all species into distinct, non-overlapping groups. Mapping individual song traits (BD, IBI, and IPI) as well as composites of these song variables onto our (*per*) gene + chromosome inversion phylogeny revealed no phylogenetic signal when different comparative mapping methods were used. Hence, the evolution of courtship songs in *D. buzzatii* cluster species was uncorrelated with the degree of species divergence. These findings reinforce previous observations that courtship songs evolve rapidly enough to erase any signature of evolutionary affinity between closely related animal species.

Chapter 7 - *Nasonia* (Hymenoptera: Pteromalidae) are small haplodiploid parasitoids of flesh- and blowfly pupae that have become model organisms for speciation research. The genus consists of four closely related species that harbor species-specific *Wolbachia* bacteria that cause postmating reproductive isolation. Antibiotic curing allows for interspecific crosses and genetic exchange between species which, together with haploidy of males, facilitates genetic analysis of fitness traits. In this chapter the author synthesizes the current knowledge on the different prezygotic isolation factors that act in the *Nasonia* genus, and on the genetic basis of these traits. A major prezygotic isolation factor is courtship behaviour. Species differ in male courtship behaviour, and there is large variation in interspecific mate discrimination depending on species pair. The author summarizes data on the strength of prezygotic isolation barriers between all possible species pairs and presents new data on mate discrimination in choice and no-choice experiments. In tests of reinforcement, the author found no stronger female mate discrimination of *N. vitripennis* strains occurring in microsympatry with *N. giraulti* compared to that of allopatric *N. vitripennis* strains. Additionally, the author presents data on the significance of cuticular hydrocarbon profiles for assortative mating in males and discusses other factors that may be involved in prezygotic isolation, including pheromone communication, within-host-mating and sneaking behaviour.

Chapter 8 - Not all genes are equally important during the process of speciation. This premise underlies a basic question in evolutionary biology – “Does divergence at any one gene, or set of genes, play a particularly important role in speciation?” The answer to this question may appear to be no for some forms of reproductive isolation, but there may indeed be ‘kinds of genes’ that routinely play a role in speciation when divergence is driven by postmating, prezygotic traits. Here, I outline the kinds of molecular pathways and interactions that underlie postmating, prezygotic phenotypes which ultimately points to the kinds of genes where the author should look for species-specificity. Interestingly, it is only when the author considers the entire system of interacting sex proteomes, cell structure, membrane dynamics, and physiological pathways that a picture of where species-specific interactions likely occur becomes clear. While this approach points to several kinds of pathways and gene types, a notable finding is that cell membrane receptors (like G-protein coupled receptors and receptor tyrosine kinases) that line the inside of the female reproductive tract and trigger post-copulation cell signaling should be considered the kinds of genes that routinely contribute to reproductive isolation and speciation when divergence is driven by postmating, prezygotic phenotypes.

Chapter 9 - Speciation, the process by which one species splits into two, involves the evolution of reproductive barriers such as the sterility or inviability of hybrids between previously interbreeding populations. One of the earliest intrinsic barriers to gene flow to evolve between geographically isolated populations is the sterility of hybrids of the heterogametic sex. The Dobzhansky-Muller model describes how hybrid incompatibilities that underlie intrinsic postzygotic reproductive barriers such as hybrid sterility may evolve. A major goal in speciation research is to identify the genes that underlie hybrid incompatibilities. The identification of such genes opens the door to understanding the molecular pathways which, when tinkered by evolution within species, lead to sterility in hybrids, and promises to reveal the biological forces that drive the evolution of hybrid sterility genes. While very few hybrid sterility genes have been identified so far, the idea that the evolution of DNA-protein interfaces driven by intragenomic conflict may cause hybrid sterility is gaining wider acceptance. Here the author describes how the molecular and evolutionary insights from two hybrid sterility genes

– *Odysseus* and *Overdrive* – have illuminated the role of heterochromatin in the molecular basis of hybrid sterility and the role of genetic conflict as a driving force in speciation.

Chapter 10 - With innovative DNA sequencing technologies has come a new appreciation for the content of animal and plant genomes. Overwhelmingly, a picture has been painted in which mobile and repetitive elements dominate the genomic landscape. Mobile genetic elements have recently been shown to contribute coding and regulatory sequences during their proliferation, leading to functional and regulatory novelty as well as element-mediated rearrangements coinciding with speciation events. Additionally, dormant elements occasionally erupt in bouts of excision and transposition in interspecific hybrids, resulting in a suite of maladaptive traits. The potential for mobile elements as key players in the evolution and diversification of genomes and species is immense yet in many respects transposable elements still remain the “dark matter” of the genome. This is particularly true of their role in speciation, and in order to fully appreciate their role, much work is still needed. In this chapter, the author investigate the evidence for transposable elements as drivers of diversification and speciation.

Chapter 11 - Recent studies suggest that much hybrid incompatibility results as a consequence of conflict between different components of the genome (genomic conflict). In this chapter, I argue that the battlegrounds for much of the conflict-driven hybrid incompatibility consist of various cellular structures and processes. These include the recombination machinery, centromeres and kinetochores, and heterochromatin and its associated proteins. I conclude with a call to integration between cell biology and evolutionary biology in the study of the evolutionary genetics of hybrid incompatibility.

Chapter 12 - In the present work, I give an all-embracing macroevolutionary perspective on processes of the evolution of life and culture on earth. First, I investigate a complementary form of natural selection that diverges from the traditional form in that it is acting independently of the external environment. This form of natural selection is found as a result of a mathematical analysis of the conditions for population growth. I extend my investigation as well to other evolutionary processes than the organic, such as the evolution of human language and the evolution of science, thereby suggesting other possible forms of underlying explanatory processes. I examine the concept of complexity and show that it implies new insights into the ways natural selection has been acting in forming the evolutionary and developmental processes. Especially, complexity is found to be growing in an accelerating mode, a process that is explained as the combined result of natural selection and a self-reinforcing feedback process. The use of the concept of complexity opens the possibility of construction of a new form of a Tree of Life, which, in contrast to traditional forms, combines complexity and time. Such an illustration of the evolutionary process explains the observation that most species live without great changes over vast periods of time. For species at the highest level of complexity there is no competition from species at still higher levels and these species can therefore, if conditions are beneficent, form new species at still higher levels. The process explains the emergence of new species and the general trend of evolution towards cumulatively higher complexity levels. The cumulative addition of species with successively higher complexity implies that the latest appearing species is the one of the highest level of complexity. At present, this species is the human species.

Chapter 13 - Advances in modern medicine enable a change in the tension of intragroup selection in human populations. Thus, implementation of insulin for type 1 diabetes mellitus (DM) treatment considerably lowered the selection tension for this symptom and converted it from the sub-lethal to the one with a lowered adaptability. Increasing variety of type 1 DM and

type 2 DM is being observed recently in different populations. Moreover, recently the heterogeneity of type 1 DM and type 2 DM has also been observed. The investigation was aimed to study the influence of the selection on the evolution of DM clinical forms. Global implementation of insulin therapy into type 1 DM treatment caused the prevalent increase of this disease. Currently, there is a positive selection trend for type 2 DM, which is the original cause for the prevalence increase within the population, and the negative one for type 1 DM determines its prevalence within the population approximately on the same level. Intra-population change of gene frequencies, susceptible for type 1 and 2 DM, predetermined the development of such DM clinical forms as LADA (latent autoimmune diabetes in adults). It also resulted in the increasing number of patients with an absolute insulin deficiency of type 2 DM, which is a more complicated form of this DM type. Polymorphisms association, changing the immune response and forming the susceptibility to type 1 (C1858T of the gene *PTPN22*, A49G of the gene *CTLA4*) and type 2 (E23K of the gene *KNJ11* participating in insulin insufficiency formation) with different DM forms illustrates the result of decreasing the selection tension against type 1 DM after insulin therapy implementation into the public health services practice.

Chapter 14 - Genetic variation is generally responsible for ethnic differences in certain diseases, including inflammatory processes. The antagonist of cytokine IL-1, IL-1Ra, has been widely studied among Caucasian and African populations for genetic polymorphisms, and interethnic differences have been documented. However, the variation and genotype distribution of polymorphisms from these genes among South American Amerindians are thus far unknown. The author present the results for a VNTR located in the IL-1Ra second intron, in a sample of 169 individuals belonging to 5 Native American populations from Argentina and Paraguay, identified as native according to their self designation, and their geographic location. The author also compare this data with the results obtained from a sample of non-native Argentinian people. Among the five known alleles of the VNTR, the author found only two (alleles 1 and 2) in the native populations from Gran Chaco, and heterozygosity was 19%. The allele 2 which is considered proinflammatory (IL-1Ra * 2) has been found in homozygosity at a considerable frequency among native individuals. However, the association of this allele with inflammatory disease previously demonstrated for other populations of the world, might not be acting in the same way for native people, probably due to local adaptation. This would indicate that the allele 2 will probably not have a negative influence on individuals of native origin who have homozygous genotype 2-2. On the contrary, few records on inflammatory disease are available for the native people. It seems that the increment on allele 2 is not related to any adaptive process but to genetic drift, that changes randomly the allele frequencies of different genetic regions along the genome. The effect of genetic drift has already been demonstrated with genetic markers located in autosomes, X and Y chromosomes. These results indicate that the author must be very cautious when studying populations that passed a process of genetic drift, which can become a confounding factor in epidemiological studies. This information will contribute to a future understanding of the association of this polymorphism with disease, and its incidence in different ethnic groups.

Chapter 15 - Alternative splicing (AS) is a fundamental mechanism of gene expression regulation that extremely expands the coding potential of genomes and the cellular transcriptomic and proteomic diversity. This dynamic and finely-tuned machinery is particularly widespread in the nervous system and is critical for both neuronal development and functions. Alternative splicing defects, therefore, frequently underlie neurological disorders. In

this chapter, the author will focus on Parkinson's disease (PD), the second most common neurodegenerative disorder worldwide. The author will provide a current overview on the impact of alternative splicing in PD by representing the multiple splicing transcripts produced from the major PD-linked genes and their regulation in PD states. Furthermore, the author will review the studies describing global splicing expression changes revealed by whole-genome transcriptomic approaches. The author will also summarize the current knowledge about the alternative splicing modulation in PD through non-coding RNAs (miRNA and lcnRNA) molecules. Assessing the role of alternative splicing on PD pathobiology may represent a central step toward an improved understanding of this complex disease.

Chapter 16 - The microtubule-associated protein (MAP) tau is essential for the development of neuronal cell polarity. Tau protein is preferentially localized in the axon, whereas MAP2, another neuron-specific microtubule-associated protein, is localized in the somatodendritic domain. Previous studies have demonstrated that the localization of these proteins depends, at least in part, on mRNA subcellular localization - of tau mRNA into the axon and MAP2 mRNA into the dendrite. Tau protein plays a pivotal role in the pathophysiology of Alzheimer's disease, in which its hyperphosphorylation promotes aggregation and microtubule destabilization. Tau undergoes alternative splicing, which generates six isoforms in the human brain, due to the inclusion/exclusion of exons 2, 3 and 10. Dysregulation of the splicing process of tau exon 10 is sufficient to cause tauopathy, and has been shown to be influenced by β -amyloid peptides, while there has been less research conducted on the splicing of other exons. This study found that the effects of β -amyloid (42) on the alternative splicing of tau exon 2/3 and 6 caused formed cell processes to retract in differentiated cells and altered the expression of exon 2/3 in cell culture. Expression of exon 6 was repressed under β -amyloid treatment. Although the molecular mechanism for this amyloid-tau interaction remains to be determined, it may have potential implications for the understanding of the underlying neuropathological processes in Alzheimer's disease.

Chapter 17 - Alternative splicing is a co-transcriptional mechanism that regulates eukaryotic gene expression that affects almost 90% of the human genes. In this mechanism, different combinations of exons and introns can be identified and removed from the pre-mRNA, allowing multiple mRNA configurations of joined sequences to arise from a single gene, increasing the coding potential of the genome. Alternative splicing events are catalyzed by a large complex known as the spliceosome, which is conformed by more than 300 proteins and ribonucleoproteins. At the catalytic core of the spliceosome, the small nuclear ribonucleoproteins (snRNPs) U1, U2, U4, U5 and U6 are found. The auxiliary factors responsible for the fine regulation of this mechanism include two major groups: the SR proteins and the hnRNP family. Malfunctions of alternative splicing events can affect the natural expression of a large number of transcripts, including several factors involved in apoptosis or cell survival, molecular processes intimately associated with cancer evolution. In many cases, specific splicing factors or mutated components of the splicing machinery are linked to an anomalous event. Moreover, a switch in specific splicing factors occurs in particular types of cancer where the concomitant outcome is the production of non-functional proteins with added, deleted, or altered domains affecting tumorigenesis. With all this evidence, several strategies have been developed to regulate alternative splicing in which central or auxiliary splicing factors are the target of modulatory molecules. Given the combination of elements needed to regulate alternative splicing, the mechanisms underlying the functional and physiological

implications of these tools are also diverse. Collectively, these strategies are intended to improve cancer prognosis, therapeutic and treatment.

Chapter 18 - The precise control of protein production is essential for the appropriate cell physiology and survival. However, single mutations or accidentally introduced errors can occur during the flow of genetic information. In eukaryotic cells, some messenger RNA (mRNA) molecules leave the nucleus after splicing but further mechanisms are involved in the ultimate outcome of the correspondent protein. One of such systems corresponds to the mRNA surveillance network that includes nonsense-mediated mRNA decay (NMD), an important quality control system that ensures the accuracy of transcripts, to maintain a healthy cellular homeostasis. NMD eliminates anomalous mRNAs harboring premature termination codons (PTCs) to prevent the production of potentially harmful truncated proteins, but it can also regulate the steady state of many physiological mRNAs. Targets for NMD are sometimes linked to mutations or introduced errors but the vast majority can be generated as a result of alternative splicing. The key components required for NMD include the UPF and SMG proteins. These factors interact with a set of proteins (the exon junction complex or EJC) that are deposited just upstream of exon-exon junctions after mRNA splicing and orchestrate a regulated mechanism in order to identify PTCs and either sequester these mRNAs or target them for degradation. Some of these factors are linked to particular disorders and could be modulated in order to correct the defect. The NMD pathway is physiologically and medically important, because an escape from NMD can result in severe clinical phenotypes. Consistent with this, it is estimated that more than 60% of human genes have alternatively spliced products that generate at least one PTC isoform and that approximately 30% of inherited genetic disorders are caused by nonsense mutations or by frameshifts that generate nonsense codons. Initially associated as a genetic cause for beta-thalassemia, the NMD process has expanded from the haematopoietic system and it has reached different kind of human disorders, including cystic fibrosis, Duchenne muscular dystrophy, Hurler syndrome, recessive spinal muscular atrophy and polycystic kidney disease. Finally, it is predicted that some cancers are aided by NMD and the repression of this mechanism may be a potential target for the treatment of certain cancers. In this regard, pharmacological agents have been developed for the treatment of diseases caused by premature stop mutations, including aminoglycosides, but the whole therapeutic potential of NMD targets to correct genetic disorders remains to be exploited.

Chapter 19 - Alzheimer's disease (AD) is the most common type of dementia in the elderly population with a higher prevalence in women. Memory impairment and cognitive decline in AD are primarily linked to its neuropathological hallmarks, i.e., cholinergic neuronal atrophy and death, presence of intraneuronal neurofibrillary tangles and accumulation of amyloid β ($A\beta$) deposits within the extracellular senile plaques. Consequently, genes encoding $A\beta$, tau and proteins involved in $A\beta$ processing received the most careful scientific attention. However, a growing body of evidence suggests that AD pathogenesis is not limited to the changes of AD genes' expression, but also depends on their alternative splicing. Therefore, the present review focuses on data concerning alternative splicing changes in AD. First of all, pivotal AD genes (APP (amyloid precursor protein), tau, presenilin 1 (PS-1) and 2 (PS-2) and apolipoprotein E (APOE)) have a number of splice variants with divergent functions that are differentially expressed in AD and normal brain tissues. Second, alternative splicing of genes involved in $A\beta$ processing and metabolism is also affected in AD. This group includes BACE-1 (β -site APP cleaving enzyme 1), nicastrin and APH-1 which are components of the γ -secretase complex, AIDA-1 (protein that binds to the intracellular domain of $A\beta$ PP following cleavage by γ -

secretase; FE65 (binds to the cytoplasmic tail of β PP). Finally, splice variants of such candidate AD genes as acetylcholinesterase (cholinergic deficit is one of the central mechanisms in AD development), estrogen receptor α (ER α) (estrogen deficiency is one of the predisposing to AD factors), BDNF (brain derived neurotrophic factor that is crucial for neuronal survival and synaptic transmission) and its receptor TrkB, excitatory aminoacid transporter 2 (EAAT2) (colocalized with tau in dystrophic neurons), genes of the ion channels and neurotransmitter receptors, synapsin, RCAN-1 (regulator of calcineurin), neuronal GFAP, ubiquilin-1 (related to the proteasomal degradation of proteins and interaction with PS-1 and PS-2) and genes involved in the regulation of the cell cycle and apoptosis (CIZ1, DENN/MADD) should be considered as important elements of the AD pathogenesis. Together, these data show that a lot of changes in alternative splicing are intimately linked to AD, which should be, thus, considered as a disorder with compromised and disregulated splicing events.

Chapter 20 - Posttranscriptional control of gene expression is crucial for biological processes. In particular, alternative splicing allows the same gene to produce multiple proteins and is thus a key generator of functional complexity. This posttranscriptional regulatory mechanism is a prevalent feature of eukaryotic genomes, being currently estimated to occur in over 50% of plant genes. RNA-binding proteins (RBPs) are known to control many aspects of RNA metabolism, from pre-mRNA splicing to the transport and stability of mRNA transcripts. The Arabidopsis and rice genomes contain about 200-250 genes predicted to encode RBPs, but few of these proteins have been characterized in plants. As sessile organisms, plants are continuously exposed to environmental challenges that affect their growth and development. The phytohormone abscisic acid (ABA) is crucial in the coordination of plant responses to various abiotic stress factors. Interestingly, several RNA-metabolism genes have recently been implicated in the ABA pathway, providing a link between mRNA processing and ABA-mediated plant stress responses. Indeed, the loss- or gain-of-function of genes encoding different classes of RBPs directly involved in constitutive and alternative pre-mRNA splicing, such as snRNP factors or SR and hnRNP proteins, has been shown to result in striking ABA-response plant phenotypes. Functional roles in ABA biosynthesis or signaling have also been reported for other RBPs, including cap-binding proteins, RNA helicases, pentatricopeptide repeat proteins or poly(A) processing enzymes. Taken together, these data support the notion that posttranscriptional networks act as central coordinators of plant stress responses, namely by targeting key components of ABA signal transduction machinery. Future identification of the direct targets of these RBPs should uncover the molecular mechanisms underlying the mode of action of these proteins in the regulation of the ABA pathway.

Chapter 21 - Alternative splicing occurs in most human genes and contributes to protein diversity by producing multiple mRNAs from each gene. Many of these alternative isoforms are expressed in a spatio-temporal manner and play important functional roles in many biological processes including neuronal events. Here, neuronal-specific splicing was comprehensively investigated by using P19 cells. GeneChip Exon Array analyses were performed of total RNA purified from cells during different stages of the differentiation process. Nine filtering conditions were used to efficiently and readily extract alternative exon candidates. A total of 262 candidate exons (236 genes) were obtained. Semi-quantitative RT-PCR results of 30 randomly selected candidates suggested that the expression levels of 87% of the candidates were at least 2-fold different between undifferentiated and differentiated cells. Gene ontology and pathway analyses also showed that many of these 236 candidate genes played important roles in neuronal events. These results suggested that this novel method to

determine alternative exons was successful and efficient. In addition to the known neuronal functions, informatics analyses demonstrated that alternative splicing events of 11 candidate genes played important cell cycle functions. These results suggest that this novel method also provides a way to determine the functional roles of previously unknown alternative splicing events.

Chapter 22 - Only sequence analysis of full-length transcripts can identify genuine alternative splicing variants. However, it was difficult to obtain full-length cDNAs for rare or long-sized transcripts. Recently, the author have developed a powerful method, named a vector-capping method, to construct a size-unbiased full-length cDNA library containing rare or very-long-sized cDNA clones with >10kbp inserts. The characteristic of the full-length cDNA contained in this library is that the intactness of the 5'-end capped site sequence of the cDNA can be assured by the presence of an additional dG at its 5' end. Since this full-length cDNA is derived from a single mRNA, this library enables us to perform in-depth analysis of genuine alternative splicing variants. Using the vector-capping method, the author prepared full-length cDNA libraries from human retina-derived cell lines and analyzed the full sequence of the clones. As a result, the author found many novel alternative-splicing variants for rare or long-sized transcripts. In this chapter, I show the examples of these variants including very-long-sized transcripts with >7kbp that were identified by us for the first time.

Chapter 23 - Genome-wide analysis indicate that alternative splicing of mRNA precursors (pre-mRNAs) affects the vast majority of human genes. Alternative splicing provides a fundamental mechanism to increase transcriptome complexity, allowing the production of two or more mRNA variants that often encode proteins with different, sometimes opposite functions. Its importance is underscored by the observation that misregulated alternative splicing can lead to human diseases. Pre-mRNA splicing has long been known to be regulated by *cis*-acting sequence elements and *trans*-acting protein factors. In higher eukaryotes, it mostly occurs co-transcriptionally so that it is not surprising that a role for chromatin and epigenetic factors in the regulation of exon inclusion is now emerging. In this review, the author will discuss the most recent findings on the roles played by chromatin structure on the modulation of the cotranscriptional splicing reactions. In particular, the author will focus our attention on how the modulation of the transcribing RNA polymerase II, the changes in nucleosome architecture and the presence of different histone modifications contribute to the regulation of the splicing process.

Chapter 24 - Alternative splicing expands transcriptome diversity and allows cells to meet the requirements of an ever-changing extracellular environment. It has been more than 30 years since nitric oxide (NO), a gaseous free radical, was recognized as a critical physiologic signaling molecule. Since then the list of known NO-directed functions has grown substantially to include regulation of smooth muscle function in vascular and gastrointestinal systems, inhibition of platelet aggregation and adhesion, neurotransmission and neuromodulation, regulation of cellular respiration and cytotoxicity, mitochondrial biogenesis and, immune defense. However, the importance of alternative splicing in regulation of enzymatic components of NO signaling pathway started to emerge only recently. Our understanding of the mechanisms governing this process remains very limited and awaits systematic investigation. In this chapter the author will attempt to summarize the available information on alternative splicing of major enzymes mediating canonical NO transduction through the secondary messenger cGMP. The author will highlight evidence accumulated from different laboratories that suggest splicing of enzymes in the NO/cGMP pathway, including nitric oxide

synthases, heterodimeric soluble guanylylcyclase and cGMP-dependent protein kinase, is very complex and strongly affects NO signaling in response to various environmental cues. Future studies will certainly bring new, exciting insights into the role that alternative splicing plays in NO/cGMP biology.

Chapter 25 - Genes could produce multiple protein-coding transcripts by alternative splicing (AS). It was known that AS is related to several diseases in generating biological and functional diversity. So, analysis of the mRNA diversity of genes would be important for understanding gene function. Previously, we obtained 1.46 million human full-length cDNAs (FLJ cDNA) and sequenced their 5'-ends. The author selected approximately 55 thousand cDNAs from FLJ cDNA and sequenced completely. Our FLJ cDNAs were constructed by an optimized oligo-capping method. Thus, by using 5'-EST data, a lot of valuable information was obtained regarding the diversity of the transcription start site (TSS) and amino acid sequences at the N-terminal ends of proteins. The author found that alternative TSSs were utilized for tissue-specific expression. Using this data, the author constructed FLJ Human cDNA Database ver. 3.2, <http://flj.lifesciencedb.jp>. But, a lot of AS-related information still remains to be extracted from our 1.4 million cDNA resources. From a huge number of human sequence information, the author selected only the reliable cDNAs for the analysis of the mRNA diversity. And, the author developed probes which can analyze the mRNA diversity. As a result, by comparing our constructed 5,784 pairs of independent probes, the author are able to detect the expression profiles of splicing patterns in 3,413 genes. Moreover, using these probes, the author analyzed the mRNA diversity of genes after inducing neuronal differentiation in human NT2 teratocarcinoma cells using all-trans retinoic acid (RA). Analyses of NT2 cells identified 452 RA-responsive genes. The mRNA diversity analysis revealed that the rate of genes that showed AS in their N-terminus, internal region, C-terminus, is almost the same, respectively.

Chapter 26 - Alternative splicing of pre-messenger RNA is nearly universal, involving more than 90% of human genes. It's an essential step in gene expression and responsible for much of the proteome diversity in mammalian genomes. The immune system requires a great diversity of functional proteins and immune cells need to respond to various foreign invasions rapidly. Alternative splicing provides one more layer of regulation that is essential for the function of human immune system. Many immune genes have been found to undergo alternative splicing, which plays an important role in the regulation of immune cell activation and function. Dysregulated splicing has been shown to be involved in various immune disorders, such as systemic lupus erythematosus (SLE) and rheumatic arthritis (RA). It may be a direct cause of the disease, or a modifier of disease susceptibility and severity. Further understanding of how alternative splicing may be used as a general mechanism in immune response is essential for our research in the pathophysiology of autoimmune diseases and development of new therapeutics for those diseases. This chapter provides an updated review about alternative splicing in human immune system as well as the relationship between dysregulated splicing and autoimmune diseases, particularly SLE and RA.

Chapter 27 - Poly(ADP-ribosyl)ation is a major post-translational modification performed by Poly(ADP-ribose) Polymerases (PARP1). PARP1 utilizes NAD as the substrate to synthesize a long linear and branching poly(ADP-ribose) (pADPr), ranging in length from 2 to 200 units of ADP-ribose. PARP1 can modify a variety of proteins by attaching pADPr to the target proteins in either a covalent or noncovalent manner. In this way, Poly(ADP-ribosyl)ation is involved in a number of biological processes, including transcription regulation, stress responses, and apoptosis. Recent studies have demonstrated that several splicing factors,

including hnRNPs and S/R proteins, are poly(ADP-ribosyl)ated via noncovalent binding by PARP1. Here, the author describe how poly(ADP-ribosyl)ation regulates alternative splicing by modulating the activities of these splicing factors. In addition, the author discuss a possible role of PARP1 in coupling transcription with splicing.

Chapter 28 - Genetic and behavioral data mining will revolutionize health/lifestyle delivery and outcomes, in much the same way the Internet delivered meaningful social outcomes by providing the infrastructure for information to be collected and connections to be made in ways that had not been possible hitherto. Genetic databases and lifestyle correlations via the cutting edge field of bioinformatics, wearable technology and personalized medicine will be the next echelon of information to be reconnoitered and used through the Internet to enrich our lives. Thanks to smart phones and wearable technology, it will all commence from the palm of our hands. In this review, the author will spotlight the emerging fields of bioinformatics, personalized medicine and biosensor technology: historical insights, current issues, and future trends. In particular, the author will embark on an in-depth look into the sub-fields of personalized medicine and wearable systems for health and lifestyle management. The health-sector, being information intensive, will exploit IT-led platforms that capture personalized health data in real-time and have connectivity with a cloud backbone to intelligently process the information to deliver real-time analytics and actions. Various roadmaps are presented offering aggressive offensive and defensive IP strategies and solutions to companies in the bioinformatics and biosensor space. How does current patent law and policy in the wake of the AIA reforms and Supreme Court decisions in *Alice* and *Mayo* impact on this field and on the strategic guidepost? What's more, the author will reopen the patent troll reform debate: will it create a sweeping sea change, or be just another talking point. How do the sweeping provisions of Obamacare touch upon the field of wearable health systems and bioinformatics? Moreover, how can the author reconcile proprietary value with the growing trends of the open-source movement and the big health data initiatives that lie at the intersection of private and non-profit sectors? Finally, all of this big health data begs an inquiry into issues of privacy and a host of other ethical concerns. With the right strategies in their IP toolkit, bioinformatics and wearable startup companies will not only bolster their current patent position, but also be able to leverage it into a significant competitive advantage. More importantly, the broader public will be able to avail themselves of the utility of these flash-bulb popping innovations that promise to capture personalized health data in real-time to deliver targeted and improved health outcomes.

Chapter 29 - Marfan syndrome is a heritable disorder of connective tissue that is transmitted as an autosomal dominant trait and is characterized by the mutation in the fibrillin 1 gene (FBN1). At present, the diagnosis of Marfan syndrome, based on Ghent criteria, considers both clinical evaluation (in which it is possible to observe alterations in eye, osteoarticular apparatus and cardiovascular system) and genetic evaluation (that reveals a FBN1 gene mutation). Moreover, the author take into account the presence of affected relatives suffering of the same syndrome. The diagnosis of Marfan syndrome is often complex because of the evolution of the phenotype with age and because of the inter-individual variation in the clinical presentation even among the affected family members. According to a recent review, Marfan syndrome is often associated with a range of psychiatric problems like anxiety disorders, depressive disorders, schizophrenia, neurodevelopmental disorders (autism spectrum disorder and attention deficit/hyperactivity disorder) and eating disorders. The author report a 16 years old female patient (MF) (kg 43, H 165 cm, BMI 16.9) who came to our outpatient service for a selective feeding successively diagnosed as Avoidant/Restrictive Food Intake

Disorder (ARFID; DSM 5). “Selective feeding” refers to children who restrict the ingestion of food to a limited number of “favourite” foods, typically five or six different foods. ML., indeed, eats only pasta, sweets and a few other foods. Beyond the altered feeding behaviour, she was diagnosed with a mild intellectual disability (IQ=57; VIQ=61; PIQ=62) and reduced adaptive levels in the socialization and communication domains of the Vineland Adaptive Behavior Scales (VABS). This patient presented ligamentous laxity, long-limbed body habitus suggestive for Marfan syndrome. Therefore, she was addressed to the cardiologist: the echocardiographic evaluation showed mild dilatation of aortic root, with rectilinear sinotubular junction and enlarged ascending aorta (z score=2). Z-scores of the aortic root at the level of the aortic annulus, sinuses of Valsalva, sinotubular junction, and ascending aorta measured from the parasternal long axis in diastole using leading-edge-to-leading-edge technique. The eye examination was normal. Sanger sequencing of the FBN1 gene identified the c.7501G>A (p.Val250Ile) variant/mutation. The missense mutation/variant is not reported in UMD, Ensemble, Exome variant server and dbSNP. The SIFT, PolyPhen2, Mutation Tester (software tool) for predicting damaging of missense mutations and variants gave the following results: PolyPhen2: benign; Mutation Tester: disease causing, SIFT: tolerant. Her mother carries the same mutation. ML. is affected by ARFID, (mild) intellectual disability and adaptive disorders often associated with Marfan syndrome, confirming and extending the results obtained in the review previously described. Moreover, the author underline the importance in multidisciplinary diagnosis and care (also) in these cases.

Chapter 30 - The author have previously shown that introduction of single genes encoding diacylglycerol acyltransferases (*DGAT1s*) or partially-silenced mitochondrial pyruvate dehydrogenase kinase (*mtPDCK*), each had the capacity to enhance seed oil content in *Arabidopsis*. In the current study, the author report the cumulative effects of expressing a two-gene stack: a site-directed mutagenized *DGAT1* from *Tropaeolum majus* (*TmDGAT1* Ser¹⁹⁷-to-Ala¹⁹⁷) and an anti-sense *mtPDCK* from *B. napus* (*A/S mtPDCK*) introduced into *B. napus* cultivar DH12075 to alter seed oil content. Compared to plasmid-only controls, the best lines of the two-gene construct showed, on average, a 23.6% proportional increase (11.6% net increase as % of DW) in oil content which is near-additive to the best results obtained in transgenic experiments with the single genes. These findings demonstrate the utility of stacking two specific transgenes controlling the key steps in two very different metabolic streams-mitochondrial carbon flux (*mtPDCK*; “push”) and triacylglycerol assembly via the Kennedy pathway (*DGAT1*; “pull”), to bring about significant increases in oil content in *B. napus*. This approach holds promise for similar use in other oilseed crops.

Chapter 31 - It is well known that genetic variations can in part affect human oral health. Periodontitis is a common dental noncommunicable disease (NCD). According to the World Health Organization, periodontitis affects 20% of the world population. Periodontal inflammation could eventually induce alveolar bone resorption, causing tooth mobility and tooth loss, which ultimately affects oral function, oral health and the individual’s quality of life. Family study, twins study and linkage analysis in the early years revealed that genetic influence contributes to periodontal disease. Since the first single nucleotide polymorphisms (SNP) study on chronic periodontitis in 1997, many genetic loci were found to be associated with periodontitis, including *IL-1*, *Fc* gamma receptor and complement component 5 genes. Population structure or earlier investigations employing a less desirable sample size may lead to various limitations or bias and therefore diverse results. Meta-analysis reports have proved the association between periodontitis and SNPs in some regions, such as *IL1* and *HLA*. With

development of the technologies, multiple genes can be analyzed in a single assay, and even whole genome SNPs can be screened. Seven genome-wide association studies (GWASs) investigated the potential genetic risk locus for periodontitis and found *GLT6D1* is significantly associated with aggressive periodontitis. In the future, whole genome sequencing, together with replication in multiple populations and functional studies, could potentially disclose the nature of periodontitis as an NCD.

Chapter 32 - Imprinted regions of the mammalian genome are commonly composed of one or more paired paternally and maternally imprinted genes and differentially methylated regions (DMRs). These DMRs are methylated in an allele-specific manner during germ-line or early embryonic development, and they regulate allele-specific gene expression. Although genes in most imprinted regions are expressed ubiquitously among tissues, some imprinted regions manifest tissue-specific gene expression patterns. The brain is a major site of tissue-specific gene expression from imprinted regions in embryonic tissues. The author summarize several imprinted regions that show neuron-specific gene expression patterns. First, the author describe the *SNRPN-UBE3A* domain, which contains three brain-specific imprinted genes, *SNORD115*, *UBE3A-ATS* and *UBE3A*, as well as one imprinted gene with a brain-specific promoter, *SNRPN*. Second, the author discuss the *DLK1-DIO3* domain, which contains the brain-dominant imprinted genes *SNORD112*, *SNORD113* and *SNORD114* as well as numerous brain-specific imprinted miRNAs. Third, the author provide an overview of *Grb10*, which also has a brain-specific promoter and switches the imprinted allele during early neurogenesis. Our chapter reviews these imprinted regions and describes the similarities and differences of the neuron-specific imprinting switch in each region.

Chapter 33 - Treatment of human diseases in general has dual goals: to treat effectively, at the cost of minimal adverse side effects. However, person-to-person differences on drug efficacy and adverse reactions have been frequently observed. Pharmacogenomic studies intend to address such differences based on personal genomic variants such as single nucleotide polymorphisms. In the past, success has been achieved on the identification of major histocompatibility complex genomic variants which are associated to immune-mediated adverse drug effects. Additionally, genomic variants on the direct drug targets have been found to be associated to drug efficacy. Besides immunological and drug-target factors, one other critical factor is on the drug metabolism mediated by various xenobiotic metabolizing and detoxification enzymes. They determine how quickly the drugs can be metabolized and then excreted from the body, thereby affecting drug efficacy and toxicity. Human xenobiotic metabolizing enzymes are categorized by their phases in the metabolizing processes: the modification phase (phase 1), the conjugation phase (phase 2), and further modification and excretion (phase 3). Previously, genes involved in phase 1 have been extensively studied in pharmacogenomics. In comparison, genes in phase 2 received less attention, despite the fact that they are no less important. These genes include Uridine Diphospho-Glucuronosyltransferases and others. This Chapter presents the genomic variants on phase 2 genes which can account for the drug efficacy and adverse reactions.

Chapter 34 - In species with separate sexes, gender differences in longevity are widespread and the extent and direction of these differences varies tremendously among taxa. To understand sexual dimorphism in longevity and explain how different forms of selection shape longevity and other fitness-related traits within and among species, it is important to obtain information on the genetic architecture (the number of genes and degree of inter- and intra-genic interactions) and various mechanistic causes which underlie mortality variation between

the sexes. Here the author review recent empirical studies on gender differences in longevity in insect species, from both mechanistic and evolutionary perspective. Whenever it was possible, the author focus on data obtained from the laboratory evolution experiments because the study of evolution under controlled conditions may provide valuable evidence not only for the effects of natural and sexual selection in shaping sex-specific longevity and mortality rates but also it offers novel insights into the mechanistic basis of these differences.

Chapter 35 - During mating, male bush-crickets transfer a complex spermatophore to the female. The spermatophore is comprised of a large nuptial gift which the female consumes while the sperm from the ejaculate-containing ampulla are transferred into her. Two main functions of the nuptial gift have been proposed: the ejaculate protection hypothesis and the parental investment hypothesis. The former, founded on sexual selection theory, predicts that the time to consume the gift is no longer than necessary to allow for full ejaculate transfer. The latter maintains that gift nutrients increase the fitness or quantity of offspring and hence the gift is likely to be larger than is necessary for complete sperm transfer. With an aim to better understanding the primary function of nuptial gifts, the author examined sperm transfer data from field populations of five *Poecilimon* bush-cricket taxa with varying spermatophore sizes. In the species with the largest spermatophore, the gift was four times larger than necessary to allow for complete sperm transfer and is thus likely to function as paternal investment. Species with medium and small gifts were respectively sufficient and insufficient to allow complete sperm transfer and are likely to represent, to various degrees, ejaculate protection. The author also found that species that produce larger spermatophores transfer greater proportions of available sperm than species producing smaller spermatophores, and thus achieve higher paternal assurance.

Chapter 36 - Mate choice copying was mostly described as a strategy employed by females to assess the quality of potential mates, but also males can copy other males' mate choice. In both cases, focal individuals show an increased propensity to copulate with a potential mating partner they could observe interact sexually with another individual (the 'model'). Sexual interactions, however, convey additional—partly conflicting—information to the choosing individual: females may try to avoid sexually active males due to a rougher courtship or coercive mating attempts, and males may respond to an increased sperm competition risk (SCR). How do females and males respond to different copying situations, in which the model individual either resides in the vicinity of a potential mating partner without physical contact (i.e., no harassment, and low SCR), or interacts sexually with the potential mating partner? Do individuals copy less in the latter situation? The author investigated these questions in the guppy (*Poecilia reticulata*), a livebearing fish with internal fertilization, strong SCR among males, and frequent sexual harassment of females. Focal individuals could choose to associate with a large or a small stimulus fish, and mate choice tests were repeated after the previously non-preferred stimulus fish could be seen associating (low harassment and low SCR) or physically interacting (high harassment and high SCR) with a model individual. In a control treatment, no model was presented. The author found both males and females to copy similarly in both copying treatments, while no response was observed in the control. This contrasts with a study reporting that Atlantic molly (*P. mexicana*) males copy less under elevated SCR. Even though the author lack a compelling explanation as to why both congeners might differ, the author are tempted to argue that strong(er) benefits arising from copying may have selected guppies to copy in a broader range of contexts, including situations where the choosing individual incurs harassment or SCR.

Chapter 37 - Across pre-industrial human societies mating is regulated, with arranged marriage, in which male parents select spouses for their female relatives, being the primary mode of long-term mating. This chapter reviews the model of sexual selection under parental choice that offers a good account for these patterns of human mating. This model postulates that parent-offspring conflict over mating induces parents to control the mate choices of their children, and the spouses they select for them are individuals who conform best to their preferences. By doing so, parents become a significant evolutionary force affecting the course of sexual selection. The model is applied in order to achieve a comprehension of the evolution of specific mating strategies. In particular, it is argued that in a context where mate choice is regulated, at least three mating strategies can thrive: addressing parental choice, addressing female choice and circumventing parental and female choice by force.

Chapter 38 - The need for germplasm banks that safeguard the melon genetic resources is more than justified by the genetic erosion aggravated in the last few decades, not only in the cultivated materials, but also in traditional landraces and wild relatives. The classical and new technologies employed in all stages, from the sample prospecting to resources management, with the conservation and evaluation of the plant resources in between, are described. An added value to these germplasm collections is the use of the genetic resources there preserved in the melon breeding programs. The access to wild genetic resources make possible to exploit them, for instance, for germplasm enhancement incorporating disease resistances in elite varieties. In this sense, conventional breeding methods have rendered unquestionable benefits to agriculture, which can be accelerated by the appearance of new biotechnological tools and genomic resources as a result of the increasing number of vegetable species whose genomes have been or are being sequenced. All germplasm banks, and those preserving vegetable resources like melon in particular, will have to face up to the challenge of characterizing genetically their collections in order to maximize their use in breeding strategies assisted by molecular tools, like molecular markers. At the same time, the use of molecular markers can help to efficiently manage the resources by the creation of core collections. This would alleviate the problems derived from the high number of entries in many of them that is compromising some of the purposes for which they have been created, the conservation and the use of the genetic diversity originally prospected. The advantages in terms of labor and economical investment of preserving, characterizing and using a reduced subset of samples without a considerable sacrifice of the genetic diversity, are undeniable.

Chapter 39 - The evolution of cultivated plants played important role in the ascent of humanity. A large number of theories exist about the evolution of the European grapevine (*Vitis vinifera* ssp. *sativa* L.), it is supposed, that woodland grape itself, or crossing with other species could be the progenitor. The woodland grape (*Vitis vinifera* ssp. *sylvestris* GMEL.) in Hungary is a protected species. The quest and preservation of its populations are significant in terms of nature conservation and reserve of biodiversity as well. In the years of 2010-2015 32 woodland grape genotypes were collected in the Szigetköz, Hungary and ex-situ preserved in the genebank National Agricultural Research and Innovation Centre, Research Institute for Viticulture and Enology, in Badacsonytomaj, Hungary. In 2015-2016 these genotypes were characterised by SSR analysis and were compared with 20 grape rootstocks and 16 *Vitis vinifera* ssp. *sativa* cultivars to ensure the true-to-typeness. Based on the results dendrogram was constructed. In the dendrogram the *Vitis vinifera* ssp. *sylvestris* accessions form an distinct group, but are closer to the *Vitis vinifera* ssp. *sativa* cultivars, than to the rootstocks. This raises the probability, that these accessions are true-to-type woodland grapes.

Chapter 40 - Two major cherry species are grown for their fruit, the diploid sweet cherry and the tetraploid sour cherry. Cherries are characterized by high genetic diversity mainly due to their self-incompatibility propagation system. Estimation of the species phenological and genetic diversity has been performed using a number of different traits and marker systems including morphological and anatomical characteristics, as well as isoenzyme and molecular markers. Different molecular markers have been used, spanning from restriction fragment length polymorphisms (RFLPs) to single nucleotide polymorphisms (SNPs), simple sequence repeat (SSR) markers being the most frequently. Moreover, molecular markers have also been used to mark and trace specific agronomic traits, such as self-(in)compatibility (*S*-alleles) or fruit weight thus developing functional markers. The markers reviewed herein will be useful not only for monitoring the genetic diversity in cherry breeding programs, but also for gene conservation, while these or other markers may permit marker-assisted selection for favorable agronomic traits.

Chapter 41 - Soybean purple seed stain (PSS) causes seed decay and purple seed discoloration, resulting in overall poor seed quality and reduced market grade and value. It is a prevalent soybean disease that also affects seed vigor and stem establishment. PSS is caused by the fungus *Cercospora kikuchii* and other *Cercospora* spp. The most common symptom of this disease occurs on the seed. Infected seeds may appear healthy or have discoloration in seed coat varying from pink to light or dark purple spots with range in sizes from a small speck to the entire seed coat. Warm and humid environments favor pathogen growth and disease development. Management strategies for this disease include crop rotation with non-legume or non-host crops, fungicides applications, and tilling the soil to disrupt spore dissemination. Along with these strategies, the use of resistant cultivars may provide more reliable and economical control of PSS, especially when environmental conditions are conducive for disease development. In this chapter, general information about the PSS and an overview of research on germplasm screening and genetic resistance are presented and discussed.

Chapter 42 - The tissue cryopreservation represents an interesting tool for the conservation of animal biodiversity. The establishment of tissue banks has been indicated as a practical approach to the preservation of species and, associated with other biotechniques, it could provide the rescue or multiplication of endangered species. In general, a large number of wild species have been having their gonadal and somatic tissue cryopreserved and for this purpose, the vitrification is the method routinely used. There is a diversity in cryobiological properties and requirements among cell types within tissues, presenting a challenge for its procedure. Nevertheless, even with those obstacles, studies have shown satisfactory results in many wild mammalian species. The gonadal tissue use involves the possibility for the reestablishment of endocrine functions of the testes and ovary allowing the preservation and posterior use of spermatozoa and/or spermatogonial stem cell and oocytes for other assisted techniques. Recent developments in the autografting and xenografting of testes and ovary clearly demonstrated the potential value of cryopreserving gonadal tissue. Already on the somatic tissue, skin samples have been widely utilized because of the possibility of sampling a large group of animals, without a dependency of limitations regarding gender or age. Moreover, this tissue can be obtained quite easily at using a simple methodology with a reduced cost. In this sense, this chapter highlights the importance of applying tissue cryopreservation to wild mammals conservation at showing the most recent studies in this area and the perspectives for its use in conservative programs.

Chapter 43 - The author sequenced mitochondrial genes (*COI*, *COII*, *Cyt-b*) of accepted Latin America tapir species (*Tapirus pinchaque*, *T. terrestris* and *T. bairdii*) as well as an alleged new species, *T. kabomani*. The mountain tapir (*T. pinchaque*) is a relatively rare large mammal species. Some population censuses indicate that no more than 2,000 mountain tapirs are left in the wilderness areas of Colombia, Ecuador and Peru. Our results showed that the gene diversity levels are medium to low with respect to other mammals sequenced for the same or similar genes. However, these gene diversity levels are not impoverished, which means that the genetic situation of this species is not as critical as its population censuses suggest. It will be crucial to determine the gene diversity levels in certain populations not included in the current work (the eastern and, possibly, western Andean Cordilleras in Colombia as well as the Tabaconas Namballe National Sanctuary in Peru), because they are probably the smallest populations of this species. On the other hand, the lowland tapir (*T. terrestris*), the species with the largest geographical distribution in Latin America, showed the highest gene diversity levels of all the other tapir species studied. Additionally, the genetic structure of *T. terrestris* is clearly more robust than that of *T. pinchaque*. Different geographic populations of both species showed different demographic trends throughout time. Our results including five samples of *T. kabomani* showed this taxon to be a haplogroup within *T. terrestris*, reducing the likelihood of *T. kabomani* being a new full species. Finally, the author also analyzed the influence of diverse Pleistocene climatic changes on the mitochondrial haplotype diversification of *T. terrestris* and *T. pinchaque*. The Pleistocene Refugia and the Recent Lake hypotheses probably played integral roles in the evolutionary history of *T. terrestris*. In contrast, the Pleistocene Refugia hypothesis involving the Andes, which probably played an important part in the genetic diversification of other mammals, did not have a significant impact on *T. pinchaque*.

Chapter 44 - Plants are responsible for a significant part of food supply for the entire world, and through agriculture they play an extremely important socio-economic role for the mankind. Therefore, the development of genetically improved crops becomes even more relevant for it aims at an everlasting enhancement of agronomic traits of interest. For many years, plant genetic improvement program has been based in empiric selection of the target traits; however, significant advances were obtained in the last years. Many tools, allowing crops to be improved with greater optimization of the time needed to reach the necessary modifications, are currently available. Regarding the methods used in the genetic improvement, molecular studies have been essential to identify which genes are important for each specific agronomic trait, such as those related to tolerance to abiotic stress. Such studies contribute not only to a better understanding of the endogenous defense mechanisms of plants at molecular level by which these organisms adapt when facing hostile conditions, but also contribute to the generation of stress-tolerant crops by genetic engineering. These programs aim a significant productivity and sustainability that can be reached through soil preservation that is directly related to less necessity of farm inputs. Better adapted crop cultivars make it possible, as well as better use and decontamination of water resources. In this chapter the author attempt at providing an overview regarding strategies that have been used for prospection of genes related to the response of plants to abiotic stress. The combination of biotechnological and bioinformatics tools used in the identification of stress-related genes and development of genetically engineered crops by silencing and/or over-expression of specific genes will be presented in this chapter. Emphasis will be given to the drought and salinity that represent a major part of abiotic

stresses by which the plants are often exposed, leading to serious production losses in many important crops worldwide.

Chapter 45 - Proteinuria is the hallmark of diabetic nephropathy (DN) and vastly increases the incidence of cardio-vascular disease and mortality. Traditional Chinese Medicine (TCM) has been used for diabetes and its complications for thousands of years and appears to be promising in the treatment of proteinuria in DN patients. Clinical trials evaluating TCM for proteinuria, either used as a monotherapy or in combination with western medicine, has produced positive results. Although a large number of clinical studies have been conducted, the clinical evidence with regard to TCM for proteinuria in patients with DN remains inconclusive. The recent progression of evidence will be introduced in this chapter. Current basic research has disclosed that TCM might affect a variety of genes regarding DN etiology. This chapter will analyze recent achievements in this area and address the issue of the association between clinical evidence and the genetic effects of TCM.

Chapter 46 - Ansamycins are composed of secondary metabolites possessing a high degree of activity against numerous types of Gram-positive and Gram-negative bacteria. Structurally, ansamycins are characterized by the presence of core structures including an aromatic moiety (benzene or naphthalene derivative) and an aliphatic chain. Most ansamycins were isolated and characterized from Actinomycetes, while a few were mined in higher plants, for example, maytansine and colubrinol. Due to the development of microbiological techniques, genetic engineering and recombinant proteins, a wealth of different types of ansamycins have been mined along with their biosynthetic gene clusters. This resulted in developed strategies for enhancement of ansamycin production as well as synthesis of novel structure derivatives. In this chapter, the author will describe the biochemistry and genetics of the most important members of ansamycin antibiotics and their applications as cytotoxic, anti-tumor, anti-parasitic and anti-bacterial agents. Thereafter, the future of ansamycins will be discussed to outline the most critically applied aspects in the last.

Chapter 47 - Ubiquitously expressed protein products of *BRCA1* and *BRCA2* genes are implicated in processes fundamental to all cells, including DNA repair and recombination, checkpoint control of cell cycle, and transcription. *BRCA* gene mutations lead to disruption of *BRCA* proteins in mutation carrier cases and induce susceptibility to specific types of cancer. Among women with germline *BRCA* mutations near 50% of mammary malignancies are triple negative breast cancer (TNBC) presenting with a high grade histologically. Among women with breast cancer, TNBC was established in 57.1% of *BRCA1*-mutation positive and in 23.3% of *BRCA2*-mutation positive cases, whereas in only 13.8% of *BRCA*-proficient women. Although *BRCA* gene mutation carrier women usually exhibit clinical symptoms of defective estrogen receptor (ER) signaling; such as anovulatory infertility and early menopause, the serum estrogen levels of these patients are consequently elevated. In these cases, a compensatory feedback mechanism aims to break through the inherited or acquired ER resistance by increased estrogen synthesis so as to maintain the cellular estrogen surveillance. The higher the estrogen overproduction of *BRCA*-mutation positive cases, the higher the possibility of tumor-free survival. In conclusion, *BRCA1* and *BRCA2* gene mutations seem to increase the breast cancer risk, particularly that of TNBC, in case of insufficient compensation of defective ER signaling. Upregulation of these genes by means of elevated estrogen levels of high parity, artificial hormonal cycle created by oral contraceptives or a pregnancy mimicking high estrogen dose may decrease the excessive cancer risk of *BRCA* mutation positive women.

Chapter 48 - Rett syndrome (RTT) is a neurodevelopmental disorder mainly caused by mutations in the MECP2 gene affecting around 1 in 10,000 female births. Mutations in the MECP2 gene have been associated with the onset of RTT. Clinical manifestations include severe linguistic and motor impairments that are the core of phenotype symptoms. Some patients show a moderate level of conservation of linguistic functions while others lose the use of functional verbal communication. The objectives of the present chapter are to study in depth the latest theoretical approaches to the link between linguistic processes and the specific RTT genotype. This chapter begins with a theoretical overview on cognitive alterations and then focuses on linguistic specific impairments characterized by the loss of articulation or the production of few functional sounds. A restricted sample shows the presence of verbal speech (Preserved Speech Variant). Renieri et al. (2009) proposed the term “Zappella variant” rather than “preserved speech variant” to describe milder forms of RTT, because other aspects, besides speech, are involved. The second part proposes a preliminary research which analyses the correlation between linguistic phenotype and specific genotype.

Chapter 49 - Marfan Syndrome was originally described by Antoine Bernard-Jean Marfan in 1896 and is an uncommon inherited connective tissue abnormality which occurs as an autosomal dominant genetic disorder with frequent mutations. The patients have normal mentation but have a characteristic increase in height, abnormally long limbs, arachnodactyly, joint hypermobility, distinctive facial features, scoliosis, ectopia lentis, dural ectasia and array of aortic and cardiac abnormalities that are often life threatening. The genetic cause of the disease has been identified. Although some medical and surgical treatments are currently in practice they are not always helpful. The orthopaedic problems are often significant and sometimes require complex surgical interventions.

Chapter 50 - Marfan syndrome (MFS) is a systemic connective tissue disorder that is caused by mutations in the extracellular matrix protein fibrillin-1. While MFS is considered to be at high risk of dental disorders and cardiovascular disease (CVD), little causal relationship has been provided to date. In this article, the author reviewed the prevalence of periodontitis in patients with MFS to assess the relationship between periodontal bacterial burden and CVD in MFS patients.

Chapter 51 - Introduction: Pectus deformities can coexist with cardiovascular diseases. This association is well known in tissue conjonctive disorders such as Marfan's syndrom. Combined procedures can be performed safely and represent an interesting alternative in such situations. Valve sparing aortic root replacement has excellent long term outcomes and has become an increasigly popular alternative to aortic root replacement especially in young marfan patients to avoid lifelong anticoagulation. The author present our serie of single-stage pectus correction and cardiac surgery, and emphasize the role of aortic valve sparing interventions in such situations by a review of the literature. Methods: A retrospective review was conducted of patients who underwent chest deformity repair and cardiac surgery at the same time from January 2007 to May 2014. All datas were collected propestively in our data base. A review of literature was conducted to collect all published cases of combined valve sparing root replacement and correction of a chest wall deformity. Results: Including our serie (4 patients) 12 patients underwent a combined Tirone David and chest wall deformity correction. 10 patients underwent a Nuss procedure, 2 patients a modified Ravitch procedure. Conclusion: Combined technique of valve sparing aortic root replacement and correction of a chest wall deformity especially by Nuss technique is safe and effective. This strategy has excellent mid-term results for both aortic and chest wall pathologies.

Chapter 52 - Marfan syndrome frequently causes cardiac complications such as, aneurysm and dilatation of the aortic root. Many Marfan syndrome patients have these cardiovascular problems, and the surgical replacement of aortic and mitral valve, and aortic roots is frequently required. In addition, cases are associated with severe periodontitis, which is a chronic inflammation of the gingiva, periodontal ligament, and alveolar bone. Because of the surgical replacement, it is essential to prevent dental infection, such as infectious endocarditis caused by the periodontitis. In Marfan syndrome, an unfavorable oral hygiene due to the crowded teeth and narrow dental arch had been thought as a cause of severe periodontitis. However, clinical and basic studies have highlighted the genetic background as a pathogenesis of the severe periodontitis. It is suggested that cell alignment and tissue architecture of periodontal ligament are impaired in the model mice of Marfan syndrome. The model mice were more susceptible to alveolar bone resorption after the infection of *Porphyromonas gingivalis*, which is known to cause chronic periodontitis. It is likely that activated TGF- β signaling upregulates IL-17 and TNF- α levels, resulted in the increased alveolar bone resorption. In this review, the perspective of the dental management and the effect of angiotensin II receptor blocker are discussed.

Chapter 53 - Preimplantation genetic diagnosis (PGD) was introduced 24 years ago with the purpose of performing genetic testing before pregnancy, in order to establish only unaffected pregnancies and avoid the need for pregnancy termination, which is the major limitation of traditional prenatal diagnosis. Despite the requirement for ovarian hyperstimulation and in vitro fertilization (IVF), needed to perform genetic testing of oocyte or embryo prior to transfer, PGD has been accepted in most parts of the world. Thousands of PGD cycles have now been performed for single gene disorders, with PGD presently offered for some indications that have never been practiced in prenatal diagnosis, such as late onset diseases with genetic predisposition, and preimplantation HLA typing. The present paper describes our experience on PGD for Marfan syndrome, caused by *FBN1* gene, which was performed in 38 cases, as part of our PGD experience of 2,860 cycles for single gene disorders, which is the world's largest PGD experience.

Chapter 54 - The order Anura currently encompasses over 6,800 amphibian species distributed in 56 families and is an interesting group for cytogenetic studies. Whereas some groups of species present conservative karyotypes, others are highly variable in diploid number or number/location of diverse chromosomal markers, such as nucleolus organizer regions, heterochromatic bands and specific satellite DNA sites. In some cases, karyotypic variation overcomes morphological diversification, turning cytogenetics into a useful tool for taxonomy. Special variation is observed with respect to sex chromosomes and sex determination systems, with both female and male heterogameties observed. Although most species already karyotyped do not show sex chromosome heteromorphism, distinct levels of differentiation are observed between the sex chromosomes of several species, which makes this group particularly interesting for studies of sex chromosome evolution. In this chapter, the author explore the use of cytogenetic data for studies of frogs as well as the insights that hypotheses of phylogenetic relationships have added to this issue. In addition, the author provide a brief review of PcP190 satellite DNA (with new data for the genus *Engystomops*), sex chromosome systems and B chromosomes found in Anura.

Chapter 55 - Humans are daily exposed to a variety of potentially harmful agents in the air they breathe, liquids they drink, food they eat and products they use. Long-standing evidence of the bond between health and environment has led to the recognition for the need of sustainable development. On the other hand, there is an increasing global awareness of the

inevitable limits of individual health care and the need to complement such services with effective public health strategies. According to World Health Organization (WHO), cancer is a leading cause of death worldwide. Additionally, at least 200 000 people die every year from occupational or work-related cancers. Exposure assessment aims at prevention. Establishing the health effects of various activities and exposures requires information about the levels of exposure and the biological effects resulting from the interaction between the organism and the chemical agent. The resulting data provides a basis for designing effective prevention and mitigation strategies. Cytogenetic endpoints have long been applied in surveillance of human genotoxic exposure and early effects of genotoxic carcinogens. Assays measuring chromosomal aberrations (CAs), micronucleus (MN) and sister chromatid exchange (SCE) in lymphocytes are well-established techniques extensively used in human biomonitoring studies to assess DNA damage at the chromosomal level. The relevance of cytogenetic alterations as a cancer risk biomarker is further supported by epidemiologic data linking CAs and MN with cancer risk in human populations. Thus, the use of cytogenetic markers in human biomonitoring is of paramount importance due to its predictability regarding deleterious effects resulting from the exposure to environmental stressors.

Chapter 56 - The great era for classical cytogenetics started sixty years ago with the description of the twenty-three pairs of human chromosomes and the discovery of the Philadelphia chromosome, the first known chromosomal defect associated with a specific type of cancer. Since then many recurrent chromosomal aberrations linked to specific hematological malignancies have been detected. The vast majority of these abnormalities can be detected by modern molecular genetic methods so it might seem there is no longer a need for microscopy. However, there is still a significant portion of hemato-oncologic patients for which the classical cytogenetic investigation is appropriate, i.e., cases with complex karyotypes. Complex karyotypes are characterized by the presence of three or more unrelated chromosomal aberrations co-existing in a single clone. Their occurrence is associated with adverse outcomes across the entire spectrum of hematologic malignancies. These aberrations are also a powerful diagnostic indicator for molecular targeted therapies, allogeneic stem cell transplantation or other generally more aggressive treatment strategies. Since the presence of complex karyotypes would be missed when using only molecular genetic methods, this highlights the irreplaceable role of classical cytogenetics as a first-tier analysis for the evaluation of complex structural chromosomal abnormalities in hemato-oncologic patients. Ideally, classical cytogenetics is then followed by more precise molecular genetic methods to identify specific chromosomal aberrations more deeply. Here the author focus on the complex karyotype issues in the myelodysplastic diseases, leukemias, lymphomas and multiple myelomas as seen daily in our center.

Chapter 57 - Initially identified as *Astyanax schubarti* on the basis of morphological characteristics, its chromosomal analysis revealed a unique diploid number in the genus. With 42 chromosomes and the impossibility of homologous pairing, the karyotype of the individual was compared to a set of haploid complements of *Astyanax schubarti* ($2n = 36$ chromosomes) and *Astyanax fasciatus* ($2n = 48$ chromosomes). Natural hybrids are rare. The viability of a hybrid between species with such chromosomal discrepancy may offer important hypotheses to explain the morphological, molecular, and cytogenetic diversity of the genus.

Chapter 58 - There are three distinct subtypes of Trichorhinophalangeal syndrome (TRPS); TRPS type I, TRPS type II and TRPS type III. Features common to all three subtypes include sparse, slowly growing scalp hair, laterally sparse eyebrows, a bulbous tip of the nose (pear-

shaped), and protruding ears. The diagnosis of TRPS is based in typical clinical and radiographic features, as well as in the identification of causative mutations in the *TRPS1* gene (TRPS type I and III) and in loss of functional copies of the *TRPS1* and *EXT1* genes (TRPS type II). The mode of inheritance in TRPS I and III is autosomal dominant, however *de novo* deletions of *TRPS1* and *EXT1* genes are the main defect of TRPS II. Parental balanced chromosomal rearrangements are an important cause of interstitial aberrations in TRPS. In cases of cytogenetically-invisible alterations, parental FISH analysis as well as aCGH should be considered as part of the clinical baseline testing. Treatment of clinical problems of TRPS types is mainly supportive and includes ectodermal and skeletal issues. The number of distinct syndromes as TRPS is rapidly increasing and their confirmation is necessary especially in cases that typical features are absent. Clinical geneticists should provide information for the families and advise them how to overcome problems.

Chapter 59 - It is generally believed that genotype and adult lifestyle elements are primary risks of some metabolic diseases such as insulin resistance, obesity and diabetes mellitus in later life. However, increasing evidence demonstrates that early life malnutrition during the period of gestation and/or lactation may increase our susceptibility to such metabolic diseases in later life. The underlying mechanism is still not very clear. Recently, epigenetics is hypothesized to be the important molecular basis of the imbalanced early life nutrition and glucose metabolism disorders, which is known as "Developmental Origin of Health and Diseases" (DOHaD). Currently, there are substantial epidemiological studies and experimental animal models that have demonstrated nutritional disturbances during the critical periods of early life development can significantly impact the predisposition to developing some metabolic diseases in later life. The fundamental mechanism is that early developmental nutrition can regulate epigenetic modifications of some genes associated with development and metabolism. DNA methylation is the first discovered and one important epigenetic modification. MicroRNAs are recognized as an important epigenetic modification and they are a major class of small non-coding RNAs (about 20-22 nucleotides) which can mediate posttranscriptional regulation of target genes with cell differentiation and apoptosis. Recent studies suggest that DNA methylation and microRNAs maybe the crucial modulators of fetal epigenetic programming in nutrition and metabolic disorders. This chapter will focus on how early life nutrition can alter the epigenome, produce different phenotypes and alter disease susceptibilities, especially for impaired glucose metabolism.

Chapter 60 - Oxidative stress is a state in which production of reactive oxygen species exceeds the capacity of antioxidant systems. Reactive oxygen species have one or more unpaired electrons, making them highly reactive with other cellular molecules such as protein, lipid, and nucleic acid. The peroxidation of polyunsaturated fatty acids in biological membranes results in impaired membrane integrity. Oxidatively modified proteins lose *their* capacity to carry out the *physiological functions* and they may form intracellular aggregates. Attacks of reactive oxygen species to DNA results in strand breakages and base oxidation. Major DNA oxidation product is 8-hydroxydeoxyguanosine which has a pro-mutagenic potential. Due to these damaging effects, oxidative stress plays an important role in various pathologies such as cancer, diabetes, chronic inflammatory diseases and neurodegenerative disorders. Epigenetic changes are regular and natural events which regulate gene expression without changing base sequences on DNA. Dysregulation of regular epigenetic mechanisms is a contributory factor for many of human pathologies. Recently, reactive oxygen species have been shown to cause epigenetic dysregulations that play a pivotal role in human disorders. The basic epigenetic

mechanisms and their dysregulation by reactive oxygen species have been reviewed in this chapter.

Chapter 61 - Cancer is one of the deadliest malignancies that have plagued mankind for decades. Epigenetic mechanism(s) play a central role in the homeostasis of normal cell proliferation and differentiation. Global epigenetic modifications are frequently associated with cancer initiation, progression, as well as metastasis. These changes include DNA methylation, histone lysine methylation/demethylation, acetylation/ deacetylation, including methylation and acetylation of non-histone proteins, and can alter the expression of various oncogenic signaling cascades, which in-turn can lead to uncontrolled proliferation. In this chapter, the author primarily focus on the major epigenetic changes that occur in oncogenes, tumor suppressor genes, transcription factors and cancer stem cells, which in turn mediate tumor growth. These modifications are controlled by regulatory enzymes such as DNA methyltransferase, histone acetyltransferases, histone deacetylase, lysine acetyltransferase, and arginine and lysine methyl transferases. In addition, the author also describe a few selected pharmacological agents that can modulate the action of these enzymes and display significant potential for cancer therapy.

Chapter 62 - Alzheimer's Disease is one of the most common neurodegenerative disorders. Many efforts have been directed to prevent AD due to its rising prevalence and the lack of an effective curative treatment. Epigenetic changes are involved in regulation of gene expression, and may mediate various pathologies. Epigenetic changes are reversible so that can be easily modulated. Modulation of dysregulated epigenetic mechanisms is a promising therapeutic approach for many diseases. There is some evidence for epigenetic dysregulation at various levels contributing to AD pathogenesis. Despite the recent rapid accumulation of knowledge about AD pathogenesis, the role of epigenetic modifications has not been understood exactly. This chapter provides a brief overview about the role of epigenetic changes that are linked to AD pathogenesis and emerging targets for new therapeutic strategies in this field.

Chapter 63 - Cardiovascular disease (CVD) is not a single condition, but an umbrella term used to describe a range of common diseases affecting the heart and the circulatory system. The term commonly includes diseases of the cardiac muscle and of the vascular system supplying the heart, brain, and other vital organs. Many of these conditions can be life-threatening. CVD is the leading cause of death in the world, affecting all populations, irrespective of demographic or socioeconomic differences and is responsible for one third of all deaths. In the present chapter, the author focus on the epigenetic control of embryonic cardiac development and the role of epigenetic mechanisms in CVD observed from results in human, animal and cell culture studies' approaches. The author discuss the main epigenetic mechanisms involved in heart development and major CVDs such as coronary heart disease (CHD), heart failure (HF), myocardial infarction (MI), hypertension, stroke, arrhythmias, cardiomyopathy and cardiac hypertrophy (CH). Additionally, this chapter also focuses on the epigenetic modifiers that are involved in the development of CVD, and the potential utility of epigenetics-based therapeutic strategies in CVD.

Chapter 64 - Extensive characterization has been performed on the genes and genetic mutations that are involved in spermatogenesis and male infertility, but the vast role of the sperm's transcriptome and epigenome in male reproduction has yet to be completely explored. Recent research has established that epigenetic remodeling of the sperm is necessary for development and for its function following fertilization. The histone- retained regions of the sperm have been recently shown to carry the bivalent marks of activating histone H3 lysine K4

trimethylation and the repressive H3 lysine K27 trimethylation. These bivalent histone marks facilitate the dynamic changes in stage-specific gene expression during sperm development. The paternal epigenome bears unique and important epigenetic modifications determined to be potentially important to the developing embryo, but the complete scope of its contribution is still emerging. Additionally, advances in assisted reproductive techniques have also suggested that alterations in the epigenetic profile in infertile men are transmitted to the developing embryo. This review will highlight the latest advances in epigenome profiling of the chromatin modifications during the development of immature to a mature sperm as well as provide a glimpse into the future role of epigenetic mechanisms in the generation of new germ cells/gametes from induced pluripotent stem cells, to treat male infertility.

Chapter 65 - Epigenetic is defined as the study of mitotically and/or meiotically heritable changes in the gene function without changing DNA sequence and playing crucial roles both in normal development and human diseases. The molecular basis of epigenetic process consists of histone modifications, DNA methylation, positioning of histone variants, and non-coding RNAs. Genome-wide patterns of DNA and chromatin modifications together called as 'epigenomes'. Epigenomes undergo precise, coordinated, reversible changes through the developmental stages and so contributes to the lineage and tissue specific expression of genes. In addition to the tissue specific impact, environmental factors such as nutrients, toxins, infections and hypoxia can also influence epigenomes. Distinct or global changes in the epigenetic landscape are hallmarks of chronic inflammation associated diseases. Alteration in methylation status of CpG sites, monoallelic silencing, and other epigenetic regulatory mechanisms have been observed in key inflammatory response genes. Epigenetic changes including DNA methylation, histone modification and noncoding RNA expression, were found associated with acute and chronic inflammatory disorders. Recently, therapies targeting epigenetic mechanisms are trending options for the treatment of chronic and degenerative disorders. Epigenetic drugs that are applied on animal models and some clinical trials are displaying positive therapeutic effects. Not only mono therapies but also combined usage of HDAC or DNMT inhibitors could be the next step for the epigenetic therapeutic modulations. Scope of this chapter is to provide an overview of the epigenetic modifications in inflammation and inflammation driven diseases.

Chapter 66 - Obesity is a public health problem leading to morbidity and mortality throughout the world. It arises from the interactions between genetics and environmental factors. In recent years, susceptibility to obesity has been also linked to epigenetic factors. Epigenetics, is the study of heritable changes in gene expression which do not involve in the underlying DNA sequence. The epigenetic mechanisms include DNA methylation, covalent histone modifications, chromatin folding, miRNAs, and polycomb group repressive complexes. Both dietary factors and individual behaviors affect obesity development via epigenetic mechanisms. Epigenetic mechanisms are also linked to programmed changes in gene expression as a result of early environmental exposures during pregnancy which alter offspring growth and development. There is evidence that nutrient and environmental exposures during pregnancy may affect fetal/newborn development and result in offspring obesity or metabolic syndrome which is a cluster of metabolic abnormalities. Obesity related genes, epiobesigenes, display methylation patterns playing important roles in the development of obesity which are potential future epigenetic biomarkers of obesity. The susceptibility genes have been reported as FGF2, PTEN, CDKN1A, and ESR1, functional in adipogenesis; SOCS1 and SOCS3, functional in inflammation, and COX7A1, LPL, CAV1, and IGFBP3 which are functional in

fat metabolism and insulin signaling. It is important to prevent this era's epidemic, obesity, since it leads to chronic diseases including hypertension, atherosclerosis, insulin resistance, and moreover metabolic syndrome. It may be easier to prevent the progress of the disease by revealing the epigenetic mechanisms especially methylation profiles of the susceptibility genes.

Chapter 67 - Entomopathogenic fungi have been mentioned as one of the best alternatives for insect pest control. These fungi cause insects death. More than 750 fungal species have been described infecting insects, some of the most utilized for insect control are: *Beauveria bassiana* and *Metarhizium anisopliae*. This is evidence of cosmopolitan distribution of entomopathogenic fungi and its evolutionary success, this type of fungi is related to a serie of interactions among fungi, plants, and insects. In this chapter, the main objective is to review and discuss the most recent information on genetics and evolution of entomopathogenic fungi. In this chapter are covered the following themes: Entomopathogens role in nature, Entomopathogenic fungi and their interactions with the insect immune system, Isolation and identification of entomopathogenic fungi, Genetic diversity among strains of entomopathogenic fungus, Genes involved in virulence of entomopathogenic fungi, Molecular phylogeny of entomopathogenic fungi and their biogeographic implications, Evolution of entomopathogenicity in fungi, Genetic improvement of entomopathogenic fungi for insect biocontrol, Future trends and Conclusions.

Chapter 68 - The author sequenced the mitochondrial (mt) *ND5* gene of 100 specimens of *Eira barbara* (Mustelidae, Carnivora). The samples represented six out of the seven putative morphological subspecies recognized for this Mustelidae species (*E. b. inserta*, *E. b. sinuensis*, *E. b. poliocephala*, *E. b. peruana*, *E. b. madeirensis*, and *E. b. barbara*) throughout Panama, Colombia, Venezuela, French Guiana, Brazil, Ecuador, Peru, Bolivia, Paraguay, and Argentina. The main results show that the genetic diversity levels for the overall samples and within each one of the aforementioned putative taxa were very high. The phylogenetic analyses showed that the ancestor of the Central and South-American *E. barbara* originated during the Miocene or Pliocene (6.3-4 millions of years ago, MYA). Furthermore, the ancestors of some geographical groups, (we detected at least four) originated during the Pliocene (3.7-2.5 MYA). These four groups (or lineages) were placed in the Cesar-Antioquia Departments (northern Colombia), Bolivia and northwestern Argentina, northern-central Peru, and in the trans-Andean area of Ecuador. However, during the Pleistocene, this species experienced a strong population expansion and many haplotypes expanded their geographical distributions. They became superimposed on the geographical areas of older geographical groups that originally differentiated during the Pliocene. Until new molecular studies are completed, including those with nuclear markers, the author proposed the existence of only two subspecies of *E. barbara* (*E. b. inserta* in southern Central America, and *E. b. barbara* for all South America). All of the demographic analyses showed a very strong population expansion for this species in the last 400,000 YA during the Pleistocene.

Chapter 69 - Like other forms of diagnostics, genetic testing comes with a retinue of costs and benefits. Significant benefits in terms of morbidity and mortality have accrued to individuals tested for more prevalent genetic conditions like cystic fibrosis and sickle cell disease, including persons seen in the emergency room or identified through public health surveillance. These benefits do not mitigate the drawbacks of genetic testing, false and missed diagnoses and sheer cost among them. Both medicine and public health have aimed at means of maximizing genetic test benefits in the interventions that they apply. The President's Precision Medicine Initiative (PMI) holds promise in that its results could be used to tailor

medical treatments to the individual characteristics of patients, “precision” implying a more accurate and precise regimen overall. The National Cancer Institute (NCI) has already launched the NCI-MATCH precision medicine trial, which assigns targeted treatments based on the genetic abnormalities in a tumor, regardless of cancer type. Other trials, such as the NCI Pediatric MATCH trial, are yet to happen. The efficacy of cancer treatments also intersects public health concerns. The Evaluation of Genomic Applications in Practice and Prevention (EGAPP) Working Group has evaluated the use of *UGT1A1* genotyping to determine the best dose of irinotecan to prevent side effects when treating patients for metastatic colorectal cancer. Analytic validity does not always equate with improved patient outcomes, however, thus the public health emphasis on development of a suitable evidence base for precision medical and public health efforts. The public health approach to precision medicine, or “precision public health”, differs from the medical approach in several important ways: (1) population-based with attention to at-risk populations, as opposed to being strictly individualized; (2) focus on primary and secondary prevention, rather than frank disease (tertiary prevention); and (3) prioritizing interventions that have already demonstrated readiness for large-scale implementation, in contrast to the undertaking of novel clinical trials. Precision public health is exemplified in the Centers for Disease Control and Prevention’s emphasis on the implementation of Tier 1 genetic tests that have passed systematic review for analytic and clinical validity and utility – the use of family history for referral for hereditary breast and ovarian cancer genetic testing (*BRCA1/2* mutations), and hereditary nonpolyposis colorectal cancer cascade screening (Lynch syndrome *MLH1*, *MSH2*, *MSH6* mutations). This paper will cross-compare the precision medical approach to cancer based on pharmacogenomic regimens using companion diagnostics, and the public health approach to precision management of hereditary cancer for 3 cancer types – lung, breast, and colorectal. It will describe methods of early detection and consider how lives can be saved through precise management – from predictive testing and cancer monitoring of the at-risk population, to tailored chemoprevention that fits the needs of the individual. In the population context, a cascade screening “multiplier effect” exists in that relatives can also be assessed and followed for mutations identified in the proband. Cost-benefit analyses (T4 translational research) of medical and public health approaches will be closely examined and compared. Points of commonality between the two approaches will also be discussed, since primary/secondary and tertiary disease prevention represent a continuum. These analyses point to the value of allocating resources towards the health of at-risk populations. Questions remain if particular forms of genetic testing are to become “universalized”, and if the needs of all at-risk groups, including racial-ethnic, are to be addressed.

Chapter 70 - Williams syndrome (WS) is a genetic neurodevelopmental disorder (prevalence close to 1 in 20,000-30,000 births) resulting from the deletion of 16-25 genes on the long arm of Chromosome 7. Individuals with WS have an intelligence quotient of 40-70. Theirs is a unique neuropsychological profile, characterized by an apparent dissociation between cognition and language, as language is relatively well preserved, compared with other cognitive skills. However, a more complex profile is now emerging, with good lexical, short-term memory (especially auditory-verbal) and face processing skills, but visuospatial (especially local processing of information), executive (planning and inhibition), memory (working memory and long-term) and attentional deficits. Individuals with WS also have specific auditory-perceptual cognitive skills (hyperacusis), category-specific perception of speech sounds, and musical skills that exceed their cognitive level. Other behavioral characteristics include hypersociability. In this chapter, the author provide a review of the

literature on the specific neuropsychological profile of individuals with WS, from a cognitive-behavioral and neuroanatomical point of view. Early studies of the neuropsychological profile in WS focused on the dissociation between cognition and language. Since then, research has shown this syndrome to be more complex. Our aim is to highlight the heterogeneity of the cognitive profiles observed in this syndrome, and to identify the factors that might explain this heterogeneity. The complexity and specific features of the neuropsychological profile in WS need to be understood in order to develop therapeutic and learning methods adapted to the developmental pace of individuals with WS.

Chapter 71 - Chromosome rearrangements are the most common genetic abnormalities in humans. Abnormal chromosomal configurations are formed among non-homologous chromosomes to allow full synapsis of homologous chromosomes at meiosis I of chromosome rearrangement carriers. These abnormal chromosomal configurations result in the production of gametes with various chromosomal complements due to malsegregation of derivative chromosomes or recombination. Most of the gametes have unbalanced chromosomal complements and only a small number of gametes have normal or balanced chromosomal complements. Carriers of balanced chromosome rearrangements are phenotypically normal but they are at an increased risk of abnormal pregnancies due to the unbalanced gametes. However, chromosome rearrangement carriers who cannot have babies or experience repeated abortions or abnormal pregnancies can have healthy babies after introduction of PGD. Balanced reciprocal translocation is the most common chromosome rearrangement. Thirty-two types of gametes can be produced in meiosis of reciprocal translocation carrier. According to the results that analyze the meiotic segregation of embryos from PGD cycles of reciprocal translocation carriers, 2:2 segregation is the main segregation mode. The meiotic segregation might be affected by the gender of carriers. The frequency of balanced embryos was not different between female and male carriers. However, the frequencies of 2:2 segregation, especially adjacent-1 segregation, 3:1 and 4:0 segregation were significantly different between female and male carriers. Robertsonian translocation is also common chromosome rearrangement in humans. Gametes with eight different chromosomal complements can be produced in Robertsonian translocation carriers. The frequency of balanced embryos was higher in male carriers than in female carriers. Carriers of complex chromosome rearrangements (CCR), very rare chromosome rearrangements, can achieve pregnancy by PGD although the number of normal or balanced embryos was extremely low. Although it is very difficult to estimate the rate of normal or balanced embryos, the rate is estimated to be less than 10% in PGD for CCR carriers. The 3:3 segregation and chaotic segregation (meiotic segregation whose meiotic segregation cannot be defined) are prevalent segregation modes in carriers of three-way translocation. In PGD cycles of CCR carriers, cycle cancellation is very frequent due to the absence of the normal or balanced embryos. Therefore, it is important that a large number of embryos are obtained in one cycle. Occasionally, abnormalities of chromosome rearrangement-unrelated chromosomes are observed in embryos or abortuses although chromosome rearrangement-related chromosomes are normal or balanced. At present, PGD that diagnose all 24 chromosomes using array-CGH, SNP array or NGS is carried out worldwide. In those PGD cycles, the risk that embryo transfer is cancelled might be increased. However, the rate of normal or balanced embryos is unexpectedly increased and pregnancy rate is improved. Surely, PGD is the very effective assisted reproductive technique in achieving pregnancy of chromosome rearrangement carriers by preventing repeated abortions that chromosome rearrangement carriers usually experience. In the field of PGD for chromosome

rearrangements, the next stage will be the development of technique that can discriminate between normal embryos and balanced embryos. Chromosome rearrangements, such as translocation (reciprocal or Robertsonian), inversion and complex chromosome rearrangements, are the most common genetic abnormalities in humans. Chromosome rearrangements are classified as balanced or unbalanced. Carriers of balanced chromosome rearrangements have the normal chromosomal complements but carriers of unbalanced chromosome rearrangements have additional or missing chromosomal material. Generally, the incidence of balanced chromosome rearrangements is about 0.19% of newborns. Carriers of balanced chromosome rearrangements are phenotypically normal because they have all the genetic materials. However, they are at an increased risk of implantation failure, repeated abortions or birth of chromosomally unbalanced offspring. Of course, not all carriers of balanced chromosome rearrangements experience the abnormal pregnancies. The abnormal pregnancies are resulted from the unbalanced gametes generated during meiosis of chromosome rearrangement carriers. Abnormal chromosomal configurations are formed to allow full synapses of homologous chromosomes during meiosis of balanced chromosome rearrangement carriers. The unbalanced gametes are produced as a results of malsegregation of these abnormal chromosomal configuration. Most of the gametes produced during meiosis of chromosome rearrangement carriers have unbalanced chromosomal complements. These unbalanced gametes result in implantation failure, repeated abortions or birth of chromosomally unbalanced offspring. After the first successful clinical application of preimplantation genetic diagnosis (PGD) in 1990, PGD has been widely used worldwide to select normal or balanced embryos in *in vitro* fertilization and embryo transfer (IVF-ET) programs of balanced chromosome rearrangement carriers. Cleavage-stage fluorescence in situ hybridization (FISH) has been widely used to PGD for carriers of balanced chromosome rearrangements till the early 2010s and PGD based on array comparative genomic hybridization (CGH) or next generation sequencing (NGS) is widely applied nowadays. The only abnormalities of rearranged chromosomes can be diagnosed in FISH-based PGD and the abnormalities of other chromosomes cannot be diagnosed. However, abnormalities of all 24 chromosomes can be diagnosed after the application of array CGH or NGS into PGD.

Chapter 72 - *Aims*: Attention deficit hyperactivity disorder (ADHD) is a multifactorial psychiatric and neurobehavioral disorder. The brain-derived neurotrophic factor gene (*BDNF*) has been proposed as a strong candidate for this pathology. The aim of this study was to determine a family-based association between three polymorphisms of the *BDNF* gene and the ADHD in a Tabascan-Mexican population. *Methods*: The author analyzed the rs6265, rs12273363 and rs11030119 polymorphism of the *BDNF* gene through a family-based association study. A total of 105 individuals grouped in family-trios (mother, father and ADHD patient) were studied. Allelic and haplotypic transmission were assessed through transmission disequilibrium test (TDT), using HaploView software. *Results*: No statistically significant association was observed between the *BDNF* gene polymorphisms and the ADHD etiology in Tabascan-Mexican families: rs6265 ($\chi^2 = 1.33$; $p = 0.24$); rs12273363 ($\chi^2 = 1.33$; $p = 0.24$); rs11030119 ($\chi^2 = 0.66$; $p = 0.41$). Furthermore, no preference of transmission was observed for any of the haplotypes. *Conclusions*: It was not possible to prove any association between the *BDNF* gene polymorphic variants and ADHD in a Mexican population. Future studies comprising larger samples are necessary to determine the potential role of the *BDNF* gene in ADHD.

Chapter 73 - Petroleum can be characterized as an oily, flammable substance, less dense than water, with a distinctive scent and color ranging from black to dark brown. The petroleum industry includes some global processes that can be highlighted by the environmental risks they present. At offshore platforms, the danger of a spill is aggravated by the density of the oil, which floats and is carried quickly through sea currents. Thus, in addition to the marine fauna, coastal fauna and flora (e.g., mangroves and estuaries) can also be affected by a leakage. Petroleum is predominantly composed of hydrocarbons, these can be degraded by several species of bacteria, which are of great interest for the bioremediation of contaminated environments. Among these species the author can mention *Pseudomonas*, *Sphingomonas*, *Mycobacterium*, *Microbacterium* and *Gordonia*. The success of a bioremediation process depends on numerous factors such as microbial biomass, population diversity, enzymatic activity, pH, temperature, and carbon source. Moreover, the strains have their genetic potential for bioremediation investigated through methods of analysis concerning genes related to the degradation of aliphatic and aromatic hydrocarbons mostly by oxygenases, such as the polycyclic aromatic hydrocarbon ring-hydroxylating dioxygenases (PAH-RHD_o), and *alkB* (for n-alkane degradation) genes. The study of autochthonous microbial communities is of crucial importance for the understanding of the genetic and biotechnological potential of bioremediation in environments close to the areas of extraction and susceptible of contamination; it is also of great interest for industry and biotechnological development. This chapter covers the main aspects of petroleum, regarding its exploitation aspects, the process of geological formation, and environmental impacts. It will include topics in microbiology and genetics; for instance, metabolic pathways of hydrocarbon biodegradation, biofilm dynamics associated to the oil industry, metagenomics of communities in marine environments, corrosion influenced by microorganisms (CIM), microbial control methods related to the petroleum industry, and bioremediation will be discussed.

Chapter 74 - In healthy individuals, the incessant activity of regulatory genetic mechanisms ensures the metabolic and proliferative equilibrium of cellular activity. In case of accidental defects occurring in any part of the system, a coordinated counteraction of numerous mediators may successfully help in the restoration of physiologic processes by means of either overexpression or hyperactivity. Even serious defects of genome stabilizer mechanisms may be kept in balance for a long duration, showing the clinical signs of good health. By contrast, due to the exhaustion of the compensatory processes, DNA defects may develop and lead to the clinical manifestations of diseases. Estrogen activated estrogen receptors (ERs) are the primary initiators and organizers of the up-regulatory circle of genome stabilization in correlation and crosstalk with aromatase enzyme and genome safeguarding proteins, such as BRCA_s. The promoter regions of *ESR1*, *BRCA1*, and *CYP19* aromatase genes exhibit strong triangular partnership for the harmonized regulation of the synthesis of ERs, BRCA proteins and aromatase enzyme, which can be reconstructed from the meticulous details of earlier scientific results. Considering the extreme capacities of ER-signaling for self-restoration, it is obvious that antiestrogen treatment, either ER binding by a false ligand or inhibition of estrogen synthesis, may provoke extreme compensatory actions in genetically proficient cases. Analyses of the results of genetic studies on tumor cells have shown that upregulation of ER-signaling induced by natural estrogen or antiestrogen is a beneficial defensive process even in tumor cells, promoting their domestication and elimination. A schematic representation of the main stream of upregulative genome stabilizing circle visualizes the possibilities for extreme counteractions against the toxic effects of antiestrogens. The presented complex genome

stabilizer mechanisms reveal that cancer cells may preserve their residual capacity for the upregulation of ER-signaling, even if the efforts are not satisfactory.

Chapter 75 - Rett syndrome (RTT) is a rare, neurodevelopmental genetic disorder that develops in early childhood and influences many functions within neurobehavioural domains. The core of phenotype symptoms includes severe linguistic and motor impairments. The onset of RTT is characterised by a gradual or sudden loss of speech and hand function followed by a slow decrease in acquired gross motor skills with subsequent severe functional dependence. RTT is associated primarily with mutations in MECP2, a gene located on the long arm of the X chromosome (Xq28). The severity of impairments depends not only on genotype, but on the extent of X inactivation. However, CDKL5 and FOXP1 gene mutations have been also identified in girls affected by atypical RTT. Despite sharing neurological features, subjects with RTT present considerable clinical variability. Research on effects of genotype of RTT is expanding in many directions. The current chapter, will discuss the correlations between genotype and motor abilities in subjects with RTT. The main aim of this chapter is to relate functional outcomes, in particular motor impairments, to mutation type in patients with RTT. This chapter begins with a theoretical overview on genetic alterations in RTT and then focuses on motor specific impairments. In the second part of this chapter, the author propose a preliminary research which analyzes the correlation between motor phenotype and specific genotype.

Chapter 76 - Fragile X syndrome (FXS) is the most common inherited cause of intellectual disability. However, findings reported in cross-sectional studies on this population are heterogeneous. This chapter focuses on the longitudinal assessment of a boy with FXS from 12 months through to 6 years of age using two tests: one that assesses psychomotor development and another that assesses neuropsychological maturity. The child had attended an Early Childhood Development Intervention Center-ECDIC since 12 months old. He was administered the *Brunet-Lézine Revised Scale of Psychomotor Development in Early Childhood* until 36 months and underwent CUMANIN neuropsychological testing from age 3 to 6. The results obtained allow us to observe trends in the boy's psychomotor development and neuropsychological maturity over time. Significant commonalities between these results and those of previous cross-sectional studies are discussed. Furthermore, some conclusions are drawn that may prove valuable to professionals and researchers interested in this syndrome.

Chapter 77 - Cornelia de Lange syndrome (CdLS), also known as Bushy syndrome, Amsterdam dwarfism and Brachmann- de Lange syndrome is a genetic multi system disorder, usually caused by spontaneous mutation. Although present from birth, it may not always be immediately diagnosed. The estimated incidence is about 1:10,000-30,000 births. Both sexes are affected. Mortality is high early in life and neurosensory, craniofacial, musculoskeletal, cardiac and gastrointestinal abnormalities are all apparent. There is no known cure and treatment is supportive, requiring a team support system.

Chapter 78 - *Background:* SCAs are the most frequently occurring chromosomal abnormalities with an incidence of 1 in 400 births. As the number of X chromosomes increases, the phenotypic severity increases as well and it is estimated that cognitive abilities decrease by 10–15 IQ points for each additional X chromosome. Aim of this paper is to illustrate clinical variability of cognitive-behavioral phenotype in the different SCAs. *Design and Methods:* The sample was composed by 53 subjects (mean age = 21.16 yrs, range: 13-54) with karyotype 47, XXY (73%), 49, XXXXY (7%), 48, XXYY (9%), mosaicism 47, XXY/48, XXXY (2%), 47, XYY (5%), 48, XXXY (2%), 49, XXXYY (2%). Only 5 subjects have been diagnosed

prenatally (4 KS and 1 XXYY). Primary caregivers completed a comprehensive questionnaire detailing birth, medical, developmental and psychological history. Cognitive and behavioral assessment was performed with clinical interviews using DSM 5 criteria and psychometric questionnaires (WISC-R, WAIS-R, CPM, Token Test, VABS, SCL90, SCQ). Twenty-one sex and age matched subjects karyotypically normal were also evaluated from the behavioural point of view. *Results:* Mean IQ in typical KS was 87.45 ± 2 ds (sd = 20.12) range 45-123, VIQ 91.74 (sd = 19.55) range 50-130 and PIQ 86.87 (sd = 20.87) range 50-126. Mean IQ in other SCAs was 68.71 (sd = 20.81) range 45-106, VIQ 69.36 (sd = 21.97) range 47-113 and PIQ 74.72 (sd = 21.70) range 45-112. In CPM KS subjects scored 27.75 (range 13-36) and 31.50 in the Token Test (range 21-35) while in CPM the other SCAs subjects scored 22.27 (range 10-35) and 22.50 (range 9-31) in the Token Test ($p < .05$). VABS scores documented more marked impairment on adaptive behavior in atypical SCAs subjects. SCL90 documented an elevation of paranoid scale in the 70% of KS subjects and 50% of other SCAs. Autistic traits were present in 67% of the other SCAs subjects and in the 18% of KS at the SCQ. *Conclusion:* A precise identification of the cognitive and behavioral phenotype in different SCAs may enhance the clinical treatment, anticipatory guidance, and care throughout the lifespan.

Chapter 79 - Transcranial direct current stimulation (tDCS) is a non-invasive, painless brain stimulation treatment that uses direct electrical currents of low intensity to stimulate specific parts of the brain. tDCS could both facilitate (anodic stimulation) and inhibit (catodic stimulation) specific areas of the brain, as many neurological and psychiatric disorders are linked to a hypofunction or hyperfunction of specific areas of the nervous system. Such phenomenon is based on two processes: rearrangement of functional neural circuits, and their reconstruction. In light of the studies mentioned above, it is assumed that tDCS can represent a useful tool to facilitate the process of neuroplasticity in subjects affected by chronic neurological diseases and genetic etiopathogenesis, such as Rett Syndrome (RTT). The aim of the present study is to examine the neurophysiological and cognitive effects of cognitive empowerment combined with tDCS in young girls and women with RTT, with chronic language impairments. Despite results in cognitive rehabilitation showing a positive trend, the efficacy of specific intervention on articulated speech is less consolidated. Lack of current research on successful outcomes in language production prompted the current study, which focuses more on intervention of articulated speech and cognitive functions. In this chapter, the author propose an integrated intervention: tDCS and cognitive empowerment applied to language in order to boost speech production (new functional sounds and new words). Given that maximal gains are usually achieved when tDCS is coupled with behavioural training, the author applied tDCS stimulation on Broca's area together with linguistic training. Fourteen young girls and women with RTT were randomly allocated into two subgroups: AtDCS ($n = 7$) or placebo tDCS ($n = 7$). tDCS was applied over Broca's area for a 20-minute session for ten consecutive days. During tDCS stimulation, speech rehabilitation was divided into two sessions: production of vowels and word associations, and discrimination between corresponding words and images. Neurophysiological and cognitive parameters were measured at baseline, post-training and one month after intervention. Results show a general enhancement in language, motor coordination and neurophysiological parameters in the AtDCS group compared to the placebo group. The present study provides evidence that tDCS combined with cognitive empowerment can improve language abilities and motor coordination and foster brain plasticity in young girls and women with RTT. Hence, this study supports the role of tDCS

stimulation as a new methodology in the rehabilitation of diseases with chronic impairment and genetic etiopathogenesis.

Chapter 80 - The present study examined the control of force and timing during finger tapping sequences of adolescents with Down syndrome. Data were obtained from two groups. An experimental group was composed of nine male adolescents with Down syndrome (15–17 years old). Two participants were moderate in intelligence level while other seven were severely intellectually disabled. A comparison group consisted of nine male high school students (16–17 years old). Participants performed both unimanual and bimanual tapping tasks with one self-paced test trial after three audible-synchronized practice trials with concurrent feedback of force output. All tasks consisted of a target force of 2N and a target intertap interval of 500 ms. Adolescents with Down syndrome exhibited a greater magnitude of positive constant error and variable error for peak force than typical adolescents. They also exhibited a greater magnitude of negative constant error and variable error for intertap interval than typical adolescents. Although normally developing comparison adolescents exhibited a linear relationship between peak force and press duration or time-to-peak force, the relationship was not familiar to adolescents with Down's syndrome. This may suggest differences in the manner of motor unit recruitment between the group with Down's syndrome and comparison adolescents. On the other hand, there was no difference between unimanual and bimanual tasks for variable error of intertap interval in adolescents with Down's syndrome. Because people with Down syndrome have exhibited a thinner corpus callosum than typical people, they may be unable to combine the output of two separate timing systems.

Chapter 81 - The author have shown promising results of Assisted Cycling Therapy (ACT) for improving executive functioning in adolescents with Down syndrome (DS). The current study examines the one month retention of executive function benefits gained by adolescents with DS. Fifteen participants were randomly assigned to voluntary cycling (VC; i.e., self-selected cadence) or Assisted Cycling Therapy (ACT; i.e., 65% faster than self-selected cadence accomplished by a motor). Both cycling groups rode a stationary bicycle, for 30 minutes, three times a week, for eight weeks. At the beginning (i.e., pre-test) and end (post-test) of the 8-week session, and at a one month retention (follow-up), three executive functions including set-switching, inhibition, and cognitive planning, were measured. The results showed improved cognitive planning and set-switching for the ACT group after 8 weeks of intervention and these improvements were maintained for one month after the intervention. However, no significant differences were found between the cycling groups for our measure of inhibition. Thus, our results suggest that, especially in regards to cognitive planning and set switching, ACT may lead to relatively permanent changes in the brain.

Chapter 82 - Down syndrome (DS), well known as trisomy 21, occurs in one in one out of 700-1000 live births in all ethnic groups. DS is a very common cause of mental impairment and children with DS present with a variety of medical issues, including cardiovascular defects, endocrine problems, neurodevelopment disorders, hematological problems, gastrointestinal and sleep dysfunctions, visual and hearing impairment. It is well known that there is an association between DS and autoimmune disorders. Thyroid dysfunction (mostly autoimmune hypothyroidism) is the most typical endocrine disorder in individuals with this syndrome (affecting 0 to 66% of these patients). In these patients, in contrast with the general population, there are not substantial differences of incidence between sexes. It can be either congenital or acquired, and usually presents as a subclinical disorder. Hyperthyroidism as well is more commonly seen in people with DS than in the general population, and the phenotypic

metamorphosis from hypothyroidism to hyperthyroidism is more frequently seen. Moreover it is described in the literature an increased frequency of shortness of stature, diabetes mellitus, and nutritional disorders (overweight and obesity) in children with DS. An increased oxidative stress, probably due to the over expression of the gene for Cu/Zn superoxide dismutase (*SOD1*) located in chromosome 21, is observed in DS people and this fact may play a role in the higher prevalence and severity of a number of clinical conditions linked with the syndrome, as well as the accelerated ageing observed in these individuals. An endocrine follow up is required for these individuals, in order to early detect any subclinical abnormalities and prevent as many consequences as possible.

Chapter 83 - Autosomal dominant polycystic kidney disease (ADPKD) is an inherited genetic disorder that results in progressive renal cyst formation and ultimately loss of renal function. Mutation in either *PKD1* or *PKD2*, which are the genes coding for polycystin-1 and polycystin-2, respectively, is the main cause of the disease. The mutation in *PKD1* accounts for 85% of all ADPKD cases, whereas only 15% of ADPKD cases result from *PKD2* mutations. ADPKD is a systemic disorder associated with cardiovascular, portal, pancreatic and gastrointestinal systems. ADPKD is a ciliopathy, a disease associated with abnormal primary cilia. Non-motile primary cilia, functioning as mechanosensory organelles, have been an intense research topic in ADPKD. It has been shown that both structural and functional defects in primary cilia result in cystic kidney and vascular hypertension. In particular, polycystin-1 and polycystin-2 are co-localized to primary cilia and are responsible for mechanosensory induced calcium influx in response to fluid-shear stress. Based on the multiple signaling pathways in ADPKD, different molecular targets have been developed for potential therapies.

Chapter 84 - Hereditary Haemorrhagic Telangiectasia (HHT) or Rendu Osler Weber syndrome (OMIM 187300/ORPHA774) is a vascular hereditary autosomic dominant multiorgan dysplasia. Prevalence is in between 1 to 5,000/8,000 inhabitants around 65,000 in Europe, and 200,000 in USA); although due to founder effect, and insulation, it is higher in some regions as the Jura in France, Funen Island in Denmark and Caribbean Dutch Antilles where the prevalence may be 1 in 1,200 inhabitants. Diagnosis is based on the clinical criteria of Curaçao: epistaxis, telangiectases, first degree relative with HHT, and visceral arteriovenous malformations (AVMs), mainly in lung, liver and brain. For a positive diagnosis, 3 out of the 4 previous criteria are required. A positive genetic test implies also a positive diagnosis. The clinical diagnosis requires then a detailed medical screening, with involvement of different medical specialties. Penetrance of the disease is variable increasing with age. Pulmonary arteriovenous malformations (PAVMs) occur in approximately 50% of patients, hepatic involvement in up to 70%, brain AVMs in 10% and spinal in 1%. However the most frequent clinical manifestation of HHT is epistaxis (nose bleeding) normally from light to moderate that affects 93% of patients and is present before the age of 21 in 90% of cases. The genetic origin of the disease is due to mutations of genes involved in the TGF- β pathway, critical for the normal development of blood vessels. The first gene identified was *Endoglin* (*ENG*), responsible for the 39-59% of the HHT cases (HHT 1); shortly after, *ALK1*/*ACVRL1* was discovered to be involved in 25-57% of cases (HHT2). In around 2% of the HHT patients, the mutation is located in the *MADH4/Smad4* gene leading to a combined syndrome of Juvenile Polyposis and HHT (JPHT). A third and a fourth locus have been mapped on chromosomes 5 and 7 with no genes identified at the moment. Endoglin plays a key role in vasculogenesis and arterial/venous differentiation in embryos, as well as in angiogenesis and neovascularization processes in the adult; ALK1 is responsible for the events occurring during the activation phase

of angiogenesis. Haploinsufficiency is accepted as the mechanism of pathogenicity for the HHT.

Chapter 85 - Osteogenesis imperfecta is the most common heritable cause of fractures in children. It is not a single disorder but a large group of diseases most, but not all, being caused by defects of the genes coding for collagen. Autosomal dominant inheritance is the most common finding in familial cases but new mutations occur. Autosomal recessive inheritance does occur and mosaicism is recognized. The great molecular heterogeneity is reflected in great clinical and radiological variation. Some cases are so severe that survival beyond intra-uterine life is impossible. At the other extreme some patients, undoubtedly affected their family history, have few fractures and live normal lives. Fractures often occur spontaneously and previously asymptomatic fractures in various stages of healing are often found radiologically. While most symptomatic fractures are diaphyseal, all types of fractures including metaphyseal fractures, rib fractures and skull fractures do occur. Modern management involves good orthopaedic surgery; it is particularly important to avoid prolonged immobilisation of limbs to avoid superimposed osteopenia. Drug therapy, particularly with pamidronate, may be appropriate in children with the more severe forms of the disorder. Specific attention may be needed to scoliosis, basilar invagination or deafness. Good occupational therapy to maximise mobility is important. Children with osteogenesis imperfecta have normal intelligence and good education is vital.

Chapter 86 - The sudden rise of new biochemical and molecular techniques, have enabled a better understanding of the physiological and biochemical bases of the tumorigenesis, leaving clear that cancer is a "*genetic condition*". Breast cancer is one of the most common cancers in women, affecting one in six women among 40-59 years in the world; so it has been widely studied. Unlike the majority of genetic diseases, in which the presence of a mutation in a particular gene is sufficient for delineation of a phenotype (monogenic); in breast cancer, the simple presence of a mutation in a particular gene is not enough to explain it. Approximately 90% of breast cancer cases occur sporadically and the majority of cases are caused by mutations in the *BRCA1* or *BRCA2* gene. However, in 5-10% of cases of breast cancer has been identify an autosomal dominant transmission, as well as mutations in specific genes such as *TP53*, *PTEN*, *CHECK2* *STK11* among others, that considerably increase the susceptibility to this condition. Autosomal dominant transmission of this disease has opened a new chapter in cancer medicine since the presence of any of these mutations in a patient, forces the doctor to carry out a deep investigation of the condition in order to establish a prognosis, as well as effective strategies for survival and family prevention. This chapter makes a brief revision of the autosomal dominant disorders Li-Fraumani syndrome, Cowden Disease and Peutz-Jeghers syndrome, which have a high susceptibility to development of breast cancer.

Chapter 87 - Fragile X syndrome (FXS) is the most common cause of familial intellectual disability and the most commonly known single gene causing autism spectrum disorders. The incidence varies according to the populations; it is well-accepted that 1/4,000 males and 1/6,000 females are affected, and 1/250 females and 1/800 males are carriers. The main clinical manifestations are intellectual disability, dimorphic traits and behavior disturbances. The syndrome is inherited as a dominant X-linked trait with reduced penetrance (80% for males and 30% for women). This syndrome is due to a functional loss of the *FMR1* gene product, Fragile X Mental retardation Protein (FMRP), and, in most cases it is caused by a CGG repeat expansion in the *FMR1* promoter. The repeat is from 6 to 55 CGGs long in the normal population while in patients with FXS the repeat number is over 200 CGGs (full mutation

FM). This number of CGGs generally leads to methylation of the repeat and the promoter region, which is accompanied by silencing of the *FMR1* gene. The absence of the FMR1 protein, FMRP, is the cause of the intellectual disability in these patients. Individuals with 55 to 200 CGGs carry a premutation (PM). All affected children have carrier mothers (full mutation [FM] or PM) with a 50% of chance of having another affected child in future pregnancies. Female PM carriers are at risk of developing primary ovarian insufficiency (FXPOI). Elderly PM carriers (males and females) may develop a progressive neurodegenerative disorder called fragile X-associated tremor/ataxia syndrome (FXTAS). The *FMR1* gene is responsible for different disorders depending on the length of the CGG tract and the molecular mechanism. The Fragile X syndrome is due to a loss of function, and FXPOI and FXTAS are due to a toxic gain of function of the mRNA or repeat-associated-non-AUG translation. At present, there is no treatment, although several studies are ongoing in both human and animal models. The present and the following chapters review the current status of the wide spectrum of different pathologies related to the *FMR1* gene.

Chapter 88 - Fragile X-associated primary ovarian insufficiency (FXPOI) is among the family of disorders caused by the expansion of a CGG triplet repeat in the *FMR1* gene. FXPOI is a new clinical entity in which, carrier premutation (PM) females (with 56 to 200 CGG repeats) present early ovarian dysfunction, with menopause occurring 5 years earlier than non-carrier family members. It has been estimated that 2-6% of all women with premature ovarian failure and a normal karyotype have a PM. Therefore, *FMR1* testing is recommended in all women with confirmed ovarian failure; i.e., cessation of the menstrual cycle during three-four months with elevated FSH levels of >30U/L. Despite abundant literature, all authors agree that only a subset of PM carriers develop FXPOI (about 13-26%), with the precise molecular mechanisms of how the *FMR1* premutation causes FXPOI not being well understood. Nonetheless, recent studies have attempted to provide some insight into these mechanisms. This chapter summarizes some of these studies. Among these reports, the most important findings are: 1) the significant positive association of repeat size with POI, demonstrating that women with fewer than 100 repeats have an increased risk of FXPOI, and 2) the hypothesis that the *FMR1* PM may have a toxic RNA gain-of-function effect on ovarian follicle dynamics. These findings have also been demonstrated in rodent models in which the *FMR1* gene protein (FMRP) is highly expressed in oocytes which are important for folliculogenesis. Indeed, the two PM mouse models studied to date have shown evidence of ovarian dysfunction and increased expression of *FMR1* mRNA in the ovary.

Chapter 89 - Fragile X-associated tremor/ataxia syndrome (FXTAS) is a late-onset inherited neuropsychiatric degenerative disorder that occurs predominantly in male carriers of the *FMR1* premutation (55-200 CGG repeats). *FMR1* premutation is relatively frequent in the general population, affecting approximately 1 out of 800 males and 1 out of 250 females, and leading to symptoms of FXTAS in up to 1 in 3,000 men older than 50 years. Clinical symptoms in FXTAS patients usually begin with an action tremor. After that, different findings including ataxia (balance problems with frequent falling), and more variably, loss of sensation in the distal lower extremities and autonomic dysfunction (e.g., impotence, hypertension, and loss of bowel and bladder function), may occur, and gradually progress. Molecular mechanism leading to FXTAS is distinct from the *FMR1* silencing mechanism and/or a deficit in FMRP operating in fragile X syndrome. Individuals with *FMR1* premutation alleles have markedly elevated levels of expanded CGG-repeat *FMR1* mRNA, which is thought to have a toxic gain-of-function. Since 2001, when FXTAS was first described, the advancement of our understanding

of the clinical phenotype as well as the molecular pathophysiology has occurred very quickly. The aim of this chapter is to present the most recent advances in the current knowledge of FXTAS.

Chapter 90 - This chapter describes the psychopathological alterations of the different phenotypes associated with the *FMR1* gene. Fragile X patients present enormous functional impairment with a behavioral phenotype of hyperactivity and impaired attention, marked anxiety, with poor eye contact, affective lability, aggression, self-injurious behavior and autistic features. Since its first description the basic formulation of the fragile X behavioral phenotype has remained intact to the present day, with substantial confirmation of these basic findings in subsequent studies. Autism is one of the most recognized and severe behavioral abnormalities observed in males with fragile X syndrome (FXS) and this syndrome is the leading known monogenic cause of autism, accounting for approximately 5% of autism cases. Approximately 30–50% of FXS individuals meet full Diagnostic and Statistical Manual of Mental Disorders DSM-IV-TR criteria for autism with 60–74% fulfilling criteria for an autism spectrum disorder (ASD). Attention deficit and hyperactivity disorder (ADHD) is the most common diagnosable condition in FXS patients, with most males meeting formal criteria at some point in their lives. In children with FXS, ADHD is reported to be characterized by more inattentiveness, restlessness, fidgetiness and impulsivity. However, affective symptoms can be severe and disruptive, and are a common target for psychopharmacologic intervention. Several clinical subgroups present a higher risk for presenting anxiety outcomes, including children with FXS. Males with FXS display a broad range of anxiety symptoms, but these symptoms often do not fit into the established categories of major anxiety disorders employed by the DSM. Regarding individuals carrying a premutation there is a growing body of literature suggesting that neuropsychiatric features of fragile X-associated tremor/ataxia syndrome (FXTAS) follow a fronto-subcortical pattern with primary impairments in executive function and increased vulnerability to mood and anxiety disorders. Increased rates of psychiatric symptoms may represent early markers of neurodegenerative diseases and have also been reported among *FMR1* premutation carriers without FXTAS. On the other hand, several studies have reported an excess of intermediate *FMR1* alleles in patients with cognitive and/or behavioral phenotypes. Numerous studies have investigated neuropsychological phenotypes among premutation allele carriers in women, but no definitive profile has been achieved. An increased risk of anxiety and mood disorders among premutation allele carriers has not been established, although they seem to be more common in these subjects than in controls.

Chapter 91 - The expansion of the CGG trinucleotide located within the 5'UTR of the *FMR1* gene is involved in a growing number of diseases; the most well-established are Fragile X Syndrome (FXS), Fragile X Tremor/Ataxia Syndrome (FXTAS) and Fragile X Primary Ovarian Insufficiency (FXPOI). Whereas full mutation alleles (>200CGGs) are responsible for the FXS, smaller expansions called premutation alleles (55-200CGGs) are associated with FXTAS and FXPOI. Numerous evidence have currently been reported suggesting that premutation alleles give rise to an increased risk for carriers of these alleles in relation to additional medical, psychiatric and cognitive features which occur at a greater frequency than what would be expected for the general population. In this chapter, the author review the clinical features including peripheral neuropathy, immune-mediated disorders, migraines and neurocognitive involvement which have been suggested to be associated with premutation alleles. In addition, the current understanding of the pathogenic molecular mechanisms that give rise to the spectrum of *FMR1* premutation associated disorders is also reviewed. Although

further research is needed in order to shed light on the factors underlying the common incomplete penetrance applicable to all phenotypes associated with the premutation, it is likely that a combination of environmental and genetic factors with differences in intrinsic susceptibility may modulate the appearance and the severity of these disorders.

Chapter 92 - Fragile X syndrome is the most common form of inherited intellectual disability, with an estimated incidence of 1 in 4,000 males and 1 in 6,000 females. Each diagnosis of an *FMR1* mutation has far reaching clinical and reproductive implications for the extended family. Until now genetic counseling was offered based on the expansion risk in premutation carrier women, but the description of *FMR1*-associated disorders has increased the complexity of the genetic counseling for FXS families, especially *FMR1* premutation carriers. Male individuals carrying full mutated alleles present with intellectual disability while the penetrance is incomplete in females (30-50%). Premutation allele carriers are intellectually unaffected, but several *FMR1* premutation-related disorders have been described. The most prevalent are fragile X-associated primary ovarian insufficiency and fragile X-associated tremor/ataxia syndrome, but behavioral features such as impaired executive function, social deficits or anxiety have also been related to several *FMR1* premutation carriers. Premutation women have 50% of risk of transmitting a premutation/full mutation allele to their offspring, depending on the CGG expansion repeat and the presence of AGG interruptions. On the contrary, premutation men carriers will only transmit the premutation allele to their daughters. Some issues such as risk assessment for intermediate alleles and the clinical prognosis for females with full mutations still remain challenging. Genetic counselors must have an updated and solid understanding of this genetic condition and the *FMR1*-associated disorders in order to cover all the counseling aspects of these disorders.

Chapter 93 - Fragile X Spectrum includes three different clinical conditions Fragile X Syndrome (FXS), Fragile X-Associated Tremor and Ataxia (FXTAS), and Fragile X-Associated Premature Ovarian Insufficiency (FXPOI). Treatment of FXS is mainly symptomatic and it is addressed to improve or make disappear some of the more dyscapacitating symptoms like hyperactivity, deficit attention disorder and behavioral problems of language anomalies. Clinical trials are in course focusing on new discovered therapeutical targets (i.e., mGluR). Treatment of FXTAS and FXPOI are also mainly symptomatic and should be individually prescribed and modified depending on the clinical evolution.

Chapter 94 - Over the course of less than a decade, whole genome sequencing has progressed from being one of our nation's boldest scientific aspirations to becoming a readily available technique for determining the complete sequence of an individual's deoxyribonucleic acid (DNA)—that person's unique genetic blueprint. With this tremendous advance comes the accumulation of vast quantities of whole genome sequence data and complex questions of how—across a multitude of clinical, research, and social environments—to protect the privacy of those whose genomes have been sequenced. Collections of whole genome sequence data have already been key to important medical breakthroughs, and they hold enormous promise to advance clinical care and general health moving forward. To realize this promise of great public good ethically, individual interests in privacy must be respected and secured. Large-scale collections of genomic data raise serious concerns for the individuals participating. One of the greatest of these concerns centers around privacy: whether and how personal, sensitive, or intimate knowledge and use of that knowledge about an individual can be limited or restricted (by means that include guarantees of confidentiality, anonymity, or secure data

protection). Because whole genome sequence data provide important insights into the medical and related life prospects of individuals as well as their relatives—who most likely did not consent to the sequencing procedure—these privacy concerns extend beyond those of the individual participating in whole genome sequencing. These concerns are compounded by the fact that whole genome sequence data gathered now may well reveal important information, entirely unanticipated and unplanned for, only after years of scientific progress. Another privacy concern associated with whole genome sequencing is the potential for unauthorized access to and misuse of information. For example, in many states someone could legally pick up a discarded coffee cup and send a saliva sample to a commercial sequencing entity in an attempt to discover an individual's predisposition to neurodegenerative disease. The information might then be misused, for example, by a contentious spouse as evidence of unfitness to parent in a custody case. Or, the information might be publicized by a malicious stranger or acquaintance without the individual's knowledge or consent in a social networking space, which could adversely affect that individual's chance of finding a spouse, achieving standing in a community, or pursuing a desired career path. Realizing the promise of whole genome sequencing requires widespread public participation and individual willingness to share genomic data and relevant medical information. This, in turn, requires public trust that any whole genome sequence data shared by individuals with clinicians and researchers will be adequately protected. Current U.S. governance and oversight of genetic and genomic data, however, do not fully protect individuals from the risks associated with sharing their whole genome sequence data and information. In particular, a great degree of variation exists in what protections states afford to their citizens regarding the collection and use of genetic data. Only about half of the states, for example, offer protections against surreptitious commercial genetic testing. Currently, the majority of the benefits anticipated from whole genome sequencing research will accrue to society, while associated risks fall to the individuals sharing their data. This report focuses on reconciling the enormous public benefits anticipated from whole genome sequencing research with the potential risks to privacy of individuals, and the protections that must be foremost in our minds as the author focus our policies to facilitate such privacy and progress.

Chapter 95 - Congress has considered, at various points in time, numerous pieces of legislation that relate to genetic and genomic technology and testing. These include bills addressing genetic discrimination in health insurance and employment; personalized medicine; the patenting of genetic material; and the quality of clinical laboratory tests, including genetic tests. The focus on these issues signals the growing importance of the public policy issues surrounding the clinical and public health implications of new genetic technology. As genetic technologies proliferate and are increasingly used to guide clinical treatment, these public policy issues are likely to continue to garner considerable attention. Understanding the basic scientific concepts underlying genetics and genetic testing may help facilitate the development of more effective public policy in this area. Most diseases have a genetic component. Some diseases, such as Huntington's Disease, are caused by a specific gene. Other diseases, such as heart disease and cancer, are caused by a complex combination of genetic and environmental factors. For this reason, the public health burden of genetic disease is substantial, as is its clinical significance. Experts note that society has recently entered a transition period in which specific genetic knowledge is becoming critical to the delivery of effective health care for everyone. Therefore, the value of and role for genetic testing in clinical medicine is likely to increase significantly in the future.

Chapter 96 - Cancer is a major health problem worldwide and an early effective diagnosis would ensure timely management, impacting on the quality of life, the patients overall well being and longevity. There is no single test that can accurately diagnose cancer. An effective diagnostic test should be able to confirm or eliminate the presence of disease, monitor the disease process, and plan for and evaluate the effectiveness of treatment. Though there is a wide array of methods to diagnose cancer, an effective biomarker is still an unmet medical need. Conventional tests use single genes or discrete pathways. Cancer is a consequence of multiple changes occurring both in parallel and in sequence in the macromolecules of the cell. Earlier only gross end points were possible to be identified and measured, but today with technological advances it has become possible to interrogate events that are very early and at scales hitherto irresolvable. Thus changes in these cellular events are studied at different levels. DNA changes in cancer are expressed in terms of gross chromosomal anomalies or could be in the sequence of the gene or gene product/protein. As researchers learn more about the mechanisms of cancer, new diagnostic tools are constantly being developed and existing methods refined. Diagnostic procedures for cancer may include imaging, laboratory tests (including tests for tumor markers), tumor biopsy, endoscopic examination, surgery or genetic testing. This chapter draws attention to the critical events in pathobiology of cancer as is resolved by today's technological progress while compiling major biomarkers enabling improved diagnosis and reviewing some putative biomarkers for translation from bench to bedside.

Chapter 97 - The advent of massively parallel sequencing has changed the interrogation process of the human genome and now provides a high resolution and global view of the genome which is beyond research applications. Together with powerful bioinformatics tools, these next generation sequencing technologies have revolutionized fundamental research and have important consequences for clinically actionable tests, diagnosis and treatment of rare diseases and cancers. Today, molecular testing is commonly used to confirm clinical diagnosis of specific diseases; it requires that a clinician specify the gene or mutation to test and, in return, will receive information only about this sequence. Despite relative successes, a large number of patients receive no accurate diagnosis, even after many expensive molecular investigations. A clear paradigm shift has taken place in the health network with the introduction of the exome sequencing in molecular diagnostic lab. In this chapter, the impact of the implementation of high throughput sequencing technologies on molecular diagnosis and on the practice of medicine, with an emphasis in paediatrics, is reviewed. The author compared well-established genetic tests, using examples from our molecular diagnostic lab, to the recent exome sequencing applications. The genetic tests can fall into three main categories: 1) Mendelian Single Gene Disorder tests that include targeted mutation and targeted gene approaches 2) Genetic Disease Panels which are composed of a few to a dozen genes and 3) Exome or Genome approaches, which interrogate either the entire coding sequences of the 22,333 human genes or the entire human genome. For each of these categories, advantages and limitations are discussed. The author devoted a section on the future of molecular diagnosis and discuss which tests will subsist and which one may be soon abandoned. Massively parallel sequencing is transforming the molecular diagnostic field: it offers personalized genetic tests and generates new ethical challenges. Important questions like incidental findings and possible forms of discrimination are addressed. Finally, the author conclude with a section on the future directions surrounding the application of these multimodal molecular approaches in general and their putative applications in neonatal intensive care units.

Chapter 98 - Viral vectors engineered to carry transgenic sequences can be delivered into discrete tissues or anatomical structures to express specific transgenes into the transduced cells. Therefore, they are useful tools to produce specific, transient and localized knockout, knockdown, ectopic expression or overexpression of a gene, leading to the possibility of analyzing both *in vitro* and *in vivo* molecular basis of relevant functions. Replication-incompetent helper-dependent amplicon vectors, derived from herpes simplex virus type-1 (HSV-1) are devoid of viral genes. Thus, these vectors have great advantages as tools for *in vitro* and *in vivo* gene transfer and, in particular (i) minimal toxicity or induction of adaptive immune responses, and (ii) large transgene capacity, being able to carry up to 150 kbp of foreign DNA. In addition, these vectors have (iii) widespread cellular tropism: amplicons can experimentally infect several cell types, either quiescent or not, though naturally HSV-1 infects mainly neurons and epithelial cells, and (iv) absence of insertional mutagenesis, since the viral genome does not integrate into the host cell genome. These vectors have been used both on basic and applied research, and they have revealed as most suitable tools to study complex functions involving the nervous system, such as anxiety, sexual behavior, learning and memory. In addition, amplicon vectors are being used for the development of new experimental gene therapy approaches, both for inherited and acquired diseases affecting the nervous system, including neurodegenerative diseases. Although several technological improvements have been achieved in the last decade, some difficulties regarding these appealing vectors remain still unresolved, such as the inability to generate large amounts of high-titer fully helper-free vectors and the fact that expression from the transgenic sequence delivered by the vectors is generally unstable, often leading to a complete silencing of expression after a few weeks. To overcome these obstacles and to improve these vectors, the author have recently modified (a) the amplicon genome, in order to fully delete bacterial sequences and (b) developed novel complementing cell lines, in order to improve helper-free vector production and to render amplicon stocks compatible with clinical trials. In this review article the author briefly review data supporting the potential of HSV-1-based amplicon vector model for gene delivery in primary cultures of neural cells and into the brain of living animals.

Chapter 99 - Papillomaviruses (PVs) infect the epithelium of amniotes, where they can cause tumours or persist asymptotically. PVs are classified in the *Papillomaviridae* family, that contains 29 genera of PVs isolated from humans (120 types), non-human mammals, birds and reptiles (69 types). PVs have circular double-stranded DNA genomes approximately 8 kb in size and typically contain eight genes. Studies aiming the identification of PVs genomes use techniques such as PCR with consensus primers, rolling circle amplification and metagenomic methods. Advances in papillomaviral genome research have allowed the knowledge of PVs diversity and evolution of the poorly known PVs genera types, revealing that there is still a limited understanding of PVs diversity. Particularly, recent studies in Bovine Papillomavirus (BPV) have shown the identification of novel BPV types and several putative new virus types in cattle. This chapter will show new contributions in PVs genome studies.

Chapter 100 - Viral particles are important tools in Molecular Biology, acting as carriers of genetic material, immunogenic antigens, adjuvants, or even directly combating antibiotic-resistant microorganisms and their biofilms in hospital and industrial environments. However, the efficient use of these particles requires extensive knowledge about their characteristics and components, including those involved in their regulatory mechanisms of genome transcription and protein synthesis. The exploration of this knowledge becomes a challenge especially for scientists analyzing the virions in their natural environment, due to their interactions with the

complex and diverse types of biological systems, which directly influence the regulation of the infective cycles. Thus, knowledge about viral genes, their function, organization, and modulation, beyond the comprehension of the viral components as parts of complex systems, consist of the main hurdles for the controlled and predictable handling and use of these particles. For this, the technology of Synthetic Synthesis of viral genomes is distinguished from traditional genetic engineering through the use of modularity and standardization to construct proof-of-principle systems and allow generalized circuits designs to be applied to different scenarios. This new technology is made possible thanks to advances in many areas of science, from the use of restriction enzymes until the development of techniques of genomic synthesis and sequencing, like the 454 Roche, Illumina, and SOLiD systems. This technology is becoming increasingly a multidisciplinary tool used in the investigations about these complex systems, as well in the engineering of new particles for the optimization of diverse viral functions and alterations in their infectivity and affinity, or even in the development of completely new organisms and features without the need for a template. Nevertheless, advances in this technology are still limited by the lack of dynamic techniques for monitoring biological systems and efficient and standardized circuits. Here, the author summarize the major characteristics of viral genomes, their organization and gene modulation, and highlight the main aspects of Synthetic Biology applied to viral genomes, as its main techniques and applications.

Chapter 101 - With increase of microbial and viral genome sequence data obtained from high-throughput DNA sequencers, novel tools are needed for comprehensive analyses of the big sequences data. An unsupervised neural network algorithm, Self-Organizing Map (SOM), is an effective tool for clustering and visualizing high-dimensional complex data on a single map. The author previously modified the conventional SOM for genome informatics on the basis of batch-learning SOM (BLSOM), by making the learning process and resulting map independent of the order of data input. Influenza virus is one of zoonotic viruses and shows clear host tropism. Important issues for bioinformatics studies of influenza viruses are prediction of genomic sequence changes in the near future and surveillance of potentially hazardous strains. To characterize sequence changes of influenza virus genomes after invasion into humans from other animal hosts and to study molecular evolutionary processes of their host adaptation, the author have constructed BLSOMs for oligonucleotide, codon, amino-acid, and peptide compositions in all genome sequences of influenza A and B viruses and found clear host-dependent clustering (self-organization) of the sequences. Viruses isolated from humans and birds differ in mononucleotide composition from each other. In addition, host-dependent oligonucleotide and peptide compositions that cannot be explained with the host-dependent mononucleotide composition are revealed by these BLSOMs. Retrospective time-dependent directional changes of oligonucleotide compositions, which are visualized for human strains on BLSOMs, can provide predictive information about sequence changes of the newly invaded viruses from other animal sources. Basing on this host-dependent oligonucleotide composition, the author have proposed a strategy for prediction of directional changes of virus sequences and for surveillance of potentially hazardous strains when introduced into human populations from nonhuman sources. Millions of genomic sequences from infectious microbes and viruses will become available in the near future because of their medical importance, and BLSOM can characterize such big data easily and support efficient knowledge discovery.

Chapter 102 - An oncogene is a modified gene, or a series of nucleotides that encode a protein, and direct the cell to the development of a neoplastic phenotype. Usually, oncogenes are involved in tumor development and increase the possibility that the development

(proliferation and differentiation) of a cell directs towards cancer. New researches indicate that small ribonucleic acids (RNA) of 21-25 nucleotides, called micro RNA (miRNA), can control these genes through down-regulation. The first oncogene was discovered in 1970 and named Src. Src was first discovered in a retrovirus of chickens. In 1976, J. Michael Bishop and Harold E. Varmus of the University of California showed that this oncogene was a defective proto-oncogene present in many organisms including humans. For their studies Bishop and Varmus were awarded of the Nobel Prize in 1989. A proto-oncogene is a normal gene that can become oncogenic due to mutations or to increased expression. Proto-oncogenes encode proteins that regulate the cell cycle and differentiation. They may also be involved in the signal transduction of the start of mitosis. A proto-oncogene becomes an oncogene even with minimum modifications of its original functions. There are two basic types of activation: 1) a mutation of nucleotides which produce a different protein causing: a) increased enzymatic activity of the protein; b) the loss of regulatory sites; and c) the creation of hybrid proteins; and 2) an increase of the concentration of proteins caused by: a) an increase of gene expression (through misregulation); b) an increase of stability (half-life) of the protein; and c) a duplication or amplification of the gene coding for the protein. Growth factors (GF), or fitogens, are usually secreted in a few cells specialized to induce proliferation in a paracrine, autocrine, or endocrine manner. If a cell that normally does not produce GF suddenly begins to produce them (because it has developed an oncogene), this induces the proliferation, without control, to adjacent cells (paracrine action) and to its cell type (autocrine action) increasing secretion. There are six known classes of protein kinases (PK) and related proteins that are potential oncogenes: 1) tyrosine kinase receptor (TKR), which becomes constitutively (permanently) activated, such as the epidermal growth factor receptor (EGFR), the platelet-derived growth factor receptor (PDGFR), and the vascular endothelial growth factor receptor (VEGFR); 2) cytoplasmic TK, such as TK enzymes of the Src family, Syk-ZAP-70 and Bruton's TK (BTK); 3) regulatory guanosine triphosphate (GTP)ase, like Ras; mutations that activate permanently Ras are found in 20-25% of all human tumors and up to 90% in some types of cancer, such as pancreatic ones; 4) cytoplasmic serine/threonine kinase and its regulatory units, such as Raf kinase and cyclin-dependent kinase (CDK); 5) adapter proteins in signal transduction (for example in the apoptotic pathway); and 6) transcription factors, for example the Myc gene. A large number of genes have been identified as proto-oncogenes. Most of them are responsible for the production of a positive signal that induces cell division. Other proto-oncogenes play an important role in the regulation of cell death. As mentioned before, the "altered" versions of these genes (oncogenes) may induce the cell to replicate unruly. Such a development can take place even in the absence of normal pro-growth signals, as for example the one provided by GF. A key feature in the activity of oncogenes is that a single altered copy is sufficient to induce an unregulated growth. Such behavior is in contrast with the one typical of tumor-suppressor genes (TSG) for which it is necessary that both copies of the gene are defective for triggering a process of abnormal cell division. Proto-oncogenes that have been identified until now have very different functions within the cell. Despite the differences in their normal functions, these genes all have the characteristic to contribute to unregulated cell division when present as mutated (oncogenic). The mutated proteins sometimes retain some features of their own, but are no longer sensitive to the regulation systems that determine and control the normal form of the protein itself. In this chapter the author overview the role of oncogenes in gynecological pathology.

Chapter 103 - Cancer is an uncontrolled cell growth caused by accumulation of genetic and epigenetic mutations in genes that normally play a role in the regulation of cell proliferation, survival, apoptosis and cell cycle. *Mutations are occurring mainly* in oncogenes, tumor-suppressor genes, microRNA genes, or as DNA repair defects and aberrant DNA methylation. Oncogenes encode proteins that control cell proliferation and/or apoptosis. Products of oncogenes can be classified into 6 groups by its biological activity: transcription factors, growth factors, growth factor receptors, signal transducers, chromatin remodeling and apoptotic regulators. The main changes related to oncogene activation are chromosomal translocations and mutations that can occur as early events or during tumor progression; whereas amplification usually occurs during the tumor progression. These alterations are usually somatic events, although germ-line mutations can also predispose to familial cancer. A single genetic change is rarely insufficient for developing a malignant tumor. Most evidences point to a multistep process of sequential alterations in a wide number of oncogenes, tumor-suppressor genes, or microRNA genes related with cancer. Recently, other mechanisms as the inflammation or alterations in the cellular metabolism have been associated with cancer. This chapter describes some of the key molecular mechanisms involved in the development and progression of cancer.

Chapter 104 - Recent understanding of oncogene regulation has uncovered an emerging new field of molecular cancer biology in which tumor suppressors and mediators are controlled through RNA regulation. Among the most critical players are microRNA (miRNA), also referred to as ‘oncomirs’, and Dicer, the major enzyme responsible for cleaving double-stranded RNA and forming the RNA induced silencing complex. These components are aberrantly expressed in cancer and among some tumor types are hypothesized to be causative to the etiology of malignancy. For example, prostate and colorectal cancers express abundantly high levels of Dicer mRNA while lung, ovarian and endometrial cancers express low levels of Dicer, which is believed to correlate to poor cancer prognosis. In addition, mutations of the Dicer encoding gene (*DICER1*) occur in non-epithelial ovarian cancers and pediatric tumor pleuropulmonary blastoma. Paradoxically, Dicer is a haploinsufficient tumor suppressor; the loss of a single allele of *DICER* enhances tumor growth where the loss of the second allele results in halting tumor proliferation. MicroRNA regulates oncogenic signaling pathways as well as the expression of tumor suppressors and oncogenes, making it a major contributor to the overall status of the cell and its malignant potential. The let-7 miRNA family members are well-known as tumor suppressor genes, which target and silence the Ras oncogene. On the other hand, miRNAs may induce tumor growth; one example is miR-17-19 cluster negatively regulating two tumor suppressor genes, PTEN and Bim. In this chapter, the regulation of oncomirs will be discussed with focus on the post-transcriptional control by miRNA biogenesis machinery, which consist Dicer as a major player.

Chapter 105 - Glioblastoma multiforme (GB) is the most common primary brain tumor among adults. Rapid tumor progression and diffuse invasion of brain tissue results in a poor prognosis despite advances in the understanding of the tumor’s molecular biology. Recently, targeted therapy has been introduced as a potential therapeutic option in several types of cancer. Dysregulation of the epidermal growth factor receptor (EGFR) has been identified in a number of different malignant tumor entities, such as GB. Despite promising reports from preclinical studies, clinical trials yielded no significant improvement of outcomes of patients with GB who were treated with anti-EGFR-targeted agents. Molecular mechanisms underlying resistance to this treatment approach have become a focus of scientific efforts in the last years. The variations and complexity of EGFR signaling pathways demand a composite therapeutic strategy to

quench alternative signaling routes which otherwise might enable cancer cell survival and subsequently tumor progression.

Chapter 106 - Glioblastoma multiforme (GB) is the most common primary brain tumor among adults. Rapid tumor progression and diffuse invasion of brain tissue results in a poor prognosis despite advances in the understanding of the tumor's molecular biology. Recently, targeted therapy has been introduced as a potential therapeutic option in several types of cancer. Dysregulation of the epidermal growth factor receptor (EGFR) has been identified in a number of different malignant tumor entities, such as GB. Despite promising reports from preclinical studies, clinical trials yielded no significant improvement of outcomes of patients with GB who were treated with anti-EGFR-targeted agents. Molecular mechanisms underlying resistance to this treatment approach have become a focus of scientific efforts in the last years. The variations and complexity of EGFR signaling pathways demand a composite therapeutic strategy to quench alternative signaling routes which otherwise might enable cancer cell survival and subsequently tumor progression.

Chapter 107 - Glioblastoma multiforme (GB) is the most common primary brain tumor among adults. Rapid tumor progression and diffuse invasion of brain tissue results in a poor prognosis despite advances in the understanding of the tumor's molecular biology. Recently, targeted therapy has been introduced as a potential therapeutic option in several types of cancer. Dysregulation of the epidermal growth factor receptor (EGFR) has been identified in a number of different malignant tumor entities, such as GB. Despite promising reports from preclinical studies, clinical trials yielded no significant improvement of outcomes of patients with GB who were treated with anti-EGFR-targeted agents. Molecular mechanisms underlying resistance to this treatment approach have become a focus of scientific efforts in the last years. The variations and complexity of EGFR signaling pathways demand a composite therapeutic strategy to quench alternative signaling routes which otherwise might enable cancer cell survival and subsequently tumor progression.

Chapter 108 - This chapter describes the radiographic appearance and distribution of plexiform neurofibromas (PNs) in NF1. Although PNs are histologically benign tumors, they are the source of significant morbidity including disfigurement, pain, functional impairment and they can undergo malignant transformation. Identifying targeted agents that slow the growth or decrease the size of PNs is an area of intense research. Volumetric MRI analysis of PNs is helpful in detecting treatment response and progression in these large and complex shaped tumors with greater sensitivity compared to 1-dimensional or 2-dimensional measurements.

Chapter 109 - Neurofibromatosis Type 1 (NF1) is an autosomal dominant systemic disease. Up to fifty percent of patients with NF1 are reported to have concomitant vascular abnormalities, incidence which increases the mortality especially in young patients. Among the complications of giant neurofibromas, spontaneous rupture and massive hemorrhage has been reported, leading to limb loss and death as well. The resection of larger neurofibromas is a challenging procedure because the risk of uncontrolled hemorrhage is much higher. In this chapter, the author present a review of the medical literature of the methods that have been used to control intraoperative bleeding in large neurofibromas. The author also discuss a novel surgical technique, which includes the ligation of the base of the giant neurofibroma tissue using a continuous loop-shaped suture. This method makes the operation field relatively bloodless and facilitates identification and ligation of the intralesional vessels. In addition, this surgical technique is less complicated and easier to perform and compared to others.

Chapter 110 - Patients with Neurofibromatosis 1 (NF 1) suffer from a myriad of medical problems including benign as well as malignant tumors. Malignancy is the most common cause of death in affected individuals, and reduces life expectancy by 10-15 years. Amongst other tumors, patients with NF1 carry an elevated risk of specific central nervous system (CNS) tumors. The most common NF1-associated CNS tumors are low-grade gliomas (LGG) such as optic pathway gliomas, hypothalamic gliomas, and other parenchymal gliomas located in the brainstem, cerebellar peduncles, globus pallidus, and midbrain. In addition to CNS tumors, patients with NF1 also develop peripheral nervous system (PNS) tumors such as neurofibromas and schwannomas. These tumors arise most commonly from major peripheral nerves such as the radial and ulnar nerve. These are benign tumors however can cause significant disfigurement, pain and depending on the location specific neurological complications. Malignant transformation of these tumors leads to malignant peripheral nerve sheath tumors (MPNST). These tumors occur in less than 5% of children with NF1, but are the leading cause of mortality in adult patients with NF1. Significant advances in our understanding of LGG have been achieved over the last several years. Several genetic aberrations, which activate the MAPK pathway have been identified in the majority of LGGs, most notably the BRAF-KIA1549 fusion protein and the activating BRAFV600E mutation. Specific inhibitors of MEK1/2, the immediate downstream target of BRAF, are currently being tested in the clinic for children with LGGs. Treatment for symptomatic plexiform neurofibromas remains a clinical challenge for most pediatric neuro-oncologists. Recent studies have shown promise to use pegylated (PEG)-interferon-alpha-2b. Other treatment strategies also target the MAPK pathway for these tumors. Treatment of MPNST remains under investigation and outcome remains poor. Most patients are currently treated with a combination of surgery, radiation and chemotherapy. In this chapter, the author will outline NF1 associated CNS and PNS tumors and discuss current treatment approaches.

Chapter 111 - Optic pathway gliomas (OPG) occur in 15% of children with Neurofibromatosis type 1 (NF1) and will lead to visual deficits in up to half. Because this tumor rarely threatens life but often leads to uncorrectable and permanent vision loss, preserving vision is the primary objective in management of OPGs. However, measuring visual function with ophthalmology exams can be challenging in young children with NF1-associated OPG. Recently, a variety of surrogate outcome measures have been investigated to determine if reliable quantifiable markers of vision can be found. In this chapter, the author review the presentation and management of OPGs in children with NF1. The author discuss endpoints commonly used in clinical trials, including assessments of visual function and radiologic progression. Finally, the author investigate and compare recent putative surrogates of vision in OPG, and the author discuss their utility in identifying and following vision loss in children with NF1. Currently, visual acuity is the most reliable, comparable and quantifiable outcome measure available in the assessment of patients with OPG. However, novel physiologic and radiologic markers of vision loss may offer an important adjunct to the assessment of the child with NF1-associated OPG in the future.

Chapter 112 - Optic pathway gliomas (OPG) occur in 15% of children with Neurofibromatosis type 1 (NF1) and will lead to visual deficits in up to half. Because this tumor rarely threatens life but often leads to uncorrectable and permanent vision loss, preserving vision is the primary objective in management of OPGs. However, measuring visual function with ophthalmology exams can be challenging in young children with NF1-associated OPG. Recently, a variety of surrogate outcome measures have been investigated to determine if

reliable quantifiable markers of vision can be found. In this chapter, the author review the presentation and management of OPGs in children with NF1. The author discuss endpoints commonly used in clinical trials, including assessments of visual function and radiologic progression. Finally, the author investigate and compare recent putative surrogates of vision in OPG, and the author discuss their utility in identifying and following vision loss in children with NF1. Currently, visual acuity is the most reliable, comparable and quantifiable outcome measure available in the assessment of patients with OPG. However, novel physiologic and radiologic markers of vision loss may offer an important adjunct to the assessment of the child with NF1-associated OPG in the future.

Chapter 113 - This chapter provides an overview of the neurocognitive, behavioral and developmental aspects of the neurofibromatosis type 1 (NF1) phenotype. Investigations into the pathogenesis of cognitive dysfunction are also presented, with a focus on human neuroimaging studies and mouse models. These studies have provided major insights into the molecular and biochemical abnormalities associated with NF1 that impact on cognitive performance. Preclinical studies suggest that pharmacological correction of these abnormalities has the potential to normalize aspects of the NF1 cognitive phenotype. These are reviewed along with the rationale for ongoing and future human clinical trials.

Chapter 114 - The purpose of this chapter is to provide an overview of communication disorders observed among children and adults with Neurofibromatosis type 1 (NF1). In each section, terminology, prevalence and information regarding assessment for individuals in the general population will be presented. The literature describing communication disorders observed in the population of individuals with NF1 will be reviewed. Following a review of the literature, options for speech-language assessments for individuals with NF1 will be discussed.

Chapter 115 - Neurofibromatosis Type 1 (NF1) is a multisystem disorder, and cognitive impairment is its most common complication in childhood. Although patients with NF1 are at risk for significant clinical illnesses, most patients are only mildly affected and live healthy and productive lives. Most research in NF1 to date has been focused on defining the physical and cognitive deficits based on standardized testing, with little emphasis on their functional correlates. Evidently, student engagement in the classroom reflects, in a large part, self regulation skills which are influenced by temperament and higher order cognitive executive function processes. In particular, sustained attention, which comprises both task-oriented attention and low impulsivity, predicts academic and behavioral adjustment. The purpose of the present chapter is to identify predictors associated with school participation in a sample of children with NF1. A unique feature of this work is the examination of children's participation across different classroom activity components such as handwriting and cognitive skills, such as attention and executive function. Each of these classroom activities may require a unique set of functional skills to meet specific demands. Findings and insights presented in this chapter may help clarify the important factors supporting effective engagement by children with NF1 in school programs. The chapter describes the functional problems of children with NF1 in the classroom according to the International Classification of Functioning, Disability and Health–Children and Youth version (ICF-CY). The chapter includes current literature on NF1, using the framework provided by the ICF-CY in highlighting the holistic evaluation and interpretation of disabilities. In the first part it presents the ICF-CY conceptual model, defines the term 'ecological validity ' and details regarding two ecologically valid assessment tools (Virtual Classroom and functional questionnaire). The chapter subsequently describes the functional academic profile of children with NF1 from three points of view: attention profile,

academic skills in relation to executive profile and handwriting performance. The set of tasks composing each activity provides a useful indication of the factors that may underlie functional participation of children with NF1 in the classroom context. The chapter illustrates the functional profile with a case study and the author conclude the chapter with suggestions for further research.

Chapter 1

SPECIATION IN DIATOMS: PATTERNS, MECHANISMS, AND ENVIRONMENTAL CHANGE

Joshua T. Cooper^{1,*} and John P. Masly^{2,†}

¹Department of Microbiology and Plant Biology,

²Department of Biology, University of Oklahoma, Norman, OK, US

ABSTRACT

Diatoms represent one of the most speciose groups of organisms within the microeukaryota— they occupy almost every aquatic niche worldwide, provide important functions within ecosystems as a major contributor of carbon dioxide fixation from the atmosphere, and also serve as a major contributor to the base of the aquatic food web. Despite this striking example of biodiversity and their ecological importance, we know relatively little about the speciation processes at work in this group of organisms. In this chapter, we review the patterns of speciation in marine and freshwater diatoms with respect to their paleontological, environmental, and genetic evolutionary histories. We summarize the evidence for sex and hybridization among diatom species, and discuss recent work on understanding the mechanisms of reproductive isolation that limit gene flow in the microeukaryota. We also discuss aspects of diatom genome evolution that provide clues to the genetic basis of speciation in this group. Finally, we end with a discussion of a natural model system that presents a unique opportunity to address some of the outstanding questions in diatom speciation.

INTRODUCTION

Speciation and adaptation are two of the most important evolutionary processes that generate biodiversity. Although each of these areas of research has enjoyed a rich history of discovery, the vast majority of what we know about the mechanisms that govern these

* Corresponding Author's Email: jtcooper@ou.edu (Joshua T. Cooper, Department of Microbiology and Plant Biology, University of Oklahoma, 770 Van Vleet Oval, Norman, OK 73019).

† Corresponding Author's Email: masly@ou.edu (John P. Masly, Department of Biology, University of Oklahoma, 730 Van Vleet Oval, Norman, OK 73019).

evolutionary forces is limited largely to multicellular eukaryotic taxa. We know comparatively little about the speciation processes in the microeukaryotes, despite microeukaryota being one of the most speciose groups of organisms on the planet and their ecologically important role within the global carbon cycle. Among the microeukaryotes, diatoms represent the most speciose group, and are particularly important for global ecology and maintenance of the oceanic biogeochemical cycling. The diatoms fix an estimated 40–45% of oceanic production and fix approximately 20 Pg of carbon yearly (Mann 1999), which makes them more productive than all the Earth's rain forests combined. In relative terms, every fifth breath you take contains oxygen atoms derived from diatoms (Nelson et al. 1995; Field et al. 1998; Armbrust 2009).

The diatoms also form the foundation of almost every aquatic food web spanning marine to freshwaters. They are found wherever light and water are present, including saline lakes, freshwater springs, cave entrances, waterfalls, hot springs, and they also inhabit micro-aquatic environments including damp soils and even desert soils (which experience periodic moisture). Since their first appearance in the fossil record during the Upper Cretaceous, approximately 500 distinct genera of diatoms have evolved and been formally described; of these, at least 365 are extant with living representatives, and 132 are presumed extinct with representatives found only in the fossil record (Gersonde and Harwood 1990; Round et al. 1990; Tapia and Harwood 2002; Guiry and Guiry 2012; Silva 2012). However, some species thought to be extinct have been re-discovered recently: occasionally "living fossils" are found living in places like New Zealand tarn lakes (Vyverman et al. 1998), Lake Baikal (Edlund et al. 2000; Williams and Reid 2006), found under the ice or attached to seaweeds in Antarctica (Sutherland 2008), or in geographically isolated ancient desert valleys of Northern Mexico (Winsborough et al. 2009). Current estimates of species richness of diatoms vary greatly, but likely reside within the range of 10^4 to as many as 10^8 species (Round et al. 1990; Mann and Droop 1996).

Diatoms are a diploid group within the Stramenopiles (synonyms include Chromalveolates, Chromista, Heterokonts) and are characterized morphologically by highly ornamented and sculpted bipartite silicon dioxide (silica) shells called the frustule (two valves plus the girdle bands comprise the frustule; Figure 1). The biogenic silica ($\text{SiO}_2 \cdot n\text{H}_2\text{O}$) that comprises the majority of the frustule material is similar in chemical properties to the gemstone opal, and is sequestered from the aquatic environment by specialized silica depositing vesicles (Li et al. 1989; Vrieling et al. 1999; Martin-Jézéquel et al. 2000). The frustule is constructed much like that of a hat box, with one large "box lid", the epitheca valve, that sits on top of and overlaps a smaller "box", the hypotheca valve (Figure 1). Located between the epitheca and hypotheca are several separate, concentric bands of silica called the girdle or cingulum. Both the epitheca and hypotheca possess a cingulum; the cingula act together to allow the cell to expand in height with the addition of each girdle band.

Although the vast diversity of diatoms has drawn the interest of evolutionary biologists since the mid-1800's, little work has been performed investigating the speciation processes and mechanisms at work in this group until relatively recently. Here, we review the progress made on understanding the process of speciation in diatoms and the evolution of reproductive isolation (RI) in relation to adaptation to the environment, and discuss prospects for future research directions in the microeukaryota.

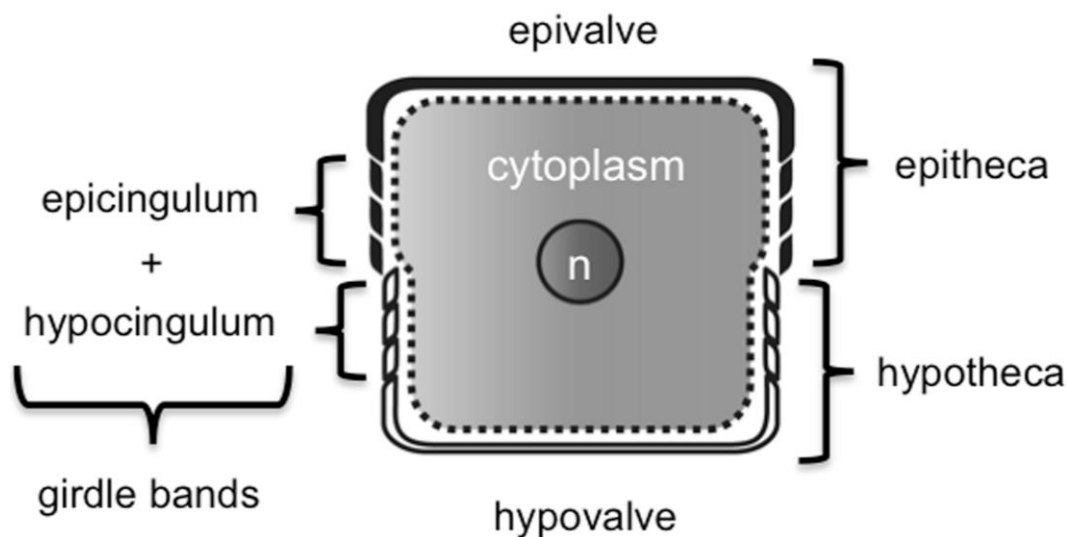


Figure 1. Major morphological features of the diatom frustule. The top of the epitheca (the epivalve plus the epicingulum) is called the valve face. The epicingulum and hypocingulum are composed of several copulae (girdle bands) that allow the cell to expand or contract vertically. The nucleus (n) is often located in the center of a diatom cell, however, it can be moved to different parts of the cytoplasm via microtubules. The cytoplasmic region is bounded by the plasmalemma and silicalemma (dashed line), which secretes the cell wall. Within the cytoplasmic space there can be numerous discoid or singular plate-like chloroplasts (not shown).

PHYLOGENETIC RELATIONSHIPS

The silica frustules that are characteristic of all diatoms are dotted with highly organized pores that extend completely through the frustule. The pores enable diatoms to interact with the external environment via diffusion or active uptake of dissolved inorganic and organic substances from the surrounding water. The frustule is often covered on the exterior with polysaccharides or sulfated proteoglycans, whereas the inside of the frustule is coated in a separate organic layer called the diatopetum that protects the silica cell wall from dissolution and binds the siliceous cell wall components together (Hoagland et al. 1993; McConville et al. 1999). Frustule architecture and morphometry have been used to classify diatoms into genera based on overall shape and degrees of symmetry (Round et al. 1990), however, they differ in species-specific patterns and their functions in various biological processes (Kröger and Poulsen 2008). Studies of diatom valve morphogenesis—the processes of frustule silicification—have identified several phylogenetically informative characters within and between groups of diatoms capable of establishing either relatedness (Cox 2010, 2011) or seemingly homoplastic characters between unrelated genera (Cox and Williams 2000; Cox 2006). Further work is needed to understand both the genetic basis of these and other siliceous morphological characters to understand ancestor-descendant relationships with assessment of apomorphic and plesiomorphic characters on a phylogeny (Theriot et al. 2006).

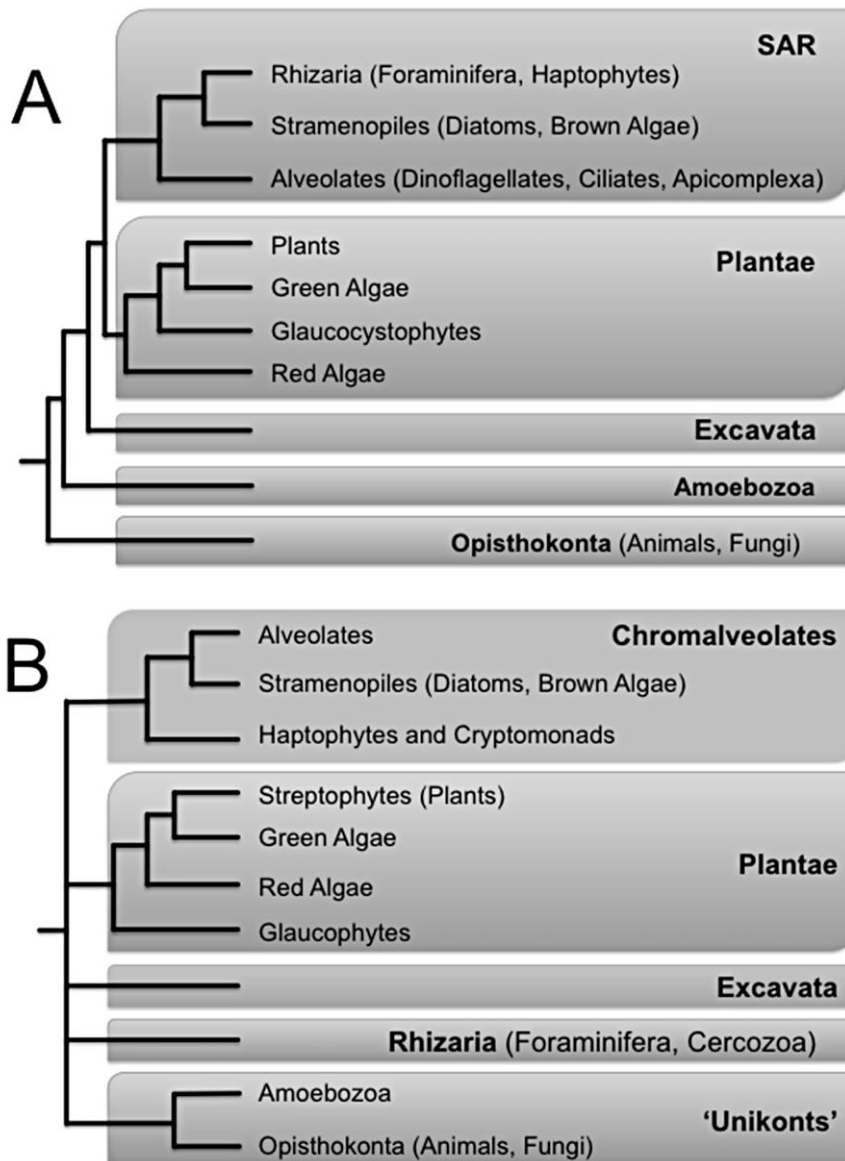


Figure 2. Two possible placements of the diatoms within the Tree of Life. A) Phylogenetic placement adapted and generalized from Parfrey et al. (2010). SAR indicates the clade composed of Stramenopiles, Alveolates, and Rhizaria. B) Phylogenetic placement adapted and generalized from Keeling et al. (2005).

The precise position of diatoms within the higher Eukaryotic Tree of Life is debated. Diatoms are single celled photoautotrophic eukaryotic algae and form a phylogenetic clade separate from plants and opisthokonts (animals and fungi). Some studies place diatoms between the Stramenopiles, Alveolates and Rhizaria group (Parfrey et al. 2010; Katz et al. 2012), whereas other studies place them within the Chromalveolates (Keeling et al. 2005) (Figure 2). Regardless of the exact position of where diatoms reside within the Tree of Life, it is clear that they evolved via secondary endosymbiotic events evidenced within their genomes (Bowler et

al. 2008; Tirichine and Bowler 2011) and by the presence of double membrane structures observed within all diatom cells. Of the two complete diatom genomes that have been sequenced and assembled (*Thalassiosira pseudonana* and *Phaeodactylum tricornutum*) the gene identities appear to be a combination of red algae, green algae, bacterial genes, and genes common to animals (heterotrophic organisms) (Armbrust et al. 2004; Bowler et al. 2008). In addition to their diverse gene ancestry, their genomes are also highly divergent at the sequence level. Results of comparative genomic analyses between *T. pseudonana* and *P. tricornutum* show that these two species fall within the range of the divergence observed between humans (*Homo sapiens*) compared to pufferfish (*Takifugu rubripes*) and humans compared to tunicate (*Ciona intestinalis*).

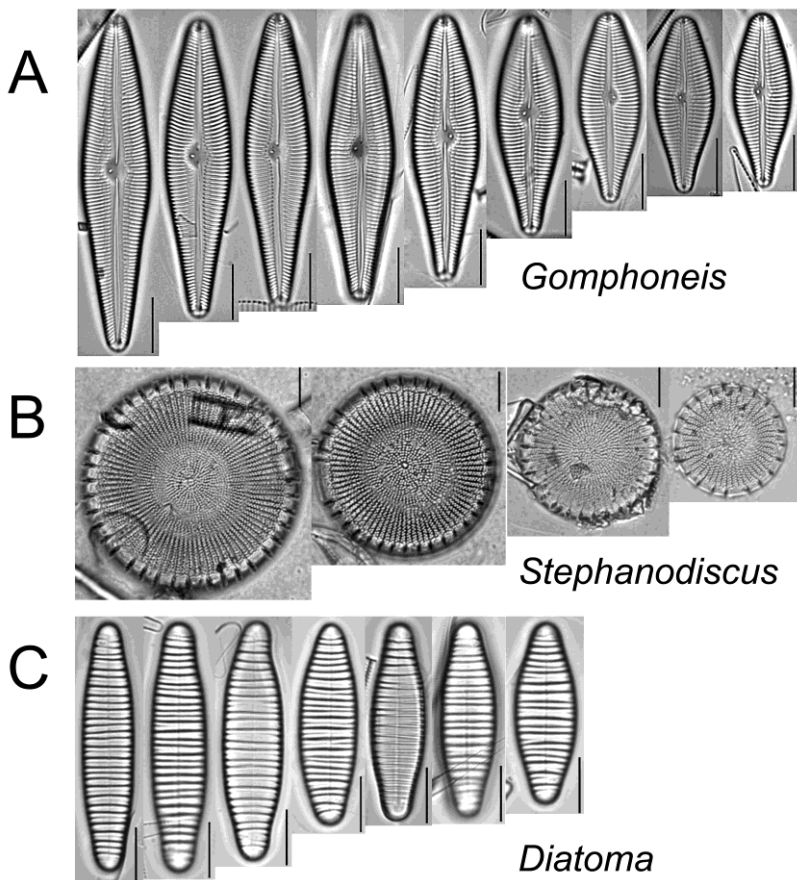


Figure 3. Diatom size diminution series in three representative diatom genera from field samples that shows the generalized pattern of cell size reduction with successive mitotic divisions proceeding from left to right. A) Raphid diatom in the genus *Gomphoneis*. The raphe is the vertical slit structure seen in the middle of the valve through which the diatom extrudes microfilaments that allows for cell motility. B) Centric diatom in the genus *Stephanodiscus*. C) Araphid diatom in the genus *Diatoma*. Scale bars are 10 µm.

T. pseudonana and *P. tricornutum* each represent two classically defined groups based on overall morphological symmetry: centric diatoms (e.g., *T. pseudonana*) that possess radial symmetry and pennate diatoms (e.g., *P. tricornutum*) that possess bilateral symmetry around a

sternum and an elongate shape (Figure 3). Within the pennate diatoms, those with a raphe (raphid) and those without (araphid) are also subdivided. The raphe is a slit-like structure through the frustule that allows diatoms to exude microfilaments and move across substrates. Many diatomists recognize the classification of Round et al. (1990) who published the first truly comprehensive work on diatom biology using a combination of light microscopy, scanning electron microscopy (SEM), and documentation of both fossil and extant taxa. Their classification further divides “centric” and “pennate” diatoms into Coscinodiscophyceae (formerly the centric group), Bacillariophyceae, and Fragilariaceae (both formerly within the pennate group). The introduction of molecular sequence analyses into constructing the diatom phylogeny (Medlin et al. 1988) has produced additional groupings and subdivisions into the centric diatoms (Coscinodiscophyceae), multipolar or bipolar diatoms (Mediophyceae), diatoms with bilateral/pennate symmetry (Bacillariophyceae) along with several other variations (Medlin and Kaczmarska 2004). Theriot (2010), however, suggests that diatoms form grades along characteristic morphologies: the radial centric grade, the polar centric grade, araphid pennate grade, and the raphid pennate grade. Regardless of the subtle differences among these proposed groupings, the overall picture that seems to be emerging from these studies is that it appears centric diatoms— both radial and polar— appear to be ancestral. Araphid diatoms share a most recent common ancestor with raphid diatoms, which together shared a common ancestor with the polar centric diatoms (Figure 2).

REPRODUCTION IN DIATOMS

To understand the mechanisms of speciation at work in diatoms, it is important to understand how diatoms reproduce. The life cycle of diatoms is unique due to the constraints of the siliceous cell wall (Figure 4). During vegetative growth, cells reproduce via mitosis, and as a consequence of the siliceous frustule, they produce slightly smaller cells within the parent cell (Figure 3, Figure 4). Diatom size diminution was first described independently by Macdonald (1869) and Pfitzer (1869, 1871). Both authors observed that many diatom species experience a binomial progression of reduction during consecutive cell divisions. This so-called Macdonald-Pfitzer rule was later formalized by Lewis (1984) as a regular decrease in the average size of a daughter cell, but with an increase in the standard deviation within the cell lineage over the course of successive mitotic divisions. The diminution process continues until the cells reach a certain species-specific size threshold, at which time they enter sexual reproduction. These size thresholds generally occur between 30% of the original cell size in radially symmetric centric diatoms and 40% of the original cell size in bilaterally symmetrical species (Drebes 1977; Lewis 1984; Potapova and Snoeijjs 1997). (Although this phenomenon is common to many diatoms species, exceptions do occur in which diatoms utilize their girdle bands to prevent mitotic cell size reduction, or in some cases, experience sudden breaks in size reduction (Geitler 1935)).

During the sexual reproductive phase, centric/polar diatoms (and some araphid and raphid diatoms) undergo a process of oogamous reproduction in which two “sexes” produce haploid gametes— either 1-2 eggs or 4-128 motile, flagellated sperm (Drebes 1977). Sperm swim toward the non-motile egg cells and upon fusion create a diploid zygote called the auxospore that separates from the parental frustules (Figure 4). The auxospore cell then swells into an

enlarged sphere that exceeds the size of the parent cell from which it was derived. Within the auxospore, the zygote deposits two new epitheca valves via a silica deposition vesicle and a mitotic division. The auxospore envelope is discarded, and the size is regenerated to the species maximum cell size.

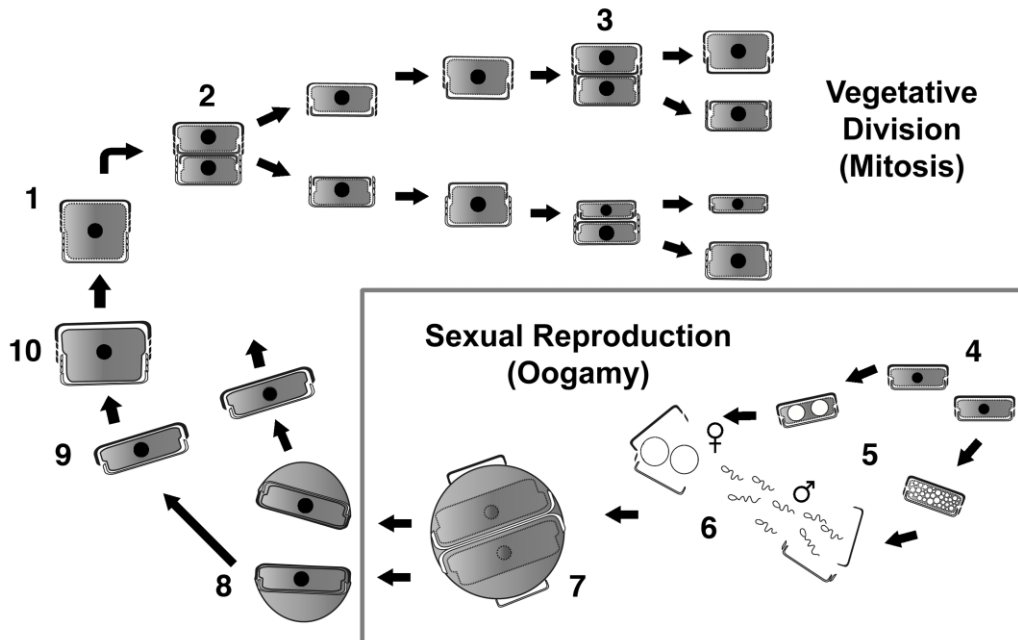


Figure 4. Life cycle of the radial centric diatom *Stephanodiscus* undergoing oogamous reproduction adapted from Round et al. (1990). 1-2) A vegetatively reproducing cell enters mitosis and develops siliceous hypotheca valves within the parent cell. The process begins near the nucleus and extends in both directions until completion. 2-3) Vegetative reproduction continues and cell size reduction occurs. The top sequence shows the fate of the original epitheca, which remains the same size, and gives rise to daughter valves of the same size class with each mitotic division. The bottom sequence shows the fate of the hypotheca, which produces smaller cells with each division, and ultimately gives rise to smaller daughter valve sizes than those derived from the epitheca. 4-5) Sexual reproduction begins when cells reach the critical size threshold and come into close proximity with one another. Each cell differentiates into either two non-motile cells (eggs) or several motile flagellated cells (sperm). 6) The frustules of sexualized cells open at the girdle bands and release the gametes. 7) Upon fertilization, the new cell swells to become the auxospore and becomes enveloped by a circular siliceous coat. One set of parental valves remains attached to the outside of the auxospore. Within the auxospore a new size regenerated cell is formed with the formation of two new epitheca valves. 8) The auxospore then splits into two hemispherical daughter valves. Within each valve, a new hypotheca valve forms to complete the frustule and the siliceous hemisphere is discarded. 9-10) The size-regenerated cells form new girdle bands between the epitheca and hypotheca, which allows the cell to expand and resume vegetative reproduction.

Most raphid diatoms, however, experience a different sexual process, in which two diatom cells generate gametangia (structures similar to those that house gametes in bryophytes) instead of gametes. An important feature of this form of reproduction is that because raphid diatoms are motile, the cells move toward one another to facilitate fertilization, instead of relying on motile gametes. Once the cells come into close proximity, each cell undergoes meiotic division, to produce four haploid nuclei. Three of these nuclei degenerate and the remaining nucleus

becomes the gametangium. The haploid gametangia are then transferred from one cell to another via copulation through a tunnel of protoplasm that connects the cells through the girdle bands (Drebes 1977). The cell then swells to create the primary auxospore wall (organic, well defined cell wall), which splits equatorially and persists as two caps on the ends of the now elongating auxospore (Figure 4; Trobajo et al. 2006). As the auxospore of pennate diatoms expands, a thin membrane called the perizonium is deposited as a set of organized bands both transverse and longitudinal appearing almost like siliceous scales that cover the mature auxospore (Trobajo et al. 2006). The zygote forms the new size-regenerated epitheca and hypotheca valves and the auxospore envelope is discarded. The cells then resume vegetative mitotic division until they again reach the critical size threshold and the entire process repeats.

Diatoms not only differ in the mechanism by which gametes are produced, or the process of zygote formation, but also in the reproductive capability of their gametes/gametangia. Some diatom species are homothallic and are capable of producing both male and female gametes/gametangia within the same individual. Other species are heterothallic, and generate either male or female gametes/gametangia, which requires another individual of the opposite mating type with which to mate (similar to + and - mating types observed in some fungi). Although the events described above represent the general reproductive events and strategies in diatoms, within each of these two general groupings several variations of these strategies also exist (see Geitler 1935 for a detailed description).

STUDYING SPECIATION IN DIATOMS

One of the greatest challenges to studying diatom speciation is understanding what constitutes a clear delimitation among diatom species. Although all speciation researchers face this consideration, it has posed a formidable problem to studying speciation in this group as the range in the number of estimated extant species of diatoms attests. Consequently, for most of the history of diatom research, discussion of speciation has focused largely on classifying diatom species, and only relatively recently have researchers begun to address questions that focus on the *process* of speciation. Species concepts and definitions have been reviewed thoroughly as applied generally to multicellular eukaryotic taxa (Hey 2001; Coyne and Orr 2004) and how these definitions apply to microeukaryotes has also been discussed at length (Mann 1989; Theriot 1992; Mann 1999; Amato 2010). Most of the arguments made to emphasize the benefits of employing one species concept or definition over another have strengths and weaknesses with regard to providing a framework in which to study speciation and we will not repeat these arguments here. Instead, we briefly discuss two species concepts in particular, that when used together, we feel provides the best framework for understanding the mechanisms that are important for speciation in diatoms.

Identifying species is, of course, necessary to study the processes by which species form. Because the exterior of the frustule is ornately decorated with striations, pores, spines, crevices, etc., it provides several characters that can be used to categorize individuals into distinct species. Beginning with the earliest diatomists, external morphology (Van Heurck 1896) or other biological characters such as the mechanism of auxospore formation (Geitler 1935) have been the primary methods used for assigning species status— thus, species status has predominantly relied on the Morphological Species Concept (MSC) in diatom taxonomy.

However, the structural variation observed among groups of diatoms to delineate species becomes problematic to studying speciation for two reasons. First, there are no clear criteria to establish what degree of morphological differences is necessary and sufficient to define good species. This is a general limitation of the MSC (Coyne and Orr 2004), but is particularly problematic given the subtle, but frequent, ultrastructural differences among diatom taxa (Mann 1981, 1984; Theriot 1987; Pappas and Stoermer 2001; Shimada et al. 2003; Mann et al. 2004). Although frustule morphology does appear to have high heritability (Mann 1984; Mann 1989; Mann et al. 1999; Evans et al. 2007; Evans et al. 2008; Trobajo et al. 2010), intraspecific variation in morphology has made species groups difficult to define. Second, the unique life cycle and life history of diatoms makes categorizing diatom species using morphology alone difficult. The reason is that many diatoms possess the diplontic life cycle discussed above: they experience asexual, vegetative reproduction punctuated by regular generations of sexual reproduction. During the progression through the vegetative phase, diatoms experience a reduction in size. As a consequence, external morphological features (*e.g.*, number of pores, striation pattern) change and can differ dramatically in both size and shape (Figure 3). To further complicate matters, the appearance of Janus valves, a phenomenon whereby the epitheca and hypotheca form different morphological patterns between consecutive cell divisions, also has the potential to obscure species boundaries in mixed populations. Although the reason for the occurrence of Janus valves is unknown, the result is clear: individuals of the same species may appear as individuals that belong to separate species (Stoermer 1967; McBride and Edgar 1998). As a consequence of these two biological phenomena, individuals within the same species may appear morphologically different, and individuals of different species may appear morphologically identical when categorized on the basis of external morphology alone (Kocielek and Stoermer 2010). This has given rise to considerable uncertainty in species delimitation, prompting many researchers to delineate a myriad of varieties and so-called phenodemes instead of what might be considered good species (Mann 1999).

Perhaps the most important limitation of the MSC for studying speciation in diatoms, is that although the MSC may provide some guidance to identifying what species are, it makes no predictions on how species evolve (Coyne and Orr 2004). For this reason, many diatomists have considered the Biological Species Concept (BSC) more useful to understand the process of speciation (Mann 1999; Amato 2010). At its core, the BSC defines species as groups of interbreeding organisms that are reproductively isolated from other groups of interbreeding organisms (Mayr 1942, 1963). The BSC is not without its limitations, one of which is the problem of applying such a definition to completely allopatric populations, but it does offer a clear advantage to understanding the speciation process (Coyne and Orr 2004): when we understand the evolution of RI, we begin to understand the process by which new species evolve. Although the extent of sexual reproduction among diatom taxa remains unclear, some diatoms do experience a routine sexual phase as a crucial part of their life cycle to restore cell size. Thus, the possibility exists for hybridization and gene flow among co-occurring diatom populations in the absence of any barriers to reproduction, particularly if these populations experience their sexual reproductive stage at the same time. It is also important to mention that although external morphology alone has limitations for defining diatom species, with regard to understanding the speciation process, differences in morphology do appear to make some contribution to restricting gene flow between some diatom populations (discussed below). For these reasons, we propose that to understand speciation in diatoms, both the evolution of

morphology and the evolution of RI need to be considered to fully investigate the speciation process.

PATTERNS OF SPECIATION IN DIATOMS

Marine Records

Marine diatom fossils occur early in the fossil record with the first reported appearance of recognizably diatom-like cells dating to the Jurassic period 175 million years ago (Myr) (Rothpletz 1896). However, this claim has been impossible to verify as the particular fossil material was lost during the bombing of Berlin in World War II along with many other scientific collections (for a historical account of this incident, see Falkowski et al. 2004). The first well-documented diatom fossils date to the Upper Cretaceous (Harwood and Nikolaev 1990), at which time diatoms appear to have already evolved considerable diversity. Diatoms during this time appear to have been either solitary or formed chains with long spines to possibly increase drag and remain in the photic zone where plankton are abundant. Recent studies of diatoms show that buoyancy can be achieved in this manner: some extant species can produce chitin threads that extrude from strutted processes on the valve, which extend outward and increase buoyancy, or can connect cells to form strings of diatoms (Durkin et al. 2009).

Molecular studies using biostratigraphic markers of unambiguous diatoms and ancient fossil lipids indicative of particular diatom lineages (*e.g.*, rhizosolenioid diatoms (Damsté et al. 2004)) suggest an origin of the group around the Permian/Triassic extinction event (Kooistra and Medlin 1996; Medlin et al. 1997). Due to the highly alkaline conditions of seawater at increasing depths and marine biochemical cycles that occur due to diagenesis of marine sediments, pre-Cretaceous intact diatom fossils are rare because silica dissolves with increasing depths and often completely degrades. The same phenomenon has been observed in the contemporary Arctic/Antarctic Oceans due to production in nutrient rich upwellings (Sarmiento et al. 2004). Moreover, when diatom fossilization does occur, although mineral replacement with pyrite leaves the general diatom cell shapes intact, it often obscures the finer details of frustule morphology. Without the fine ultrastructure, genus or even species assessment is limited to only the valve outline. The fossil record is also skewed toward species that produced heavily silicified frustules or resting spores— a dormancy strategy in which large amounts of silica are deposited onto the valve, or additional valves are formed within the parent frustule to survive abiotic stresses. This ultimately creates gaps in the fossil record for some species and makes the biostratigraphic origin of species that are lightly silicified (which includes many recent genera) difficult to assign. However, renewed interest in diatom micropaleontology has begun to fill in some of these gaps for parts of the diatom Tree of Life (Chacon-Baca et al. 2002; Harwood et al. 2004; Singh et al. 2006; Witkowski et al. 2011). Re-examination of previously published records would also prove useful to verify the origin of genera that have recently undergone taxonomic splitting, as well as discovering deposits aged between the ambiguous Jurassic record and the first unambiguous Cretaceous record to help fill any remaining gaps.

It also appears from the Upper Cretaceous fossil record (with support from molecular phylogenies) that diatoms are ancestrally marine and planktonic, are thought to have first evolved with a radially symmetric cell shape and little to absent motility, and show a directional

trend toward elongate shape and motility (Alverson et al. 2006; but see Harwood et al. 2004 for a potential terrestrial origin of diatoms in shallow ephemeral pools). Another major trend in diatom evolution was from a purely planktonic existence to a close association with a substrate that required the evolution of motility and/or attachment. Motility on a substrate is made possible by the raphe (Figure 3A), which appears as a longitudinal fissure on the valve face (epitheca/hypotheca) of one or both valves; motility is achieved via actin/microtubule-based movement through gaps in the raphe (Edgar and Pickett-Heaps 1983; Poulsen et al. 1999). The evolution of the raphe generally seems to be an adaptation to sediment burial and perhaps a strategy to access buried nutrients such as nitrate/nitrite fixed by bacteria. Several studies have also shown that the raphe is important in maintaining diurnal movement of sediment-associated diatoms such as *Pinnularia* that migrate to the surface of pond sediment toward the light before midday by phototaxis, and then bury themselves before the high light part of the day (Harper 1976). Raphe also allow diatoms in the genus *Hantzschia* to respond to other environmental change such as emerge when the tide recedes and bury themselves prior to flooding (Kingston 1999).

Species diversity from the Jurassic to the Cenozoic appears to have remained largely unchanged with the addition of only few new genera (Sims et al. 2006). It appears that diatoms were also one of the few groups that seem to have survived relatively unscathed by the Cretaceous/Tertiary (*i.e.*, KT Boundary) mass extinction. Studies of Ocean Drilling Project cores and terrestrial deposits of marine origin suggest that diatom diversity increased throughout the Cenozoic with two bursts of rapid diversification during the Eocene/Oligocene and one at the middle to late Miocene (Falkowski et al. 2004). However, this initial pattern might reflect a biased sampling of those fossil sediments that were more commonly exposed during sampling. Rabosky and Sorhannus (2009) provide a different estimate that suggests marine diatom biodiversity reached its highest level between the Eocene and Oligocene (35-30 Myr), then decreased dramatically during the late Oligocene, and subsequently rose to half the maximum observed during the late Eocene. Falkowski et al. (2004) also suggest that the pattern of increasing diatom biodiversity follows a trend with the diversification of grasses during the Miocene. They hypothesize that this correlation is potentially tied to the increased dissolution of terrestrial silica from soils by grass that would introduce greater amounts of silica into estuaries (oceans are typically silica-limited), which might have allowed diatoms to diversify. However, the strength of this relationship is unclear, as the analysis performed by Rabosky and Sorhannus (2009) shows no correlation between grassland expansion and diatom diversity.

Evidence of speciation events in the fossil record of marine diatoms has been carefully examined in only a few examples. These studies typically focus on the biostratigraphy of species in ocean drilling core records and use a strict analysis of transitions in morphology to infer speciation events. Koizumi and Yanisagawa (1990) characterized the evolution of *Pseudoeunotia dolious* from oceanic core records east of Japan, and observed a gradually increasing convex dorsal margin (asymmetrical form) from a species with *Nitzschia fossilis*-like morphology (strictly bilateral and symmetrical form) that appeared to coincide with climatic deterioration during the Pleistocene. However, this apparent speciation event does not correspond precisely with the climatic deterioration during the Pleistocene and the authors speculate that development of the convex dorsal margin is likely the result of adaptation to some other factor(s) (Koizumi and Yanagisawa 1990). They also claim that *N. fossilis*, the presumed ancestor of *P. dolious*, was once abundant in North Pacific during the Pliocene, but was rapidly replaced by *P. dolious* over the course of roughly 600,000 years. Koizumi and

Yanisagawa hypothesize that *P. dolious* evolved its characteristic morphology at the beginning of the Pleistocene when surface water temperature decreased, and they observe increased asymmetry in younger and younger core layers. Curiously, the occurrence of *P. dolious* is slightly earlier in the lower latitudes (2.0 Myr) than in middle latitudes (1.89 Myr). Although this study presents some potentially interesting patterns of speciation between these two species, the results provide only suggestive evidence of the evolutionary relationships between these two taxa, and are based strictly on phenetic similarity. It is plausible that *P. dolious* is simply a morphological variant of *N. fossilis* and not a distinct species. But, if these two populations do represent an example of anagenesis in the fossil record, then the tempo of speciation (as well as morphological evolution) in marine diatoms appears slow, at least in this case, as 10^6 years pass before complete replacement of *N. fossilis* by *P. dolious* occurs.

Another example of speciation in the marine paleontological record comes from the genus *Rhizosolenia*, in which a cladogenetic event that occurred around 3.2-3.1 Myr (end of the Pliocene) gave rise to two modern diatom lineages *R. praebergonii* and *R. bergonii* in the Pacific Ocean (Sorhannus et al. 1988). An examination of core records shows that *R. praebergonii* and *R. bergonii* split from their common ancestor and experienced different amounts of morphological evolution. In particular, *R. bergonii* appears to have retained a morphology that is closer to that of the ancestral form, whereas *R. praebergonii* appears to have experienced substantial morphological evolution. It also appears that the tempo of morphological change fluctuated dramatically over time: *R. praebergonii* began to diverge gradually from *R. bergonii* in the Pacific Ocean around 3.1 Myr, then appears to have experienced several bursts of morphological evolution leading to its present morphology (Sorhannus et al. 1988).

R. praebergonii first appears in cores sampled from the Equatorial Pacific, and co-occurs with *R. bergonii*. Two possible scenarios have been offered to explain speciation events in *R. praebergonii* based on patterns observed in core samples obtained from locations around the Pacific (Sorhannus et al. 1988; Sorhannus 1990). First, the appearance of *R. praebergonii* might represent a potential case of localized sympatric speciation, followed by rapid dispersal throughout the Pacific region. Second, parallel cladogenetic events might have occurred over the entire range of the Equatorial Pacific to give rise to new species. Sorhannus (1990) favors the sympatric speciation-migration scenario over parallel allopatric speciation, and reasons that because *R. praebergonii* shows little variation in morphology it seems less likely the species experienced multiple independent cases of morphological evolution. He also suggests that one possible mechanism that might have given rise to the pattern observed in the marine record is size-selective predation and anti-predation adaptation that decreased cell size, as core records show that the length of the apical spine possessed by these species decreased in length in *R. praebergonii* over time. These subtle changes in valve morphology with rapid decreases in cell size in *R. praebergonii* followed by stasis were consequently interpreted as a rapid transition across an adaptive valley to a new adaptive peak. However, little historical or contemporary evidence exists for even a suggested anti-predator or grazing pressure on these species. In fact, if the absence of predation pressure drove the evolution of morphology, its effects should also be observed in *R. bergonii*, yet this species shows little change in apical spine length compared to the ancestral form. Experimental studies with modern *Rhizosolenia* taxa could prove informative for testing the possible selective advantage of shorter versus longer apical spines (with longer spines increasing drag, reducing sinking potential) to grazing pressure. Adaptation to predation represents only one possible explanation for the morphological change in *R.*

praebergonii, as many environmental factors are known to influence the silica morphology of diatom species, such as changes in salinity (Schultz 1971; Tuchman et al. 1984; Saros and Fritz 2000; Trobajo et al. 2011), silica availability (Finkel et al. 2010) and nutrient availability (Theriot et al. 1988b).

A third morphologically distinct *Rhizosolenia* species, *R. sigmoida*, also appears to have evolved from a *R. bergonii* ancestor, and all three species coexisted briefly (Sorhannus et al. 1991). Interestingly, *R. sigmoida* is found only in the ocean drill core taken from the coast of Peru, and disappears around 3.13-3.07 Myr after persisting for just 60,000 years. Sorhannus et al. (1991) suggest that the environmental factors around coastal Peru at that time (cooler waters, greater range of salinity and central equatorial water mass) may have driven adaptations that ultimately provided barriers to gene flow among populations of *R. bergonii*. They also speculate that the disappearance of *R. sigmoida* may have been the result of either competitive exclusion for some nutrient/resource that ultimately led to extinction, or that *R. sigmoida* might have hybridized and merged with either one or both of its two sister taxa.

Freshwater Records

Although most paleontological work has focused on marine sediments, some data do exist that describe patterns of speciation in freshwater diatom taxa. Lake Baikal, one of the Earth's oldest (approximately 20 Myr; Mackay 2007) and largest freshwater lakes, provides a detailed diatom history revealed in sediment core samples (Mackay et al. 1998; Khursevich et al. 2001; Khursevich et al. 2002; Khursevich et al. 2005; Mackay 2007). Baikal lake cores detail dramatic and long-term changes in the lake's diatom community structure over time, which include several species appearances followed by their subsequent extinctions. In particular, the record chronicles the evolution of centric diatoms associated with processes of extinction and renewal that appears closely associated with periods of global climate changes (Khursevich et al. 2000). Lake Baikal's fossil record is also unique in both the age and preservation, which has provided a high temporal resolution for studying the evolution of freshwater centric diatoms. The lake's well-dated sediments and long geological record have thus allowed researchers to designate biostratigraphic zones carefully defined by the appearance, abundance, and extinction of particular diatom genera or species (Khursevich et al. 2001; Khursevich et al. 2005). For example, the sediment zones that belong to the Late Miocene-Early Pliocene periods show the appearance, speciation, and extinction of the genus *Mesodictyopsis*, a genus found only in the Lake Baikal (Khursevich et al. 2004). The Pliocene epoch sediments are characterized by the appearance, speciation, and extinction of another endemic genus, *Stephanopsis*.

The Pleistocene epoch sediments are also further distinguished by the appearance and active speciation within the genus *Stephanodiscus* (Khursevich et al. 2004). Throughout its 5 million year history, Lake Baikal has been home to more than 21 species of *Stephanodiscus* (Khursevich et al. 2000; Khursevich et al. 2005). The diatom flora of Lake Baikal experience frequent solar insolation fluctuations caused by the 23,000 and 43,000 cycles of the Earth's precession and obliquity (Khursevich et al. 2001; Karabanov et al. 2004). Changes in global ice volume also have a substantial impact on local precipitation, which gives rise to increases in nutrient loading and terrestrial inputs during wetter years, and decreases in nutrient availability that leads to stratification or earlier ice-out resulting in higher productivity during drier years. This close relationship between the global and local aquatic environments is one of

the many reasons diatoms are, and have been, used to track changes in water chemistry for paleolimnology, water quality, and general ecosystem perturbation (Smol et al. 1986; Dixit et al. 1992; Stevenson and Pan 1999; Potapova and Charles 2002, 2003, 2007; Smol and Stoermer 2010; Potapova 2011). Because diatom species are particularly sensitive to alternating environmental conditions, these fluctuations provide permissive conditions that have the potential to drive periods of rapid speciation or extinction.

The rapid appearances and extinctions of *Stephanodiscus* species, in particular, are thus likely linked to the pronounced periods of solar insolation that might have occurred as a consequence of the rising of the Tibetan Plateau in the Late Miocene-Early Pliocene (Khursevich et al. 2000). Khursevich et al. (2000) suggest the rapid speciation in the genus was also the result of frequent changes in climatic conditions during a period of alternating warming and cooling phases. These temperature phases were the product of Siberian glacial expansion and retreat that affected the temperature of the region. Lake Baikal's five million year old record of diatom biostratigraphy and long climatic cycles suggests a coupling of speciation events to these erratic climatic conditions. Reorganizations of the diatom communities coincide with the climate changes that alter the lakes abiotic conditions through increases or decreases in nutrient availability, light, water column mixing, temperature, and periods of ice cover could preferentially select for differential survival and reproduction of diatom clonal lineages.

REPRODUCTIVE ISOLATION

Much of what evolutionary biologists have learned about the processes that give rise to speciation has come from studying several reproductive isolating mechanisms originally described by Dobzhansky (1937, 1951). These include barriers to gene flow that might exist between two populations that act either before fertilization (prezygotic isolation) or after fertilization (postzygotic isolation). Among the marine and freshwater species of diatoms where RI has been studied, both forms of RI are observed. Although prezygotic barriers appear more common in preventing gene flow than postzygotic barriers, it is difficult to estimate the importance of postzygotic isolation because in most of these cases prezygotic isolation is complete, which precludes the possibility of studying fertilization defects, or generating interspecific F_1 hybrid individuals. Two interesting patterns of RI are also emerging from recent studies. One pattern is that it appears considerable polymorphism for RI exists and segregates within previously defined species groups, similar to observations made among populations of some multicellular eukaryotic species (Takahashi et al. 2001; Sweigart et al. 2007; Good et al. 2008; Reed et al. 2008; Yukilevich and True 2008; Nolte et al. 2009; Vyskočilová et al. 2009; Lachance and True 2010; Martin and Willis 2010; Teeter et al. 2010; Kozłowska et al. 2012). (Caution is needed in this interpretation, however, as this might simply reflect the nature of species definitions in diatoms.) The other emerging pattern is that cell morphology seems to play an important role in contributing to RI in some cases, which suggests that understanding the evolution of cell morphology could be important to understanding speciation in diatom taxa.

Mate Discrimination

Among co-occurring diatom populations, conspecific mate recognition mechanisms within species or demes (varieties) within species appear common in preventing hybridization. Indeed, the results of crosses between morphologically dissimilar diatoms were first recorded beginning in the late 1950's by Geitler who observed that hybridization failed to occur in many of these cases (Geitler 1958a, 1958b, 1975). The first quantitative experiments to test premating isolation in diatoms were made by crossing sympatric populations of *Navicula pupula* (current taxonomic designation of this species is *Sellaphora pupula*) and *Amphora ovalis* that inhabit Blackford Pond in Edinburgh, Scotland (Mann 1984). Mann (1984) studied the pairing that occurs during the sexual phases among four morphologically distinct phenodemes of *S. pupula* (phenodemes 1-4) and two morphologically distinct varieties of *A. ovalis* (var. *affinis* and var. *pediculus*). Populations of each phenodeme or variety were induced to enter the sexual phase of the lifecycle by adjusting the environmental conditions in the laboratory (diatom cells that have entered the sexual reproduction life stage are hereafter referred to as "sexualized"), and upon sexualization, auxosporulation became synchronized, which allowed for crosses among the different populations to be performed. When the sexualized populations were mixed together, pairing between individuals from the same phenodeme or variety would occur readily. However, pairing was never observed among sexualized individuals when crosses were performed between different phenodemes of *S. pupula* or different varieties of *A. ovalis*.

A lack of pairing among morphologically distinct demes has also been observed in interdeme crosses of the freshwater diatoms *S. pupula* and *S. laevissima* (Behnke et al. 2004). Behnke et al. (2004) studied 11 strains of freshwater *S. pupula* and 2 strains of the closely related species *S. laevissima*. Morphologies observed among the *S. pupula* demes fell into three categories: a "rectangular" morph, which is longer and more robust than the shorter, more slender "capitate" and "pseudocapitate" morphs, and a "small" morph, which possesses a much reduced cell size compared to the rectangular and capitate/pseudocapitate morphs. The two *S. laevissima* demes both differ in size and striation patterns compared to all three *S. pupula* demes. Premating RI was observed when interdeme crosses were performed among several demes; in each cross individual cells failed to form pairs required to exchange genetic material. Interestingly, in contrast to the crosses performed among the *S. pupula* and *A. ovalis* demes, cell morphology in these experiments was not a good indicator of pairing capability, as sexualized demes that possessed similar morphologies would sometimes fail to pair, whereas demes with different morphologies would sometimes pair. These results show that premating barriers have evolved in the absence of any substantial morphological evolution within *Sellaphora*.

However, in some diatom taxa morphology does appear to have a large effect on preventing mating. Mann and Chepurinov (2005) observed sexual pairing in three demes of the freshwater raphid diatoms that belong to the *Neidium ampliutum sensu lato* complex: *N. iridis*, and two morphologically distinct demes of *N. ampliutum*, which contains a "major" deme and "minor" deme that differ on the basis of cell size and frustule striation density. The major deme is characterized by wider frustule sizes and less dense striation patterns compared to the minor deme. Mating within these demes occurs when sexualized individuals pair and orient themselves so that they align girdle-to-girdle, which allows fertilization to occur. Interactions in mixed populations of *N. ampliutum* and *N. iridis* never resulted in attempted pairing. In the *N. ampliutum* phenodemes, mating and auxospore formation occurred in crosses within both

the major and minor demes of *N. ampliatum* provided that individuals were a mix of opposite mating types. However, crosses in mixed populations of individuals from the major and minor demes failed to align correctly and never appeared to mate. In particular, pairing between a major cell and a minor cell was observed only three times in mixed populations. In each of these cases there was no conclusive evidence that fertilization had occurred. Interestingly, the strength of mating preference on the basis of cell size appears to be different between the two demes. Individuals that belong to the major deme would mate readily with other major cells even when their cell sizes were considerably different, but individuals that belong to the minor deme preferentially pair with other minor deme cells if they were similarly sized. Although this suggests that mate discrimination on the basis of cell size might operate within the *N. ampliatum* minor deme, this result might be an artifact of cell size required to enter the sexualized reproductive stage, as sexualization in the minor deme becomes more frequent as cells become smaller (Mann and Chepurinov 2005).

Because diatoms lack any developed visual sensory system, the existence of apparent mate recognition systems suggests that diatoms secrete chemical cues into the surrounding water column to attract potential mates. Recent work has begun to dissect the nature of these signals. Sato et al. (2011) studied sexual reproduction in araphid pennate diatom *Pseudostaurosira trainorii*, which produces motile male gametes and immotile female gametes. Motility of the male sperm is generated by extrusion and retrieval of microtubules through pores in the frustule. Through a series of elegant experiments that paired vegetative and sexualized individuals of each sex, Sato and colleagues demonstrated the existence of both male and female sex pheromones that act to guide fertilization. Sexual reproduction begins when vegetative females secrete ph-1 pheromone, which induces sexualization in males. Males then release their gametes, and the sperm cells explore the nearby area by swimming in a random walk. Once induced, sexualized male cells also secrete the ph-2 pheromone, which stimulates sexualization of vegetative female cells in the nearby area. The sexualized female cells then direct fertilization when they secrete a putative ph-3 pheromone, which results in directed ameboid movement of sperm toward the female's eggs. Thus, the possibility exists that divergence between sex pheromones and their recognition systems might give rise to species-specific pheromonal cues and cause RI.

Temporal Isolation

Differences in the timing during which some diatoms enter the sexualized life stage also act to limit gene flow between species. Two closely related centric diatoms *Aulacoseira skvortzowii* and *A. baicalensis* reside in Lake Baikal and experience environmentally driven differences that govern the timing of their respective life stages (Jewson et al. 2008; Jewson et al. 2010). Many diatom lineages have an additional component to their life cycle that involves the formation of resting spores or resting cells used to survive unfavorable environmental conditions (McQuoid and Hobson 1996). Both mechanisms operate much like dormancy in vascular plants by enabling bet-hedging population emergence and differential survival of lineages (Finch-Savage and Leubner-Metzger 2006). Resting cells are physiologically dormant cells that are packed with reserve nutrients, whereas resting spores are characterized by a thickening of the siliceous wall causing the cells to sink to greater depths or to sink out of the water column (Sicko-Goad et al. 1986; Sicko-Goad and Andresen 1991; Edlund et al. 1996). It

is thought that resting spore formation is the ancestral state in diatoms, and might also function as a survival mechanism against desiccation, grazing, and dissolution (Hargraves and French 1983). Environmental factors such as light limitation, nutrient depletion, or changes in temperature can act as triggers to initiate resting dormancy strategies (McQuoid and Hobson 1996).

A. skvortzowii and *A. baicalensis* both produce resting cells, however only *A. skvortzowii* produces thick resting spores as a perennation strategy (Jewson et al. 2008; Jewson et al. 2010). *A. baicalensis* grows in the pelagic zone of Lake Baikal under the January ice until May/June when it forms resting cells as the lake water becomes stratified due to warmer summer temperatures. *A. baicalensis* resting cells contain high concentrations of nutrients required to survive at lower lake depths. However, due to silica dissolution with increasing lake depth, this diatom also experiences a physiological trade-off as a consequence of resting cell formation by increasing wall thickness to reduce risk of dissolution at the expense of decreasing cell volume and limiting nutrient reserves (Jewson et al. 2010). *A. skvortzowii*, which resides nearshore and at thermal bar regions within the water column, appears in the plankton from May until early autumn. Due to reduced light during autumn-winter, *A. skvortzowii* produces resting spores in response to the reduction in phosphate concentration during summer. The spores sink to an intermediate depth of ~50 meters where they may persist until the next spring water strata overturn.

During late May to early June before stratification of the lake, phosphate levels drop below 6 µg/L as a consequence of the increase in picoplankton growth (Nagata et al. 1994; Belykh and Sorokovikova 2003; Fietz et al. 2005). In *A. skvortzowii*, the decrease in phosphate concentration induces resting spore formation in individuals that have cell diameters that exceed 10 µm. However, for smaller cells, the drop in phosphate concentration induces auxosporulation, and consequently an increase in cell size, before resting spore formation (Jewson et al. 2008). In contrast, *A. baicalensis* forms resting cells in early to mid-June when water temperatures in the south basin exceed 6 °C. Auxosporulation in this species occurs between February and April while a sheet of ice still exists on the lake surface, thus preventing these two species from interbreeding.

F₁ Hybrid Sterility and Inviability

Because premating isolation is complete between many species of diatoms that have been studied for RI, studying the frequency and the importance of postzygotic isolation has been difficult. However, in the handful of cases where prezygotic isolation is incomplete, it appears that postzygotic isolation also exists as a barrier to gene flow between species. Because of the nature of the diatom sexual reproduction, it is difficult to separate intrinsic postzygotic isolation into F₁ hybrid sterility and F₁ hybrid inviability: viable F₁ hybrids ultimately die if they are sterile because they are unable to form auxospores. An example of this is observed in crosses among demes of *Eunotia bilunaris sensu lato* species complex isolated from freshwaters in Belgium where two particular demes, the "slender" and the "robust" demes, show strong postzygotic isolation (Vanormelingen et al. 2008). Crosses within each deme proceed easily, but interdeme crosses are considerably more difficult to perform (~20–400 times less frequent), although not impossible. Vanormelingen and colleagues observed that these crosses between slender and robust individuals often need to be repeated to obtain F₁ hybrid progeny due to lack

of sexual activity, which suggests that pheromonal cues might govern mating in these species. Crosses within each deme resulted in pairing and the production of F₁ hybrid individuals that formed healthy auxospores. In contrast, only 12 interdeme F₁ hybrids were obtained, and 9 of 12 ceased vegetative reproduction after a few mitotic divisions and died (Vanormelingen et al. 2008).

Postzygotic isolation also appears to act between populations of some marine diatoms. Amato et al. (2007) collected 95 strains of *Pseudo-nitzschia* genus from the Gulf of Naples. Their collection included several isolates of six different species: *P. delicatissima*, *P. pseudodelicatissima*, *P. dolorosa*, *P. cuspidata*, *P. calliantha*, and *P. cacialantha*. Because these demes are difficult to distinguish morphologically (many external characters overlap among the groups), they were classified using characteristic DNA sequence variation at the ribosomal Internal Transcribed Spacer 2 (ITS-2) locus, which proves to be a good indicator of deme identity. For the majority of demes, sexualization could be induced and crosses within demes occurred readily and produced viable, fertile F₁ progeny. Most interdeme crosses resulted in no sexual reproduction due to the failure of cells to pair. In crosses between two demes of *P. calliantha*, however, prezygotic isolation was incomplete and F₁ individuals were obtained, but failed to develop into auxospores and consequently died (Amato et al. 2007).

THE GENETICS OF REPRODUCTIVE ISOLATION

Dissecting the genetic basis of RI in diatoms is a burgeoning field. Although they are usually not considered to be model genetic organisms, the availability of both modern molecular techniques and genome sequence data (Armbrust et al. 2004; Bowler et al. 2008) are promising developments for beginning to understand speciation at the molecular level in the microeukaryotes. Much of the work on the genetics of speciation in diatoms has thus far focused on divergence at the ribosomal ITS-2 locus as an indicator of reproductive compatibility (Coleman 2009). ITS-2 sequences are often used to assign diatoms to demes or varieties within species (*e.g.*, Amato et al. 2007) or construct diatom phylogenies (*e.g.*, Godhe et al. 2006; Evans et al. 2007; Härnström et al. 2011). ITS-2 sequences also appear to correlate with the degree of RI observed among populations: demes that possess similar ITS-2 sequences will mate and produce viable, fertile offspring and those with increasingly divergent sequences fail to mate (Behnke et al. 2004; Amato et al. 2007). Although the ribosomal ITS RNAs are required to construct ribosomal subunits, they are not required for the function of mature ribosomes (Tschochner and Hurt 2003; Staley and Woolford Jr 2009). Because it provides important functions for viability, it might be reasonable to speculate that in interspecific hybrid genotypes, divergence at the ITS-2 locus could give rise to incompatible epistatic interactions that are a common cause of postzygotic hybrid incompatibilities (Dobzhansky 1937; Muller 1942; Coyne and Orr 2004). However, it is unclear exactly how potential incompatibilities involving ITS-2 would give rise to *prezygotic* isolation observed in crosses (Behnke et al. 2004; Amato et al. 2007). One possibility is that postzygotic isolation via incompatibilities involving ITS-2 might evolve early during diatom speciation, followed by reinforcement that drives the evolution of premating isolation. At the current level of resolution, however, it is difficult to ascertain whether the pattern of evolution observed at ITS-2 is a cause or a consequence of

divergence between populations, and whether ITS-2 RNA is involved in any hybrid incompatibilities.

Another set of potential candidate genes that might cause RI are members of the *Sexually induced gene* (*Sig*) family that were first identified in the centric diatom *Thalassiosira weissflogii* (Armbrust 1999). *Sig1*, *Sig2*, and *Sig3* are all transcriptionally upregulated within a few hours after entry into the sexual reproduction phase of the life cycle, immediately preceding gamete formation. Molecular sequence analysis shows that each of the *Sig* genes encodes a protein with a putative hydrophobic signal sequence, but lacks long stretches of hydrophobic residues, which suggests that *Sig* proteins are likely secreted rather than membrane-bound. The predicted protein sequence also possesses several epithelial growth factor-like domains and bears strong similarity to vertebrate matrix glycoprotein tenascin X, an extracellular matrix (ECM) protein known to be functionally important for mediating cell-cell interactions (Armbrust 1999). It has thus been hypothesized that the *Sig* proteins could be important during gamete recognition and/or fusion. Consistent with patterns of molecular evolution observed at other genes known to mediate sperm-egg interaction and cause RI between species (Vacquier and Swanson 2011), *Sig1* also shows molecular signature of positive selection among *Thalassiosira* species (Armbrust and Galindo 2001; Sorhannus and Kosakovsky Pond 2006), but shows no signature of positive selection among *T. weissflogii* isolates (Suzuki and Nei 2004). Although the *Sig* proteins could prove important for fertilization success and cause RI as a byproduct of divergence between species, further work is needed to determine if these proteins localize to the ECM of the gametes themselves and whether they do indeed function during sperm-egg attraction or interaction.

Whole genome duplication and polyploidy may also prove to be important genetic mechanisms of speciation similar to what is observed for speciation in higher plants (*e.g.*, Soltis et al. 2009). Indeed, variation in ploidy level might be fairly common among diatoms: triploid and tetraploid lineages have been identified in several pennate diatom genera (Mann and Stickle 1991; Mann 1994; Chepurnov and Roshchin 1995). Studies of the marine planktonic diatom *Ditylum brightwellii* suggest that large-scale genomic changes such as these might contribute to rapid diversification between populations. Koester et al. (2010) examined three populations of *D. brightwellii* and quantified differences in genome size: one population was sampled from Akaroa Harbor in New Zealand, and two different populations (population 1 and population 2) were sampled from Puget Sound, Washington in the United States. Flow cytometry estimates indicate that although no significant differences in genome size exist between the isolate from New Zealand and population 1 from Washington, population 2 from Washington shows a roughly two-fold increase in genomic DNA content compared to the other two populations (Koester et al. 2010). Sequence comparisons of ITS-1 also show no difference between New Zealand and population 1, but characteristic differences at ITS-1 exist between populations 1 and 2. This result is also consistent with high F_{st} values between Washington populations 1 and 2, which indicates that limited gene flow occurs between them (Rynearson et al. 2006). Although the authors did not perform any laboratory crosses among these populations, comparisons of growth rates and cell sizes among the populations suggest that population 2 might be reproductively isolated from both population 1 and the New Zealand population. In particular, Washington population 2 possesses a larger cell size and slower growth rates, which suggests the possibility that the timing of sexual reproduction in population 2 might differ considerably from that of population 1 or from the New Zealand population.

DIATOMS: A MODEL SYSTEM TO STUDY THE CONSEQUENCES OF CLIMATIC CHANGE ON SPECIATION

As discussed above, diatoms are acutely sensitive to changes in their environment, and environmental change appears correlated in some cases with speciation events in the fossil record. Understanding how changes in the environment give rise to the evolution of RI therefore seems to be an important factor for understanding speciation in diatoms. Research efforts using both extant marine and freshwater species have added to our understanding of how environmental differences can give rise to RI. In addition to freshwater systems with well-documented environmental histories like Lake Baikal, another freshwater ecosystem might also provide a unique opportunity to study the role of climate change in driving speciation in microeukaryotes and test hypotheses about the precise climatic events—and their molecular consequences—that cause morphological change and RI in this group of organisms.

The Greater Yellowstone Ecosystem (GYE) spans the states of Idaho, Montana and Wyoming in the continental United States. Several large freshwater lakes exist within the GYE: Yellowstone Lake, Lewis Lake, Heart Lake, and Shoshone Lake are all located within Yellowstone National Park, and Jackson Lake is located within Grand Teton National Park (Figure 5). All five of these lakes are characterized by high concentrations of silica and phosphorus and lower concentrations of nitrogen (Kilham et al. 1996; Theriot et al. 1997). However, the lakes differ in the timing and amount of nitrogen limitation, and periodically experience silica and minor phosphorus limitation during the summer months.

Yellowstone Lake is located 2.35 km (7,733 feet) above sea level, and is the largest high-elevation lake in North America. The lake has a surface area of 342 km² and formed after the collapse of the Yellowstone Caldera eruption approximately 630 thousand years ago (kyr) during the Late Cenozoic (Smith and Siegel 2000). The main lake body is located on the eastern margin of the former caldera and the southern part of the lake lies outside the caldera. Because of its high elevation and the cooler climate during the Pleistocene, the caldera was covered under a 1 km (~0.62 miles) thick ice cap at the top of the Yellowstone Plateau that spanned an area of more than 160 km (~99 miles) north to south, and approximately 113 km (~70 miles) east to west (Pierce et al. 2007). Several glacial cycles followed the caldera's formation, in which ice formed during cooling dry periods, and melted during warmer and wetter periods, and ended with the last glacial retreat (Pinedale Glaciation) roughly 14-16 kyr (Smith and Siegel 2000; Pierce et al. 2007). These repeated glacial cycles deepened Yellowstone Lake and nearby Jackson Lake, and also gave rise to many of the smaller surrounding lakes that dot the GYE. Glacial melt waters that ran through newly exposed sediments and bedrock supplied the lakes with large amounts of dissolved nutrients and potentially resulted in a continuous mixing of the lake waters. These conditions were ideal for the colonization by microeukaryotes as well as other groups of organisms that occupied Yellowstone Lake at the end of the Pleistocene.

Stephanodiscus niagarae and *S. yellowstonensis* are two closely related centric diatom species that reside in the lakes of the GYE (Kilham et al. 1996) and present an attractive model to study speciation in freshwater diatoms based on their morphological (Theriot and Stoermer 1984; Theriot 1992), genetic (Zechman et al. 1994; Theriot et al. 2006; Alverson 2007; Alverson et al. 2007), and ecological histories. *S. niagarae* is a widely distributed planktonic species that occupies a broad range of habitats from oligotrophic to eutrophic freshwater. The species is found in North America from the states of Alaska to California and from California

to Pennsylvania, and up to the Canadian Shield (Theriot 1992). The lineage appears to be at least two million years old (Theriot et al. 1988a), and fossils of this species are found as far as east as China and as far south as Guatemala (Theriot et al. 1988a; Theriot et al. 2006). *S. yellowstonensis* has been found only in Yellowstone Lake, where it co-occurs with 7 of the 12 common diatom species present in all lakes within the GYE (*Aulacoseira subarctica*, *Asterionella formosa*, *Cyclotella bodanica*, *Diatoma elongatum* var. *tenue*, *Rhizosolenia eriensis*, *Stephanodiscus minutulus*, and *Synedra* sp.1; Interlandi et al. 1999). *S. niagarae* and *S. yellowstonensis* are distinguished by striking differences in frustule morphology: *S. yellowstonensis* possesses fewer spines on the valve face and fewer pores compared to *S. niagarae*. Ancient lake sediments from Yellowstone Lake show fossilized cells that bear the modern *S. yellowstonensis* morphology occur throughout the last 9,800 years (Theriot et al. 2006). Prior to this period in which *S. yellowstonensis* appears to be the dominant species, various intermediate valve morphologies occur between a *S. yellowstonensis*-like and a *S. niagarae*-like morphology. The present-day species occupy allopatric distributions in the modern-day GYE: *S. niagarae* is found in most lakes of the GYE (Lewis, Heart and Jackson) except for Yellowstone Lake, whereas *S. yellowstonensis* is found only in Yellowstone Lake.

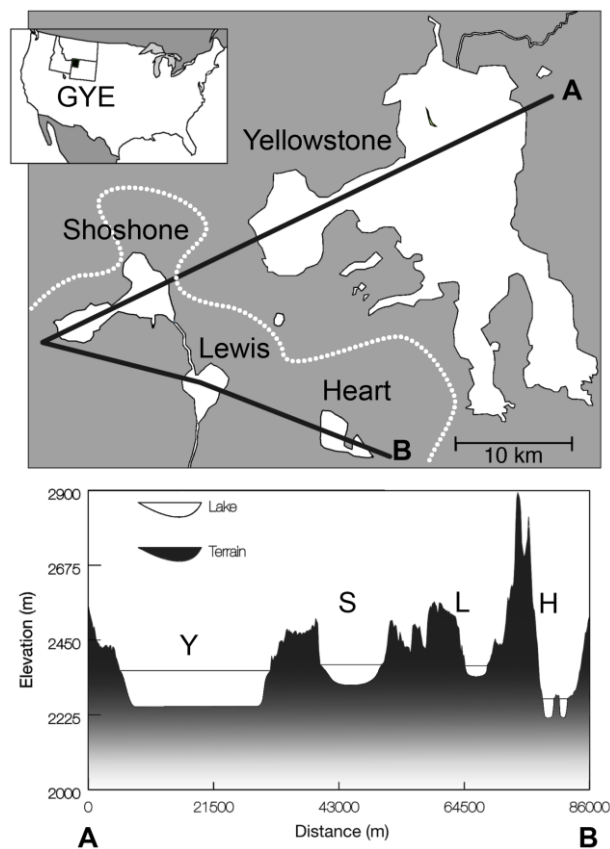


Figure 5. Map and hydrological profile of the Greater Yellowstone Ecosystem (GYE). The solid black line on the top map that extends from A to B corresponds to the vertical elevation shown on the bottom map. Lake depths and basin profiles were approximated using data from Theriot et al. (1997). The white dotted line on the top map marks the location of the Continental Divide. Y=Yellowstone Lake, S=Shoshone Lake, L=Lewis Lake, H=Heart Lake.

Paleontological History

The fossil record of the lake sediments dates back to around 14 kyr and shows that a *S. niagarae*-like ancestral species initially colonized the lakes of the GYE shortly after the glacial meltwaters filled the basin (Theriot et al. 2006). The fossil record also shows that its occupancy in Yellowstone Lake was short-lived (~13.7-12.2 kyr), and it was quickly replaced by *S. yellowstonensis*. The interval following colonization and before the disappearance of the *S. niagarae*-like ancestral species shows a transition between *S. niagarae* and *S. yellowstonensis* between 12.2-10.1 kyr, and shows a series of frustule morphologies characterized by a decreasing number of spines in the forms that precede modern *S. yellowstonensis* (Theriot et al. 2006). At 10.1 kyr, all cells in the record possess *S. yellowstonensis* valve morphology, coincident with a dramatically rapid increase in the size of the valve from a mean diameter of 40 μm to 64.3 μm . Valve diameter then steadily declines for the next 6,700 years until it returns to a mean of 41.3 μm around 3.4 kyr. Measurements of biogenic silica from the same core sediment (a rough indicator of diatom abundance) suggest a period of high productivity around 10.1 kyr. However, these data are difficult to interpret because strata of this age are marked not only by *S. yellowstonensis*, but also by *A. subarctica*, which possess more heavily silicified valves (Theriot et al. 2006).

The transition from *S. niagarae* to *S. yellowstonensis* thus occurs in a remarkably short period of approximately 1,500 years (Theriot et al. 2006). (It is worth mentioning that this also appears to be the one of the fastest examples of morphological character evolution in eukaryotes.) Theriot (1992) refers to this abrupt transition as a case of “time-transgressive morphological continuity” between fossils that appeared to be clearly *S. niagarae*-like and those that more closely resemble the extant *S. yellowstonensis*. Further studies show that *S. yellowstonensis* is not only morphologically divergent, but is also physiologically divergent: *S. yellowstonensis* can tolerate lower nitrogen levels than *S. niagarae* (Taylor 1994). Inorganic nitrogen is typically the growth-limiting nutrient in freshwaters, and diatom growth decreases substantially when nitrogen levels fall below the concentration of 10 μM (Interlandi et al. 1999). Diatoms in particular also become growth limited when silicon is low, because of the high silica requirement for frustule formation. *S. yellowstonensis* has a higher capacity for nitrogen utilization with an optimum intracellular ratio of silicon to nitrogen (Si:N) of 20:1, whereas *S. niagarae* maintains an optimum intracellular Si:N ratio of 2.5:1. *S. yellowstonensis* appears uniquely adapted to low nitrogen conditions; the species can survive and continue to grow under limited nitrogen conditions in culture, whereas low nitrogen conditions are lethal for *S. niagarae* (Kilham and Taylor 2002).

After the appearance of *S. yellowstonensis*, the core record shows the species was most abundant during two major historical droughts of the late 19th century and the Dust Bowl of the 1930's. Kilham et al. (1996) suggest that due to high light, low nitrogen to phosphorus ratio (N:P), and possibly low Si:P ratio caused by the decrease in precipitation, that these conditions may have shifted the competitive resource advantage to *S. yellowstonensis*, which appears to have experienced rapid population growth during these events. When conditions returned to wetter years with lower light, higher N:P ratio, and higher Si:P ratio due to external loading of nutrients, the coexisting diatom species *A. subarctica* began to dominate the plankton community as a presumed consequence of the shift in resource ratios. Although *A. subarctica* coexists with *S. yellowstonensis* in Yellowstone Lake, they are temporally isolated: *A. subarctica* blooms in early spring after ice-out when nutrients are high and light is low, whereas

S. yellowstonensis blooms in late summer when nutrients are lower and fewer diatom competitors are active (Kilham et al. 1996; Interlandi et al. 1999).

Events Leading to Speciation

The paleolimnological record of the lake suggests a sequence of climatic events that may have initiated the evolution of *S. yellowstonensis* and the subsequent morphological and physiological divergence between *S. niagarae* and *S. yellowstonensis*. The climate surrounding Yellowstone Lake during its formation was cool and wet, and the lake was either potentially continuously mixing or stratification was otherwise reduced— conditions that produced a turbid water column. The environmental change that occurred during the Younger Dryas Chronozone of North America had a dramatic impact on precipitation patterns of lake, which led to an alteration of the nutrient regime and changes in the mixing regime via thermal stratification of the epilimnion (upper most part of the water column). The Younger Dryas was a period of drying at the end of deglaciation from the Pleistocene, during which the climate shifted from cool and wet, to warm and dry. The rapid climate change altered the moisture balance of the region, and resulted in less runoff into the lake, which ultimately reduced the dissolved nutrient load (nitrogen, phosphorus, and silicon) being supplied from weathering bedrock. The Younger Dryas Chronozone spans ~12.9-11.7 kyr (Briles et al. 2012) and thus corresponds well with the timing of the rapid morphological changes observed between *S. niagarae* and *S. yellowstonensis* in the sediment transition zone.

Ecological studies of the modern Yellowstone Lake show that water chemistry is tightly linked to climatic conditions (Kilham et al. 1996). Increased cloud cover (decreasing light levels for photosynthesis) and increased precipitation (increasing nutrient influx to the lake) alter the abundance of the dominant phytoplankton in the lake. The consequences of these events are changes in lake dynamics toward different species composition with different optimal growth conditions (Kilham et al. 1996; Interlandi et al. 2003). The *S. niagarae*-like ancestor of *S. yellowstonensis* colonized the lake during a period when the lake was surrounded by sage and grass vegetation (13.7-12.0 kyr). Around 12.0 kyr the vegetation changed to a Lodgepole pine forest dominated by *Pinus contortus*, *Pinus albicaulis*, *Artemisia*, *Abies*, *Juniperus*, and *Picea* species, which marked expansion of the closed forest and shrinking of meadows that surrounded the lake. Warmer, drier summers also contributed to increased summer insolation of the lake and earlier ice-off of the lake following winter, which occurred 11.0-6.0 kyr. During warmer periods characterized by temperatures above 15 °C, Yellowstone Lake also becomes thermally unstable, which causes the thermocline (the water layer between the warm upper water column and cold lower water column) to move, whereas Lewis Lake and Jackson Lake, which *S. niagarae* inhabits, have more stable summer thermoclines (Interlandi et al. 1999). The most important consequence of these changes, was the decrease in nitrogen availability in the lake. It is during this period that the morphological transition between the two *Stephanodiscus* species is first observed in the sediments of Yellowstone Lake.

This raises the question of a possible connection between adaptation to a lower nitrogen environment and the change in valve morphology experienced by *S. yellowstonensis*. Transitional forms in the lake core records show what appear to be several clonal lineages altering their valve morphologies as lake stability and nitrogen concentrations decreased. Interestingly, differences in nitrogen abundance causes a change in valve morphology in a sister

species to *S. niagarae* and *S. yellowstonensis*, *S. alpinus* (Theriot et al. 1988b). Specifically, differences in the N:P ratio alters the valve costae:spine ratio in *S. alpinus*— high N:P ratio gives rise to lower costae:spine ratio and low N:P ratio gives rise to higher costae:spine ratio. Thus, as nitrogen level in Yellowstone Lake began to decrease, it seems possible that a coincident change in valve morphology may have given rise to morphological changes observed in *S. yellowstonensis*. Because *S. yellowstonensis* can survive nitrogen rich water, the possibility exists to directly test the consequence of this morphological change on RI. In particular, future research might test the hypothesis that morphological evolution drove the evolution of RI, by adjusting nitrogen levels to alter the valve morphology in *S. yellowstonensis* and measure its effect on RI between *S. yellowstonensis* and *S. niagarae*. The *S. niagarae*-*S. yellowstonensis* species pair is also a prime opportunity to bring new molecular technologies to bear on dissecting the genetic and molecular bases of RI in freshwater diatoms. Sequence data from whole genomes and whole transcriptomes of both species from all of the lakes in the GYE would facilitate identifying regions of the genome under selection, and narrow down potential candidate genes that contribute to morphological differences and RI between these two species.

CONCLUSION

Although the mechanisms of speciation in diatoms have received little attention over the past two centuries, recent work has initiated the requisite momentum to begin to thoroughly investigate processes of speciation in this large group of eukaryotes. Importantly, understanding the nature of environmental changes experienced by diatom communities is likely to prove crucial to identifying general mechanisms by which speciation occurs in diatoms.

One potentially informative avenue for future research efforts will be understanding the nature and the frequency of environmental stresses on the evolution of different types of prezygotic isolation between species. Although isolating mechanisms other than those discussed above are also likely to be at work in diatoms, detailed studies of their potential importance in diatom speciation have not yet been performed. Two barriers to gene flow in particular will likely provide fruitful avenues of future research. Diatoms occupy every global aquatic niche, which suggests that ecological isolation may play a predominant role in diatom speciation. Even among co-occurring taxa that reside within a single freshwater pond, adaptation to different niches presents the possibility of microallopatry among populations that occupy the epilimnion (top layer of benthic mud), epilithic (rocks), epiphyton (plant vegetation), and epizoic (animals) zones within the same body of water. Another mechanism of RI that could prove important is gametic isolation, particularly among those diatom species that produce motile gametes.

Perhaps most interesting is the general role that differences in cell morphology might play in the evolution of RI. Among some groups of diatoms, there is growing evidence that cell morphology appears to be important for successful sexual reproduction. If changes in frustule morphology are a common response to changes in the environment, understanding how the environment directs the evolution of morphology might provide great insight into mechanisms of speciation in the microeukaryota.

ACKNOWLEDGMENTS

We are grateful the editor for the invitation to contribute this chapter to the book. We thank D. Mann, T. Rynearson, and E. Theriot for helpful comments on the manuscript. This work was supported by funds from the University of Oklahoma.

REFERENCES

- Alverson AJ 2007. Strong Purifying Selection in the Silicon Transporters of Marine and Freshwater Diatoms. *Limnology and Oceanography* 52: 1420-1429.
- Alverson AJ, Cannone JJ, Gutell RR, Theriot EC 2006. The evolution of elongate shape in diatoms. *Journal of Phycology* 42: 655-668. doi: 10.1111/j.1529-8817.2006.00228.x.
- Alverson AJ, Jansen RK, Theriot EC 2007. Bridging the Rubicon: phylogenetic analysis reveals repeated colonizations of marine and fresh waters by thalassiosiroid diatoms. *Molecular Phylogenetics and Evolution* 45: 193-210. doi: 10.1016/j.ympev.2007.03.024.
- Amato A 2010. Species concepts and definitions: Reproductive isolation as a tool to reveal species boundaries. *International Journal of Plant Reproductive Biology* 2: 114-126.
- Amato A, Kooistra WHCF, Levialdi Ghiron JH, Mann DG, Pröschold T, Montresor M 2007. Reproductive Isolation among Sympatric Cryptic Species in Marine Diatoms. *Protist* 158: 193-207. doi: 10.1016/j.protis.2006.10.001.
- Armbrust EV 1999. Identification of a New Gene Family Expressed during the Onset of Sexual Reproduction in the Centric Diatom *Thalassiosira weissflogii*. *Applied and Environmental Microbiology* 65: 3121-3128.
- Armbrust EV 2009. The life of diatoms in the world's oceans. *Nature* 459: 185-192. doi: 10.1038/nature08057.
- Armbrust EV, Berges JA, Bowler C, Green BR, Martinez D, Putnam NH, Zhou S, Allen AE, Apt KE, Bechner M 2004. The Genome of the Diatom *Thalassiosira pseudonana*: Ecology, Evolution, and Metabolism. *Science* 306: 79-86.
- Armbrust EV, Galindo HM 2001. Rapid evolution of a sexual reproduction gene in centric diatoms of the genus *Thalassiosira*. *Applied and Environmental Microbiology* 67: 3501-3513. doi: 10.1128/AEM.67.8.3501-3513.2001.
- Behnke A, Friedl T, Chepurnov VA, Mann DG 2004. Reproductive compatibility and rDNA sequence analyses in the *Sellaphora pupula* species complex (Bacillariophyta). *Journal of Phycology* 40: 193-208.
- Belykh OI, Sorokovikova EG 2003. Autotrophic picoplankton in Lake Baikal: Abundance, dynamics and distribution. *Aquatic Ecosystems Health and Management* 6: 251-261.
- Bowler C, Allen AE, Badger JH, Grimwood J, Jabbari K, Kuo A, Maheswari U, Martens C, Maumus F, Otillar RP, et al. 2008. The *Phaeodactylum* genome reveals the evolutionary history of diatom genomes. *Nature* 456: 239-244. doi: 10.1038/nature07410.
- Briles CE, Whitlock C, Meltzer DJ 2012. Last glacial-interglacial environments in the southern Rocky Mountains, USA and implications for Younger Dryas-age human occupation. *Quaternary Research* 77: 96-103. doi: 10.1016/j.yqres. 2011.10.002.
- Chacon-Baca E, Beraldi-Campesi H, Cevallos-Ferriz SRS, Knoll AH, Golubic S 2002. 70 Ma nonmarine diatoms from northern Mexico. *Geology* 30: 279-281.

- Chepurnov VA, Roshchin AM 1995. Inbreeding influence on sexual reproduction of *Achnanthes longipes* Ag. (Bacillariophyta). *Diatom Research* 10: 21-25.
- Coleman AW 2009. Is there a molecular key to the level of "biological species" in eukaryotes? A DNA guide. *Molecular Phylogenetics and Evolution* 50: 197-203. doi: 10.1016/j.ympev.2008.10.008.
- Cox EJ 2006. *Achnanthes sensu stricto* belongs with genera of the Mastogloiales rather than with other monoraphid diatoms (Bacillariophyta). *European Journal of Phycology* 41: 67-81.
- Cox EJ 2010. Morphogenetic information and the selection of taxonomic characters for raphid diatom systematics. *Plant Ecology and Evolution* 143: 271-277. doi: 10.5091/plecevo.2010.403.
- Cox EJ 2011. Morphology, Cell Wall, Cytology, Ultrastructure and Morphogenetic Studies. In: Seckbach J, Kociolek P, editors. *The Diatom World*: Springer Netherlands. p. 21-45.
- Cox EJ, Williams D 2000. Systematics of naviculoid diatoms: the interrelationships of some taxa with a stauros. *European Journal of Phycology* 35: 273-282. doi: 10.1080/09670260010001735871.
- Coyne JA, Orr HA. 2004. *Speciation*. Sunderland, MA: Sinauer Associates.
- Damsté JSS, Muyzer G, Abbas B, Rampen SW, Massé G, Allard WG, Simon TB, Robert J-M, Rowland SJ, Moldowan JM, et al. 2004. The Rise of the Rhizosolenid Diatoms. *Science* 304: 584-587.
- Dixit SS, Smol JP, Kingston JC, Charles DF 1992. Diatoms: powerful indicators of environmental change. *Environmental Science & Technology* 26: 23-33.
- Dobzhansky T 1937. *Genetics and the Origin of Species*. New York: Columbia University Press.
- Dobzhansky T 1951. *Genetics and the Origin of Species*. New York: Columbia University Press.
- Drebes G 1977. Sexuality. In: Werner D, editor. *The biology of diatoms*. Berkeley: University of California Press. p. 250-283.
- Durkin CA, Mock T, Armbrust EV 2009. Chitin in Diatoms and Its Association with the Cell Wall. *Eukaryotic Cell* 8: 1038-1050. doi: 10.1128/ec.00079-09.
- Edgar LA, Pickett-Heaps JD 1983. The mechanism of diatom locomotion. I. An ultrastructural study of the motility apparatus. *Proceedings of the Royal Society London* 218: 331-343.
- Edlund MB, Soninkhishig N, Williams RM, Stoermer EF 2000. Morphology and taxonomy of *Eunotia clevei*. *Diatom Research* 15: 209-219.
- Edlund MB, Stoermer EF, Taylor CM 1996. *Aulacoseira skvortzowii* sp. nov (Bacillariophyta), a poorly understood diatom from Lake Baikal, Russia. *Journal of Phycology* 32: 165-175.
- Evans KM, Wortley AH, Mann DG 2007. An assessment of potential diatom "barcode" genes (cox1, rbcL, 18S and ITS rDNA) and their effectiveness in determining relationships in *Sellaphora* (Bacillariophyta). *Protist* 158: 349-364. doi: 10.1016/j.protis.2007.04.001.
- Evans KM, Wortley AH, Simpson GE, Chepurnov VA, Mann DG 2008. A molecular systematic approach to explore diversity within the *Sellaphora pupula* species complex (Bacillariophyta). *Journal of Phycology* 44: 215-231. doi: 10.1111/j.1529-8817.2007.00454.x.
- Falkowski PG, Katz ME, Knoll AH, Quigg A, Raven JA, Schofield O, Taylor FJR 2004. The Evolution of Modern Eukaryotic Phytoplankton. *Science* 305: 354-360.

- Field CB, Behrenfeld MJ, Randerson JT, Falkowski P 1998. Primary Production of the Biosphere: Integrating Terrestrial and Oceanic Components. *Science* 281: 237-240. doi: 10.1126/science.281.5374.237.
- Fietz S, Kobanova G, Izmet'seva L, Nicklisch A 2005. Regional, vertical and seasonal distribution of phytoplankton and photosynthetic pigments in Lake Baikal. *Journal of Plankton Research* 27: 793-810. doi: 10.1093/plankt/fbi054.
- Finch-Savage WE, Leubner-Metzger G 2006. Seed dormancy and the control of germination. *New Phytologist* 171: 501-523. doi: 10.1111/j.1469-8137.2006.01787.x.
- Finkel ZV, Matheson KA, Regan KS, Irwin AJ 2010. Genotypic and phenotypic variation in diatom silicification under paleo-oceanographic conditions. *Geobiology* 8: 433-445. doi: 10.1111/j.1472-4669.2010.00250.x.
- Geitler L 1935. Reproduction and Life History in Diatoms. *Botanical Review* 1: 149-161.
- Geitler L 1958a. Fortpflanzungsbiologische Eigentümlichkeiten von *Cocconeis* und Vorarbeiten zu einer systematischen Gliederung von *Cocconeis placentula* nebst Beobachtungen an Bastarden. [Biology of *Cocconeis* and preliminary work on a systematic classification of *Cocconeis placentula* together with observations on bastards]. *Osterreichischen Botanischen Zeitschrift* 105: 350-379.
- Geitler L 1958b. Notizen über Rassenbildung, Fortpflanzung, Formwechsel und morphologische Eigentümlichkeiten bei pennaten Diatomeen. [Notes on race formation, reproduction, change of form and morphological features in pennate diatoms]. *Osterreichischen Botanischen Zeitschrift* 105: 408-442.
- Geitler L 1975. Formwechsel, sippenspezifischer Paarungsmodus und Systematik bei einigen pennaten Diatomeen. [Change of shape, sip-specific mating mode and systematics of some pennate diatoms]. *Plant Systematics and Evolution* 124: 7-30.
- Gersonde R, Harwood DM 1990. Lower Cretaceous diatoms from ODP Leg 113 Site 693 Weddell Sea. Part 1: Vegetative cells. *Proceedings of the Ocean Drilling Program, Scientific Results* 113: 365-402.
- Godhe A, McQuoid MR, Karunasagar I, Karunasagar I, Rehnstam-Holm A-S 2006. Comparison of three common molecular tools for distinguishing among geographically separated clones of the diatom *Skeletonema marinoi* et Zingone (Bacillariophyceae). *Journal of Phycology* 42: 280-291. doi: 10.1111/j.1529-8817.2006.00197.x.
- Good JM, Dean MD, Nachman MW 2008. A complex genetic basis to X-linked hybrid male sterility between two species of house mice. *Genetics* 179: 2213-2228. doi: genetics.107.085340/10.1534/genetics.107.085340.
- AlgaeBase*. World-wide electronic publication, National University of Ireland, Galway. <http://www.algaebase.org>; searched on 28 May 2012. [Internet]. 2012 searched on 28 May 2012.]. Available from: <http://www.algaebase.org>.
- Hargraves PE, French FW. 1983. Diatom resting spores: significance and strategies. In: Fryxell GA, editor. *Survival Strategies of the Algae*. New York: Cambridge University Press. p. 49-68.
- Härnström K, Ellegaard M, Andersen TJ, Godhe A 2011. Hundred years of genetic structure in a sediment revived diatom population. *Proceedings of the National Academy of Sciences* 108: 4252-4257. doi: 10.1073/pnas.1013528108.
- Harper MA 1976 Migration rhythm of the benthic diatom *Pinnularia viridis* on pond silt. *New Zealand Journal of Marine and Freshwater Research* 10: 381-384.

- Harwood DM, Chang KH, Nikolaev VA 2004. Late Jurassic to earliest Cretaceous diatoms from Jasong Synthem, Southern Korea: evidence for a terrestrial origin. In: A. Witkowski, Radziejewska T, Wawrzyniak-Wydrowska B, Daniszewska-Kowalczyk G, Bak M, editors. *Abstracts, 18th International Diatom Symposium*, Miedzyzdroje, Poland, p. 81.
- Harwood DM, Nikolaev VA 1990. Cretaceous diatoms: morphology, taxonomy, biostratigraphy. In: Blome CD, Whalen PM, Reed KM, editors. *Siliceous Microfossils: Paleontological Society Short Course in Paleontology*.
- Hey J 2001. *Genes Categories and Species*. New York, NY: Oxford University Press.
- Hoagland KD, Rosowski JR, Gretz MR, Roemer SC 1993. Diatom extracellular polymeric substances: function, fine structure, chemistry and physiology. *Journal of Phycology* 29: 537-566. doi: 10.1111/j.0022-3646.1993.00537.x.
- Interlandi SJ, Kilham SS, Theriot EC 1999. Responses of phytoplankton to varied resource availability in large lakes of the Greater Yellowstone Ecosystem. *Limnology and Oceanography* 44: 668-682.
- Interlandi SJ, Kilham SS, Theriot EC 2003. Diatom-Chemistry Relationships in Yellowstone Lake (Wyoming) Sediments: Implications for Climatic and Aquatic Processes Research. *Limnology and Oceanography* 48: 79-92.
- Jewson DH, Granin NG, Zhdanov AA, Gorbunova LA, Bondarenko NA, Gnatovsky RY 2008a. Resting stages and ecology of the planktonic diatom *Aulacoseira skortzowii* in Lake Baikal. *Limnology and Oceanography* 53: 1125-1136.
- Jewson DH, Granin NG, Zhdanov AA, Gorbunova LA, Gnatovsky RY 2010. Vertical mixing, size change and resting stage formation of the planktonic diatom *Aulacoseira baicalensis*. *European Journal of Phycology* 45: 354-364. doi: 10.1080/09670262.2010.492915.
- Karabanov E, Williams D, Kuzmin M, Sideleva V, Khursevich G, Prokopenko A, Solotchina E, Tkachenko L, Fedenya S, Kerber E, et al. 2004. Ecological collapse of Lake Baikal and Lake Hovsgol ecosystems during the Last Glacial and consequences for aquatic species diversity. *Palaeogeography, Palaeoclimatology, Palaeoecology* 209: 227-243. doi: 10.1016/j.palaeo.2004.02.017.
- Katz LA, Grant JR, Parfrey LW, Burleigh JG 2012. Turning the Crown Upside Down: Gene Tree Parsimony Roots the Eukaryotic Tree of Life. *Systematic Biology*. doi: 10.1093/sysbio/sys026.
- Keeling PJ, Burger G, Durnford DG, Lang BF, Lee RW, Pearlman RE, Roger AJ, Gray MW 2005. The tree of eukaryotes. *Trends in Ecology Evolution* 20: 670-676. doi: 10.1016/j.tree.2005.09.005.
- Khursevich GK, Karabanov EB, Prokopenko AA, Williams DF, Kuzmin MI, Fedenya SA, Gvozdkov AA 2001. Insolation regime in Siberia as a major factor controlling diatom production in Lake Baikal during the past 800,000 years. *Quaternary International* 80-81: 47-58. doi: 10.1016/S1040-6182(01)00018-0.
- Khursevich GK, Karabanov EB, Williams DF, Kuzmin MI, Prokopenko AA. 2000. Evolution of freshwater centric diatoms within the Baikal rift zone during the Cenozoic. In: Minoura K, editor. *Lake Baikal. A Mirror in Time and Space for Understanding Global Change Processes*. Amsterdam: Elsevier. p. 146-154.
- Khursevich GK, Kociolek JP, Fedenya SA 2002. A new genus of fossil freshwater diatoms (Bacillariophyta: Stephanodiscaceae) from the sediments of Lake Baikal. *Proceedings of the California Academy of Sciences* 53: 1-10.

- Khursevich GK, Kociolek JP, Iwashita T, Fedenya SA, Kuzmin MI, Kawai T, Williams DF, Karabanov EB, Prokopenko AA, Minoura K 2004. Mesodictyopsis Khursevich, Iwashita, Kociolek and Fedenya A New Genus of Diatoms in the Class Coscinodiscophyceae (Bacillariophyta) from Upper Miocene Sediments of Lake Baikal, Siberia. *Proceedings of the California Academy of Science* 55: 338-357.
- Khursevich GK, Prokopenko AA, Fedenya SA, Tkachenko LI, Williams DF 2005. Diatom biostratigraphy of Lake Baikal during the past 1.25 Ma: new results from BDP-96-2 and BDP-99 drill cores. *Quaternary International* 136: 95-104. doi: 10.1016/j.quaint.2004.11.011.
- Kilham SS, Taylor SM 2002. Resource requirements and competitive interactions of two closely related *Stephanodiscus* species for silicon and nitrogen. *Verhandlungen Internationale Vereinigung für theoretische und angewandte Limnologie* 28: 928-931.
- Kilham SS, Theriot EC, Fritz SC 1996. Linking Planktonic Diatoms and Climate Change in the Large Lakes of the Yellowstone Ecosystem Using Resource Theory. *Limnology and Oceanography* 41: 1052-1062.
- Kingston MB 1999. Wave Effects on the Vertical Migration of Two Benthic Microalgae: *Hantzschia virgata* var. *intermedia* and *Euglena proxima*. *Estuaries* 22: 81-91.
- Kociolek JP, Stoermer EF 2010. Variation and polymorphism in diatoms: the triple helix of development, genetics and environment. A review of the literature. *Vie et Milieu - Life and Environment* 60: 75-87.
- Koester J, Swalwell J, von Dassow P, Armbrust EV 2010. Genome size differentiates co-occurring populations of the planktonic diatom *Ditylum brightwellii* (Bacillariophyta). *BMC Evolutionary Biology* 10: 1.
- Koizumi I, Yanagisawa Y 1990. Evolutionary change in diatom morphology -- an example from *Nitzschia fossilis* to *Pseudoeunotia doliolus*. *Transactions and Proceedings of the Palaeontological Society of Japan* 157.
- Kooistra WHCF, Medlin LK 1996. Evolution of the diatoms (Bacillariophyta). IV. Reconstruction of their age from small subunit rRNA coding regions and the fossil record. *Molecular Phylogenetics and Evolution* 6: 391-407.
- Kozłowska JL, Ahmad AR, Jahesh E, Cutter AD 2012. Genetic variation for postzygotic reproductive isolation between *Caenorhabditis briggsae* and *Caenorhabditis* sp. 9. *Evolution* 66: 1180-1195. doi: 10.1111/j.1558-5646.2011.01514.x.
- Kröger N, Poulsen N 2008. Diatoms--From Cell Wall Biogenesis to Nanotechnology. *Annual Review of Genetics* 42: 83-107. doi: doi:10.1146/annurev.genet.41.110306.130109.
- Lachance J, True JR 2010. X-autosome incompatibilities in *Drosophila melanogaster*: Tests of Haldane's rule and geographic patterns within species. *Evolution* 64: 3035-3046. doi: 10.1111/j.1558-5646.2010.01028.x.
- Lewis WM 1984. The Diatom Sex Clock and Its Evolutionary Significance. *The American Naturalist* 123: 73-80.
- Li CW, Chu S, Lee M 1989. Characterizing the silica deposition vesicle of diatoms. *Protoplasma* 151: 158-163.
- Macdonald JD 1869. I. On the structure of the diatomaceous frustule, and its genetic cycle. *The Annals and Magazine of Natural History Series 4* 3: 1-8. doi: 10.1080/00222936908695866.
- Mackay AW 2007. The paleoclimatology of Lake Baikal: A diatom synthesis and prospectus. *Earth-Science Reviews* 82: 181-215. doi: 10.1016/j.earscirev.2007.03.002.

- Mackay AW, Flower RJ, Kuzmina AE, Granina LZ, Rose NL, Appleby PG, Boyle JF, Battarbee RW 1998. Diatom Succession Trends in Recent Sediments from Lake Baikal and Their Relation to Atmospheric Pollution and to Climate Change. *Philosophical Transactions: Biological Sciences* 353: 1011-1055.
- Mann DG 1981. Studies in the Diatom Genus *Hantzschia* 3. Intraspecific Variation in *H. virgata*. *Annals of Botany* 47: 377-395.
- Mann DG 1984. Observations on Copulation in *Navicula pupula* and *Amphora ovalis* in Relation to the Nature of Diatom Species. *Annals of Botany* 54: 429-438.
- Mann DG 1989. The species concept in diatoms: evidence for morphologically distinct, sympatric gamodemes in four epipelagic species. *Plant Systematics and Evolution* 164: 215-237.
- Mann DG 1994. Auxospore formation, reproductive plasticity and cell structure in *Navicula ulvacea* and the resurrection of the genus *Dickieia* (Bacillariophyta). *European Journal of Phycology* 29: 141-157. doi: 10.1080/09670269400650591.
- Mann DG 1999. Species concepts in diatoms. *Phycologia* 38: 437-495.
- Mann DG, Chepurnov VA 2005. Auxosporulation, mating system, and reproductive isolation in *Neidium* (Bacillariophyta). *Phycologia* 44: 335-350. doi: 10.2216/0031-8884(2005)44[335:amsari]2.0.co;2.
- Mann DG, Chepurnov VA, Droop SJM 1999. Sexuality, incompatibility, size variation, and preferential polyandry in natural populations and clones of *Sellaphora pupula* (Bacillariophyceae). *Journal of Phycology* 35: 152-170. doi: 10.1046/j.1529-8817.1999.3510152.x.
- Mann DG, Droop SJM 1996. Biodiversity, biogeography and conservation of diatoms. *Hydrobiologia* 336: 19-32.
- Mann DG, McDonald SM, Bayer MM, Droop SJM, Chepurnov VA, Loke RE, Ciobanu A, Hans due Buf JM 2004. The *Sellaphora pupula* species complex (Bacillariophyceae): morphometric analysis, ultrastructure and mating data provide evidence for five new species. *Phycologia* 43: 459-482.
- Mann DG, Stickle AJ 1991. The genus *Craticula*. *Diatom Research* 6: 79-107.
- Martin NH, Willis JH 2010. Geographical variation in postzygotic isolation and its genetic basis within and between two *Mimulus* species. *Philosophical Transactions of the Royal Society B: Biological Sciences* 365: 2469-2478. doi: 10.1098/rstb.2010.0030.
- Martin-Jézéquel V, Hildebrand M, Brzezinski MA 2000. Silicon metabolism in diatoms: implications for growth. *Journal of Phycology* 36: 821-840.
- Mayr E 1942. *Systematics and the Origin of Species*. New York: Columbia University Press.
- Mayr E 1963. *Animal Species and Evolution*. Cambridge, Mass.: The Belknap Press of Harvard University Press.
- McBride SA, Edgar RK 1998. Janus cells unveiled: frustular morphometric variability in *Gomphonema angustatum*. *Diatom Research* 13: 293-310. doi: 10.1080/0269249x.1998.9705452.
- McConville MJ, Wetherbee R, Bacic A 1999. Subcellular location and composition of the wall and secreted extracellular sulphated polysaccharides/proteoglycans of the diatom *Stauroneis amphioxys* Gregory. *Protoplasma* 206: 188-200. doi: 10.1007/bf01279266.
- McQuoid MR, Hobson LA 1996. Diatom resting stages. *Journal of Phycology* 32: 889.

- Medlin L, Elwood HJ, Stickel S, Sogin ML 1988. The characterization of enzymatically amplified eukaryotic 16S-like rRNA-coding regions. *Gene* 71: 491-499. doi: 10.1016/0378-1119(88)90066-2.
- Medlin LK, Kaczmarek I 2004. Evolution of the diatoms: V. Morphological and cytological support for the major clades and a taxonomic revision. *Phycologia* 43: 245-270.
- Medlin LK, Kooistra WHCF, Gersonde R, Sims PA, Wellbrock U 1997. Is the origin of the diatoms related to the end-Permian mass extinction? *Nova Hedwigia* 65: 1-11.
- Muller HJ 1942. Isolating mechanisms, evolution, and temperature. *Biological Symposia* 6: 71-125.
- Nagata T, Takai K, Kawanobe K, Kim D-S, Nakazato R, Guselnikova N, Bondarenko N, Mologawaya O, Kostrova T, Drucker V, et al. 1994. Autotrophic picoplankton in southern Lake Baikal: abundance, growth and grazing mortality during summer. *Journal of Plankton Research* 16: 945-959. doi: 10.1093/plankt/16.8.945.
- Nelson DM, Treguer P, Brzezinski MA, Leynaert A, Queguiner B 1995. Production and dissolution of biogenic silica in the ocean: Revised global estimates, comparison with regional data and relationship to biogenic sedimentation. *Global Biogeochemical Cycle* 9: 359-372. doi: 10.1029/95gb01070.
- Nolte AW, Gompert Z, Buerkle CA 2009. Variable patterns of introgression in two sculpin hybrid zones suggest that genomic isolation differs among populations. *Molecular Ecology* 18: 2615-2627. doi: 10.1111/j.1365-294X.2009.04208.x.
- Pappas JL, Stoermer EF 2001. *Asterionella* Hassall (Heterokontophyta, Bacillariophyceae) Taxonomic history and quantitative methods as an aid to valve shape differentiation. *Diatom* 17: 47-58.
- Parfrey LW, Grant J, Tekle YI, Lasek-Nesselquist E, Morrison HG, Sogin ML, Patterson DJ, Katz LA 2010. Broadly Sampled Multigene Analyses Yield a Well-Resolved Eukaryotic Tree of Life. *Systematic Biology* 59: 518-533. doi: 10.1093/sysbio/syq037.
- Pfitzer E 1869. Über den Bau und die Zellteilung der Diatomeen. [About the construction and cell division of diatoms]. *Botanische Zeitung* 27: 774-776.
- Pfitzer E 1871. Untersuchungen über Bau und Entwicklung der Bacillariaceen (Diatomaceen). [Studies on the construction and development of Bacillariaceen (diatomaceous plants)]. *Botanische Abhandlungen* 2: 1-189.
- Pierce KL, Cannon KP, Meyer GA, Trebesch MJ, Watts RD 2007. Postglacial inflation-deflation cycles, tilting, and faulting in the Yellowstone Caldera based on Yellowstone Lake shorelines. In: Morgan LA, editor. *Integrated Geoscience Studies in the Greater Yellowstone Area—Volcanic, Tectonic, and Hydrothermal Processes in the Yellowstone Geocosystem: U.S. Geological Survey Professional Paper* 1717, 532 p. p. 128-168.
- Potapova M 2011. Patterns of Diatom Distribution in Relation to Salinity. In: Seckbach J, Kociolek P, editors. *The Diatom World*: Springer Netherlands. p. 313-332.
- Potapova M, Charles DF 2007. Diatom metrics for monitoring eutrophication in rivers of the United States. *Ecological Indicators* 7: 48-70. doi: 10.1016/j.ecolind.2005.10.001.
- Potapova M, Charles DF 2003. Distribution of benthic diatoms in U.S. rivers in relation to conductivity and ionic composition. *Freshwater Biology* 48: 1311-1328.
- Potapova M, Snoeijis P 1997. The natural life cycle in wild populations of *Diatoma moniliformis* (Bacillariophyceae) and its disruption in an aberrant environment. *Journal of Phycology* 33: 924-937. doi: 10.1111/j.0022-3646.1997.00924.x.

- Potapova MG, Charles DF 2002. Benthic diatoms in USA rivers: distributions along spatial and environmental gradients. *Journal of Biogeography* 29: 167-187.
- Poulsen NC, Spector I, Spurck TP, Schultz TF, Wetherbee R 1999. Diatom gliding is the result of an actin-myosin motility system. *Cell Motility and the Cytoskeleton* 44: 23-33.
- Rabosky DL, Sorhannus U 2009. Diversity dynamics of marine planktonic diatoms across the Cenozoic. *Nature* 457: 183-186. doi: 10.1038/nature07435.
- Reed LK, LaFlamme BA, Markow TA 2008. Genetic Architecture of Hybrid Male Sterility in *Drosophila*: Analysis of Intraspecific Variation for Interspecific Isolation. *PLoS ONE* 3: e3076. doi: 10.1371/journal.pone.0003076.
- Rothpletz A 1896. Über die Flysch-Fucoiden und einige andere fossile Algen, sowie über liasische, Diatomeen führende Hornschwämme. [About flysch fucoids and some other fossil algae, as well as about liasic, diatomaceous horn sponges]. *Zeitschrift der Deutschen Geologischen Gesellschaft* 48: 854-914.
- Round FE, Crawford RM, Mann DG. 1990. *The Diatoms: Morphology and Biology of the Genera*. Cambridge: Cambridge University Press.
- Rynearson TA, Newton JA, Armbrust EV 2006. Spring bloom development, genetic variation, and population succession in the planktonic diatom *Ditylum brightwellii*. *Limnology and Oceanography* 51: 1249-1261.
- Sarmiento JL, Gruber N, Brzezinski MA, Dunne JP 2004. High-latitude controls of thermocline nutrients and low latitude biological productivity. *Nature* 427: 56-60.
- Saros JE, Fritz SC 2000. Changes in the growth rates of saline-lake diatoms in response to variation in salinity, brine type and nitrogen form. *Journal of Plankton Research* 22: 1071-1083. doi: 10.1093/plankt/22.6.1071.
- Sato S, Beakes G, Idei M, Nagumo T, Mann DG 2011. Novel Sex Cells and Evidence for Sex Pheromones in Diatoms. *PLoS ONE* 6: e26923.
- Schultz ME 1971. Salinity-related polymorphism in the brackish-water diatom *Cyclotella cryptica*. *Canadian Journal of Botany* 49: 1285-1289.
- Shimada C, Burckle LH, Tanimura Y 2003. Morphological variability in *Neodenticula seminae*, a marine planktonic diatom in the North Pacific Ocean and Bering Sea. *Diatom Research* 18: 307-322. doi: 10.1080/0269249x.2003.9705594.
- Sicko-Goad L, Andresen NA 1991. Effect of growth and light/dark cycles on diatom lipid content and composition. *Journal of Phycology* 27: 710-718.
- Sicko-Goad L, Stoermer EF, Fahnstiel G 1986. Rejuvenation of *Melosira granulata* (Bacillariophyceae) Resting Cells from the Anoxic Sediments of Douglas Lake, Michigan. I. Light Microscopy and ¹⁴C Uptake. *Journal of Phycology* 22: 22-28.
- Index Nominum Algarum, University Herbarium, University of California, Berkeley. Compiled by Paul Silva. Available online at <http://ucjeps.berkeley.edu/INA.html> accessed May 2012 [Internet]. 2012.
- Sims PA, Mann DG, Medlin LK 2006. Evolution of the diatoms: insights from fossil, biological and molecular data. *Phycologia* 45: 361-402. doi: 10.2216/05-22.1.
- Singh RS, Stoermer EF, Kar R 2006. Earliest Freshwater Diatoms from the Deccan Intertrappean (Maastrichtian) Sediments of India. *Micropaleontology* 52: 545-551.
- Smith RB, Siegel LJ. 2000. *Windows into the earth: the geologic story of Yellowstone and Grand Teton national parks*. New York: Oxford University Press, USA.
- Smol JP, Battarbee RW, Davis RB, Meriläinen J 1986. Diatoms and lake acidity reconstructing pH from siliceous algal remains in lake sediments. In. *Developments in Hydrobiology*.

- Dordrecht, Netherlands; Boston; Hingham, MA, USA: W. Junk; Distributors for the U.S. and Canada: Kluwer Academic Publishers. p. 307.
- Smol JP, Stoermer EF 2010. *The diatoms: applications for the environmental and earth sciences*: Cambridge Univ Press.
- Soltis DE, Albert VA, Leebens-Mack J, Bell CD, Paterson AH, Zheng C, Sankoff D, dePamphilis CW, Wall PK, Soltis PS 2009. Polyploidy and angiosperm diversification. *American Journal of Botany* 96: 336-348. doi: 10.3732/ajb.0800079.
- Sorhannus U 1990. Punctuated morphological change in a Neogene diatom lineage: "Local" evolution or migration? *Historical Biology* 3: 241-247. doi: 10.1080/08912969009386524.
- Sorhannus U, Fenster EJ, Burckle LH, Hoffman A 1988. Cladogenetic and anagenetic changes in the morphology of *Rhizosolenia praebergonii* Mukhina. *Historical Biology* 1: 185-205. doi: 10.1080/08912968809386474.
- Sorhannus U, Kosakovsky P, Pond SL 2006. Evidence for positive selection on a sexual reproduction gene in the diatom genus *Thalassiosira* (Bacillariophyta). *Journal of Molecular Evolution* 63: 231-239.
- Sorhannus U, Fenster EJ, Hoffman A, Burckle L 1991. Iterative evolution in the diatom genus *Rhizosolenia* Ehrenberg. *Lethaia* 24: 39-44. doi: 10.1111/j.1502-3931.1991.tb01178.x.
- Staley JP, Woolford Jr JL 2009. Assembly of ribosomes and spliceosomes: complex ribonucleoprotein machines. *Current Opinion in Cell Biology* 21: 109-118. doi: 10.1016/j.ceb.2009.01.003.
- Stevenson RJ, Pan Y 1999. Assessing environmental conditions in rivers and streams with diatoms. In: Stoermer EF, Smol J, editors. *The Diatoms: Applications for the Environmental and Earth Sciences*. New York: Cambridge University Press.
- Stoermer EF 1967. Polymorphism in *Mastogloia*. *Journal of Phycology* 3: 73-77. doi: 10.1111/j.1529-8817.1967.tb04633.x.
- Sutherland D 2008. Surface-associated diatoms from marine habitats at Cape Evans, Antarctica, including the first record of living *Eunotogramma marginopunctatum*. *Polar Biology* 31: 879-888. doi: 10.1007/s00300-008-0426-z.
- Suzuki Y, Nei M 2004. False-positive selection identified by ML-based methods: examples from the Sig1 gene of the diatom *Thalassiosira weissflogii* and the tax gene of a human T-cell lymphotropic virus. *Molecular Biology and Evolution* 21: 914-921.
- Sweigart AL, Mason AR, Willis JH 2007. Natural variation for a hybrid incompatibility between two species of *Mimulus*. *Evolution* 61: 141-151. doi: EVO011/10.1111/j.1558-5646.2007.00011.x.
- Takahashi A, Tsaur S-C, Coyne JA, Wu C-I 2001. The nucleotide changes governing cuticular hydrocarbon variation and their evolution in *Drosophila melanogaster*. *Proceedings of the National Academy of Sciences* 98: 3920-3925. doi: 10.1073/pnas.061465098.
- Tapia PM, Harwood DM 2002. Upper Cretaceous Diatom Biostratigraphy of the Arctic Archipelago and Northern Continental Margin, Canada. *Micropaleontology* 48: 303-342.
- Taylor SM 1994. *Resource requirements and competitive interactions of two closely related Stephanodiscus species for nitrogen and silica*. [M.S. thesis]: Drexel University.
- Teeter KC, Thibodeau LM, Gompert Z, Buerkle CA, Nachman MW, Tucker PK 2010. The variable genomic architecture of isolation between hybridizing species of house mice. *Evolution* 64: 472-485. doi: 10.1111/j.1558-5646.2009.00846.x.
- Theriot E 1992. Clusters, species concepts, and morphological evolution of diatoms. *Systematic Biology* 41: 141-157.

- Theriot E, Qi Y, Yang J, Ling L 1988a. Taxonomy of the diatom *Stephanodiscus niagarae* from a fossil deposit in Jingyu County, Jilin Province, China. *Diatom Research* 3: 159-167.
- Theriot EC 1987. Principal component analysis and taxonomic interpretation of environmentally related variation in silicification in *Stephanodiscus* (Bacillariophyceae). *European Journal of Phycology* 22: 359-373. doi: 10.1080/00071618700650431.
- Theriot EC 2010. A preliminary multigene phylogeny of the diatoms (Bacillariophyta): challenges for future research. *Plant Ecology and Evolution* 143: 278-296. doi: 10.5091/plecevo.2010.418.
- Theriot EC, Fritz SC, Gresswell RE 1997. Long-Term Limnological Data from the Larger Lakes of Yellowstone National Park, Wyoming, U.S.A. *Arctic and Alpine Research* 29: 304-314.
- Theriot EC, Fritz SC, Whitlock C, Conley DJ 2006. Late Quaternary rapid morphological evolution of an endemic diatom in Yellowstone Lake, Wyoming. *Paleobiology* 32: 38-54. doi: 10.1666/02075.1.
- Theriot EC, Håkansson H, Stoermer EF 1988b. Morphometric analysis of *Stephanodiscus alpinus* (Bacillariophyceae) and its morphology as an indicator of lake trophic status. *Phycologia* 27: 485-493.
- Theriot EC, Stoermer EF 1984. Principal component analysis of *Stephanodiscus*: Observations on two new species from the *Stephanodiscus niagarae* complex. *Bacillaria* 7: 37-53.
- Tirichine L, Bowler C 2011. Decoding algal genomes: tracing back the history of photosynthetic life on Earth. *The Plant Journal* 66: 45-57. doi: 10.1111/j.1365-313X.2011.04540.x.
- Trobajo R, Mann DG, Chepurnov VA, Clavero E, Cox EJ 2006. Taxonomy, Life Cycle, and Auxosporulation of *Nitzschia fonticola* (Bacillariophyta). *Journal of Phycology* 42: 1353-1372. doi: 10.1111/j.1529-8817.2006.00291.x.
- Trobajo R, Mann DG, Clavero E, Evans KM, Vanormelingen P, McGregor RC 2010. The use of partial *cox1*, *rbcL* and *LSU rDNA* sequences for phylogenetics and species identification within the *Nitzschia palea* species complex (Bacillariophyceae). *European Journal of Phycology* 45: 413-425. doi: 10.1080/09670262.2010.498586.
- Trobajo R, Rovira L, Mann DG, Cox EJ 2011. Effects of salinity on growth and on valve morphology of five estuarine diatoms. *Phycological Research* 59: 83-90. doi: 10.1111/j.1440-1835.2010.00603.x.
- Tschochner H, Hurt E 2003. Pre-ribosomes on the road from the nucleolus to the cytoplasm. *Trends in Cell Biology* 13: 255-263. doi: 10.1016/s0962-8924(03) 00054-0.
- Tuchman ML, Theriot EC, Stoermer EF 1984. Effects of low-level salinity concentrations on the growth of *Cyclotella meneghiniana*. *Kütz Archiv für Protistenkunde* 128: 319-326.
- Vacquier VD, Swanson WJ 2011. Selection in the Rapid Evolution of Gamete Recognition Proteins in Marine Invertebrates. *Cold Spring Harbor Perspectives in Biology* 3: a002931. doi: 10.1101/cshperspect.a002931.
- Van Heurck H 1896. *A treatise on the diatomaceae containing introductory remarks on the structure life history, collection, cultivation and preparation of diatoms*. London: Wesley & Son.
- Vanormelingen P, Chepurnov VA, Mann DG, Sabbe K, Vyverman W 2008. Genetic Divergence and Reproductive Barriers among Morphologically Heterogeneous Sympatric Clones of *Eunotia bilunaris sensu lato* (Bacillariophyta). *Protist* 159: 73-90. doi: 10.1016/j.protis.2007.08.004.

- Vrieling EG, Gieskes WWC, Beelen TPM 1999. Silicon deposition in diatoms: Control by the pH inside the silicon deposition vesicle. *Journal of Phycology* 35: 548-559. doi: 10.1046/j.1529-8817.1999.3530548.x.
- Vyskočilová M, Pražanová G, Piálek J 2009. Polymorphism in hybrid male sterility in wild-derived *Mus musculus musculus* strains on proximal chromosome 17. *Mammalian Genome* 20: 83-91. doi: 10.1007/s00335-008-9164-3.
- Vyverman W, Sabbe K, Mann D, Kilroy C, Vyverman R, Vanhoutte K, Hodgson D 1998. *Eunophora* gen. nov. (Bacillariophyta) from Tasmania and New Zealand: description and comparison with Eunotia and amphoroid diatoms. *European Journal of Phycology* 33: 95-111. doi: 10.1080/09670269810001736593.
- Williams DM, Reid G 2006. *Amphorotia* nov. gen., a new genus in the family Eunotiaceae (Bacillariophyceae), based on *Eunotia clevei* Grunow in Cleve et Grunow. *Diatom Monographs* 6.
- Winsborough BM, Theriot E, Czarnecki DB 2009. Diatoms on a continental "island": Lazarus species, marine disjuncts and other endemic diatoms of the Cuatro Ciénegas basin, Coahuila, México. *Nova Hedwigia, Suppl.* 135: 257-274.
- Witkowski J, Sims PA, Harwood DM 2011. Rutilariaceae redefined: a review of fossil bipolar diatom genera with centrally positioned linking structures, with implications for the origin of pennate diatoms. *European Journal of Phycology* 46: 378-398. doi: 10.1080/09670262.2011.629685.
- Yukilevich R, True JR 2008. Incipient sexual isolation among cosmopolitan *Drosophila melanogaster* populations. *Evolution* 62: 2112-2121. doi: 10.1111/j.1558-5646.2008.00427.x.
- Zechman FW, Zimmer EA, Theriot EC 1994. Use of Ribosomal DNA Internal Transcribed Spaces for Phylogenetic Studies in Diatoms. *Journal of Phycology* 30: 507-512.

Chapter 2

SYMPATRIC SPECIATION OF ISLAND PLANTS: THE NATURAL LABORATORY OF LORD HOWE ISLAND

***Alexander S. T. Papadopoulos¹, William J. Baker²
and Vincent Savolainen^{1,2}***

¹Imperial College London, Silwood Park Campus, Ascot,
Berkshire, UK

²Royal Botanic Gardens, Kew, Richmond,
Surrey, UK

ABSTRACT

The evolution of new species without geographic isolation has received a great deal of attention over the past decade. The mechanisms that can lead to speciation of vascular plants in biogeographic sympatry fall broadly into three categories; speciation with ongoing gene flow (which is often, but not always, synonymous with ecological speciation), hybrid speciation and polyploid speciation. Here, with a particular focus on plant species, we briefly review the concepts and research on sympatric speciation and its causes. Hybrid speciation and polyploid speciation have occasionally been dismissed as special (“simple”) cases of sympatric speciation and we consider that, although they are distinctive, there are significant obstacles to overcome before new species can be generated by these mechanisms. As such they are deserving of as much research attention as is currently afforded to speciation with gene flow. We argue that it remains important to continue identifying cases of speciation in sympatry, in order to gain a better understanding of how all of these mechanisms can drive speciation in restricted areas (rather than simply to quantify sympatric speciation per se) and we elaborate on the usefulness of Lord Howe Island as a model system for speciation research with this in mind.

INTRODUCTION

Allopatric and Peripatric Speciation

Species divergence in geographic isolation (allopatry) continues to be widely accepted as the most common geographic mode by which speciation can occur (Coyne and Orr 2004; Mayr 1963; Mayr 1982). Peripatric speciation is a special case of allopatric speciation where one of the initial populations is very small (Mayr 1982). Ongoing gene exchange is considered to be a major hindrance to population splitting and, ultimately, to the evolution of new species (Coyne 1992; Mayr 1963). Complete geographic isolation of populations prevents gene flow between them, allowing the processes of genetic drift and mutation to cause genetic divergence and, given enough time, to the evolution of genetic incompatibilities causing intrinsic prezygotic or postzygotic reproductive isolation of the groups (Coyne and Orr 2004). Under this scenario, divergent or sexual selection are not actually necessary for new species to evolve, although they are likely to play important roles in extrinsic prezygotic and postzygotic reproductive isolation and reinforcement if species' ranges shift and they come into secondary contact (Coyne and Orr 2004; Schluter and Conte 2009; Sobel, et al. 2009).

Sympatric and Parapatric Speciation

“But from reasons already assigned I can by no means agree with this naturalist [Moritz Wagner], that migration and isolation are necessary elements for the formation of new species.”
- Charles Darwin (1859; p79, sixth edition).

Sympatric and parapatric speciation differ from the other modes in that geographic isolation of diverging populations is not complete and speciation occurs despite some genetic exchange (Coyne and Orr 2004; Fitzpatrick, et al. 2008; Mallet, et al. 2009). In order for species to diverge without geographic isolation, a biological mechanism is required to counter-balance the homogenising effect of gene flow (Coyne 1992; Dieckmann and Doebeli 1999; Tregenza and Butlin 1999). Definitions of sympatric and parapatric speciation come in two forms. Biogeographic definitions focus on the spatial geographic pattern under which speciation is initiated. Alternatively, population genetic definitions, are concerned with the probabilities of mating and migration between populations during divergence (for examples of each see Fitzpatrick, et al. 2008). The relative merits and disadvantages of both views continue to be debated thoroughly (Bird, et al. 2012; Butlin, et al. 2008; Coyne and Orr 2004; Fitzpatrick, et al. 2009, 2008; Gavrillets 2003; Kisel and Barraclough 2010; Mallet, et al. 2009).

In population genetic definitions allopatric and sympatric speciation form two extreme ends of a scale. At one end of the scale, two species evolve from completely isolated populations (allopatric speciation), at the other, two species arise from a single panmictic population (sympatric speciation). Speciation events that take place under the varying levels of migration and mating that exist between these extremes constitute parapatric speciation (Butlin, et al. 2008; Fitzpatrick, et al. 2009; Gavrillets 2004; Gavrillets 2003; Mallet, et al. 2009). Population genetic definitions are sufficiently precise to be applied in mathematical models of speciation and have been applied in this context to demonstrate that speciation under these conditions is theoretically possible (Gavrillets and Vose 2007; Gavrillets, et al. 2007; Tregenza

and Butlin 1999), but they contain no explicit spatial component (Fitzpatrick, et al. 2009, 2008; Mallet, et al. 2009). This has lead some researchers to hypothesise that parapatric speciation is the dominant force in nature as the criteria for allopatric and sympatric speciation are so specific that they must be by definition rare (Butlin, et al. 2008; Fitzpatrick, et al. 2008; Mallet, et al. 2009; Nosil 2008). Additionally, the requirement of initial panmixia for sympatric speciation is nearly impossible to translate to the study of wild populations, as we cannot directly observe the event and mating in wild organisms may never truly be random (Butlin, et al. 2008; Endler 1977; Mallet, et al. 2009).

As others have noted, this slightly trivialises the debate over the role of geographic isolation in the speciation process (Mallet, et al. 2009). Despite the criticism that biogeographic definitions contain artificially discrete categories (Butlin, et al. 2008; Mayr 1982), they allow us to address Mayr's proposition that allopatric speciation has been the dominant mode, as well as limiting the evolutionary forces that might be acting in different situations (Coyne and Orr 2004; Mallet, et al. 2009). Consequently, empirical studies addressing the geography of speciation (Barracough and Vogler 2000; Coyne and Price 2000; Fitzpatrick and Turelli 2006; Kisel and Barracough 2010; Krug 2011; Papadopoulos, et al. 2011; Phillimore, et al. 2008; Stuessy, et al. 2006) or investigating specific speciation events (Babik, et al. 2009; Barluenga, et al. 2006; Bird, et al. 2011; Crow, et al. 2010; Savolainen, et al. 2006a) commonly adopt biogeographic definitions. In some cases this is not explicitly stated, instead authors refer to divergence in allopatry or sympatry, e.g., Quesada *et al.* (2007). For the remainder of this chapter the broadest, biogeographic definition of sympatric speciation (see Bird, et al. 2012) is used unless otherwise stated.

Empirical Studies of Speciation

Studying natural populations offers a way to bridge gaps in our scientific knowledge that it is currently not possible to address in laboratories or with mathematical modelling. In speciation research, studies of wild populations generally take advantage of four approaches, either singly or in unison. (i) Isolated islands and archipelagos are particularly useful for speciation research as they often harbour high proportions of endemic species and can be considered as relatively closed systems (Coyne and Price 2000; Kisel and Barracough 2010; Losos and Schluter 2000; Savolainen, et al. 2006a; Stuessy, et al. 2006). (ii) In some groups of organisms, parallel evolution of morphotypes/ecotypes from separate origins allows direct comparison of the processes acting in independent replicates. Despite being relatively rare occurrences, several of these natural experiments have proved to be incredibly fruitful sources of research (e.g., Butlin, et al. 2008; Losos and Miles 1994; McKinnon, et al. 2004; Nosil, et al. 2008; Peccoud, et al. 2009; Rundle, et al. 2000). (iii) Identification of recently diverged sister species or ecologically and phenotypically divergent populations of the same species have provided evidence for the ecological drivers and genetic basis of divergent/disruptive selection and reproductive isolation (Bonin, et al. 2006; Chamberlain, et al. 2009; Hoekstra, et al. 2006; Jiggins, et al. 2001; Ramsey, et al. 2003; Szymura and Barton 1991). (iv) Phylogenetic comparative methods (Felsenstein 1985) have been used to investigate the geographic mode of speciation (Barracough and Vogler 2000; Fitzpatrick and Turelli 2006; Losos and Glor 2003), ecologically driven trait evolution (Losos 1994) and increased rates of species divergence (Barracough, et al. 1999; Barracough and Nee 2001; Barracough, et al. 1998a; Barracough,

et al. 1998b), and adaptive radiations (Baldwin 1997; Schluter 2000). In terms of studying sympatric speciation these approaches can, and have, been applied to macroevolutionary studies addressing the geographic setting of speciation and microevolutionary studies concerned with both identifying case studies and investigating the mechanisms and barriers driving divergence.

Approaches Studying Sympatric Speciation

It is well established that geography plays a key role in the formation of new species (Mayr 1963), however, there is still controversy as to whether geographic separation is actually necessary for speciation to occur and as a result, it is unclear to what extent each of the geographic modes of speciation (allopatric, sympatric, parapatric and peripatric) have contributed to current diversity patterns (Barracough and Nee 2001; Barracough and Vogler 2000; Coyne and Orr 2004). A major obstacle is that as time accumulates following a speciation event, so too can adaptive changes allowing sister species to co-exist. As a consequence post-speciation range shifts can mask the original geographic setting for speciation events (Berlocher 1998). This confounds attempts to unravel geography's influence on speciation events in both macro-evolutionary (Barracough and Nee 2001; Barracough and Vogler 2000; Fitzpatrick and Turelli 2006) and micro-evolutionary studies (Crow, et al. 2010; Rundle, et al. 2000; Schliewen, et al. 1994). Two main approaches have been applied to studies assessing the frequency of speciation modes: (i) comparative studies using current species distributions to infer the geographic distribution at speciation events, making the assumption that current distributions reflect ancestral distributions (Barracough and Vogler 2000; Fitzpatrick and Turelli 2006; Krug 2011; Losos and Schluter 2000; Phillimore, et al. 2008) and (ii) assessments of the frequency of co-occurring endemic island species belonging to the same genus, under the assumption that these cases constitute *in situ* speciation events (Coyne and Price 2000; Kisel and Barracough 2010; Stuessy, et al. 2006).

Overall, very few studies have attempted to address the frequency of speciation without geographic isolation within whole communities or taxonomic groups, and most that have indicate that it is exceptionally rare (Barracough and Vogler 2000; Coyne and Price 2000; Fitzpatrick and Turelli 2006; Losos and Schluter 2000; Phillimore, et al. 2008). Kisel and Barracough (2010) demonstrated that the geographic scale required for speciation is related to the spatial scale of gene flow, and so, for organisms to speciate on small islands they must have very restricted dispersal. In animals, speciation at small spatial scales has been ruled out for some taxa, e.g., in island birds (Coyne and Price 2000) and Caribbean *Anolis* lizards (Losos and Schluter 2000), but may be more common in others, e.g., marine gastropods (Krug 2011). Until recently, evidence from comparative studies that sympatric speciation is uncommon appeared to be borne out by the scarcity of convincing case studies found in nature. However, this may have been a reflection of the difficulties faced when attempting to prove that sympatric speciation has taken place, rather than due to its limited occurrence (Berlocher 1998; Coyne and Orr 2004). For reasons mentioned above, current sympatry of species is insufficient to be confident of the geographic setting during speciation. Coyne and Orr (2004) proposed four conditions that need to be fulfilled for confirmation of a sympatric speciation event; (1) species must be sister taxa, (2) species must occur in sympatry, (3) species must demonstrate reproductive isolation and (4) an allopatric phase in their divergence must be highly unlikely.

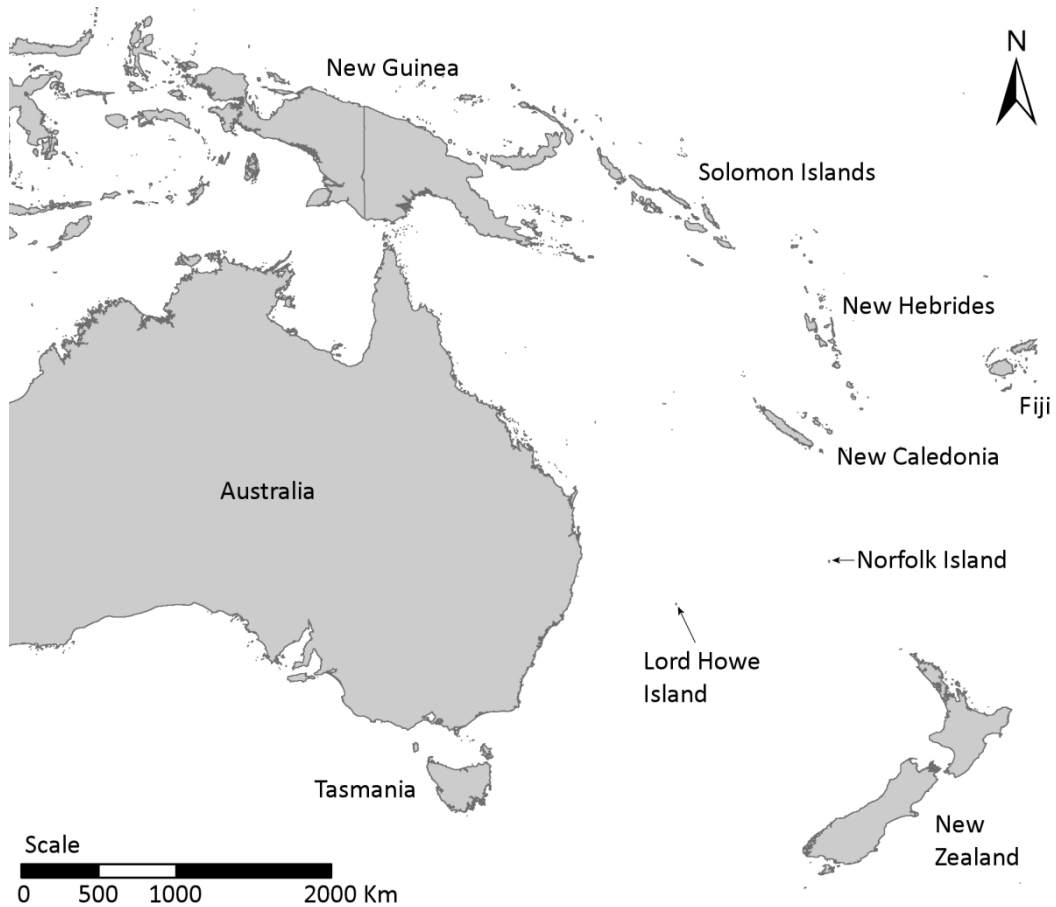


Figure 1. Map showing the location of Lord Howe Island in the Tasman Sea and surrounding landmasses.

To evaluate these criteria studies have typically used: phylogenetic evidence of species relationships to address the first criterion; surveys of species distributions to assess the second; crosses, reciprocal transplants and population genetic approaches to determine the extent of reproductive isolation and; focusing on species pairs from recently formed, and isolated lakes and islands to confirm the fourth. Few studies have managed to conclusively demonstrate all four conditions, often struggling to successfully confirm the last - which is considered the strictest and most critical. Studies of cichlids in crater lakes (Barluenga, et al. 2006; Elmer, et al. 2009), palms on an oceanic island (Babik, et al. 2009; Savolainen, et al. 2006a) and host switching in African indigo birds (Sorenson, et al. 2003) constitute some of the most convincing cases of biogeographic sympatric speciation described, however, even these are not without their detractors (Bolnick and Fitzpatrick 2007; Stuessy 2006a). The discovery of these cases and mathematical modelling of sympatric speciation (Dieckmann and Doebeli 1999; Gavrillets 2003; Gavrillets and Vose 2007; Gavrillets, et al. 2007; Tregenza and Butlin 1999) has led some researchers, particularly those investigating speciation in aquatic environments, to suggest that

this criterion is too strict and limits our ability to effectively study an interesting phenomenon (Bird, et al. 2012; Bird, et al. 2011; Crow, et al. 2010; Krug 2011). In the case of broadcast-spawning Hawaiian limpets (Bird, et al. 2011) this is certainly a justified viewpoint, however, it is currently unclear whether this view will garner widespread support (Elmer and Meyer 2010), particularly, as confirmed cases of sympatric speciation seem to be sporadic and at low frequency. Recent work investigating sympatric speciation on Lord Howe Island (LHI; Figures 1 and 2) may tip the balance in this argument. Papadopoulos et al. (2011) built upon earlier research into palm trees and used Coyne and Orr's criteria to confirm whether or not congeneric species of plants on the island had speciated in sympatry - providing the first assessment of the frequency of confirmed sympatric speciation events at a specific location. By examining the phylogenetic relatedness of multiple species pairs/groups, this research demonstrated that at least 4.5% of species and perhaps as much as 8.2% of the extant species on LHI were the products of sympatric speciation events.

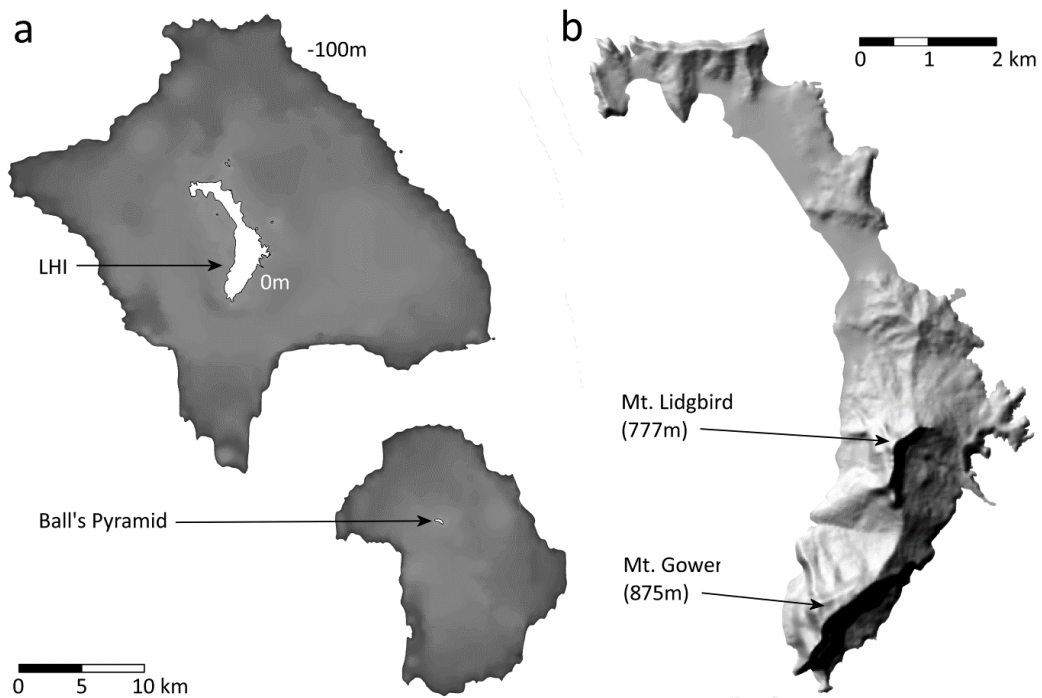


Figure 2. (a) Reconstruction of LHI and Ball's pyramid past and present developed from bathymetric readings (Kennedy 2002). White areas surrounded by black outline denote the current extent of LHI and Ball's pyramid. Shaded areas, depict the largest area of the islands in the past, are currently between 0-100m below sea level. Darker shades indicate increased depth. The historical maximum distance from one point on LHI to another on Ball's Pyramid has been shown to be too small for allopatric speciation to occur (Papadopoulos, et al. 2011) (b) Shaded relief map of LHI.

Causes of Reproductive Isolation during Sympatric and Parapatric Speciation

Table 1. Data collated from literature regarding speciation in island angiosperms. ^a relationships of endemics tested using molecular phylogenetic methods. The number of species belonging to genera endemic to the island or archipelago. ^b Assessed using population genetic techniques. ^c Number of genera for which more than one island endemic has been subject to chromosome counts

Island	Land Area (km ²)	Elevation (m)	Distance to nearest island (km)	No. of endemics with congeneric endemic / No. of endemics	No. of endemics with sister species. / No. of endemics relationships tested ^a	No. of Hybrid species ^b	No. of polyploid genera / No. of genera assessed ^c
LHI	14.6	875	600	28 / 72	14/24	2	0/3
Pagalu	15.7	654	180	2 / 16	-	-	-
Isla del Coco	24	630	500	9 / 18	-	-	1/1
Alejandro Selkirk	55.5	1649	160	8 / 27	-	-	0/5
Ullung	73	984	130	4 / 33	0/2	-	0/2
Robinson Crusoe	93	914	160	26 / 50	5/8 (+2)	-	0/5
Ascension	97	859	1300	2 / 4	-	-	-
Rodrigues	107.8	396	650	6 / 43	-	-	-
St. Helena	125.5	819	1200	18 / 38	7/7 (+3)	-	-
Christmas Island	135	357	400	2 / 15	-	-	-
Principe	148.5	948	150	10 / 31	-	-	-
Guadalupe	264.2	1298	260	8 / 27	2/2	-	-
Sao Tome	854.8	2024	150	45 / 93	4/6	-	-
Mauritius	1873.8	828	250	197 / 243	6/18 (+18)	-	-
Réunion	2535.2	3069	150	126 / 178	6/21 (+6)	-	-
Archipelago	Land Area (km ²)	Elevation (m)	Distance to nearest island (km)	No. of endemics with congeneric endemic present / No. of endemics	No. of endemics with sister spp. / No. of endemics relationships tested ^a	No. of Hybrid species ^b	No. of polyploid genera / No. of genera assessed
Norfolk	64	318	670	7 / 35	0/2	-	1/2
Osagawara	99	916	800	55 / 118	6/13	-	3/9
Socorro	426	1130	300	4 / 28	-	-	-
Tristan da Cunha	705	2060	350	8 / 10	4/4	-	1/2
Chatham Islands	1416	283	600	14 / 36	8/15	-	1/4
Palau	1480	245	850	56 / 119	-	-	-
Madeira	2637	1861	450	49 / 97	12/16	-	0/5
Cape verde	4033	2829	570	43 / 68	2/4 (+5)	-	2/5
Canary Islands	7601	3710	100	360 / 429	7/10 (+16)	2	8/9
Falkland Islands	8500	705	410	4 / 14	-	-	1/2
Hawaii	16885	4250	3660	770 / 828	222/240	1	8/54
Galapagos	42560	1707	850	83 / 148	3/5 (+16)	1	0/3

In sexual organisms, speciation in sympatry/parapatry can occur via speciation-with-gene-flow, homoploid hybrid speciation (sometimes known as recombinational speciation) and allo- and auto-polyploid speciation. In each case there are several mechanisms that can potentially initiate divergence and lead to speciation but these classes can be distinguished in two ways. First, in both forms of polyploid speciation, reproductive isolation of the emergent species is a nearly instantaneous by-product of polyploidisation (Coyne and Orr 2004; Mallet 2007;

Ramsey and Schemske 1998; Rieseberg and Willis 2007). On the other hand, the evolution of reproductive isolation in homoploid hybrid speciation and speciation-with-gene-flow may or may not be reliant on some form of divergent selection and is unlikely to occur in a single step (Coyne and Orr 2004; Dieckmann and Doebeli 1999; Gompert, et al. 2006; Mallet 2008; Nosil, et al. 2009; Rundle and Nosil 2005; Tregenza and Butlin 1999). It is important to note that divergent selection is almost certainly instrumental in the persistence of the species generated by all of the above processes, but does not play a role in the evolution of reproductive isolation during polyploidisation (Coyne and Orr 2004; Rundle and Nosil 2005; Schluter 2001, 2000; Sobel, et al. 2009). The second difference relates to the number of species required to initiate a new speciation event. When autopolyploidisation and speciation-with-gene-flow occur, two distinct species arise from a single ancestral population. Allopolyploid and homoploid hybrid speciation both require the crossing of two distinct species to generate a third lineage (Coyne and Orr 2004; Gavrillets 2004; Mallet 2007; Ramsey and Schemske 1998; Rieseberg and Willis 2007).

We collated data from a variety of sources to evaluate the extent to which these processes have been studied in island systems (Table 1). Using the island angiosperm data collected by Kisel and Barraclough (2010) and Steussy et al. (2006) as a starting point, we reviewed angiosperm species lists for islands without data in the previous studies. We then expanded on the phylogenetic data examined by Kisel and Barraclough (2010) by performing a systematic search to find evidence of within island sister species. This was carried out using the web of science with search terms “phylogenetic” and the name of each genus or the name of the island (for regions larger than 5000km²). We recorded the number of species whose relationships had been examined using molecular phylogenetic approaches and the number of species with a sister species within the same island/archipelago recovered unequivocally in these phylogenetic reconstructions. We then searched for population genetic research into homoploid hybrid speciation on these islands using the island name and “homoploid hybrid” and “diploid hybrid” as search terms. Finally, using the Index to Plant Chromosome numbers database (IPCN, <http://www.tropicos.org/Project/IPCN/>) we recorded the number of island genera for which ploidy assessments have been carried out and the number of these studies which found evidence for chromosome number variation within genera on a single island. All data is presented in Table 1.

Speciation-with-Gene-Flow

One of the central questions in speciation research is how ecologically driven divergent or disruptive selection might lead to population differentiation, reproductive isolation and the eventual evolution of new species in the face of ongoing gene flow (Butlin, et al. 2008; Hendry, et al. 2007; Kirkpatrick and Ravigné 2002; Mallet 2008; McKinnon, et al. 2004; Rundle and Nosil 2005; Schluter 2001). Divergent or disruptive selection pressures imposed by ecological interactions can have a considerable impact on divergence of adaptive morphological and physiological traits. For this to develop into “ecological speciation” this environmentally driven divergence needs also to be the root mechanism by which barriers to gene flow evolve between populations (Coyne and Orr 2004; Rundle and Nosil 2005; Schluter 2001). The term ecological speciation has been accused of being a misnomer as ecology is expected to play some role in

all speciation events (Sobel, et al. 2009), however, we use it here as the concept is clearly defined and widely applied (Nosil 2012).

Three components are required to cause ecological speciation with gene flow; an ecological source of divergent selection, a reproductive isolating mechanism and a genetic mechanism to link the divergent selection and isolation (Kirkpatrick and Ravigné 2002; McKinnon, et al. 2004; Rundle and Nosil 2005; Schluter 2001, 2000). Ecological divergent selection can stem from numerous sources that are not mutually exclusive. Factors promoting divergent selection may be abiotic - such as climate, habitat structure and resource abundance - or biotic - such as sexual selection, inter- or intra-specific competition and predation (Rundle and Nosil 2005; Schluter 2000). Similarly, reproductive barriers can take various forms, both prezygotic (e.g., habitat isolation, temporal isolation, pollinator isolation, behavioural isolation) or postzygotic (e.g., Hybrid inviability or sexual selection against hybrids; Coyne and Orr 2004; Rundle and Nosil 2005; Schluter 2000; Sobel, et al. 2009). In order for divergent selection and reproductive barriers to interact and lead to population divergence a genetic mechanism is required to link the two processes, making them heritable. Theoretically, this can occur via the evolution of a pleiotropic ‘magic trait’, through linkage disequilibrium between genes involved in local adaptation and assortative mating, or because ecological differences within a species range produce divergent genetic adaptations as well as plastic responses that confer reproductive isolation (e.g., environmentally controlled shifts in flowering time; Coyne and Orr 2004; Devaux and Lande 2008; Dieckmann and Doebeli 1999; Gavrillets 2003; Mallet, et al. 2009; Tregenza and Butlin 1999). Uniform selection (mutation order speciation) and genetic drift may play important roles during ecological speciation in allopatry, but not in sympatry or parapatry (Coyne and Orr 2004; Schluter 2009). Evidently, the genetic basis of divergence and the nature of the isolating barriers may differ between specific speciation events, nevertheless, divergent natural selection is likely to be the driving force behind the evolution of reproductive isolation in many instances of speciation-with-gene-flow (Dieckmann and Doebeli 1999; Doebeli and Dieckmann 2003; Doebeli, et al. 2005; Higashi, et al. 1999; Kirkpatrick and Ravigné 2002), although other mechanisms may lead to speciation in some instances (e.g., through sexual selection as Fisherian runaway; van Doorn, et al. 1998).

Determining whether sympatric speciation with gene flow has occurred requires, not only, that Coyne and Orr’s criteria are fulfilled but, additionally, the sister species relationship recovered in phylogenetic reconstructions must not be an artefact of hybridization. Although it is not stated explicitly in the criteria, hybrid speciation (homoploid or polyploid) would be excluded by criterion one if both progenitors have been sampled in a phylogenetic study, as long as the resolution of the gene regions used is sufficient. Furthermore, if polyploidisation (assessed using chromosome counts or genome size estimation) is not evident in the sympatrically diverging species then fulfilment of these criteria is a strong indication that speciation in the face of strong gene flow has taken place (Coyne 2011; Papadopoulos, et al. 2011).

In plants, cases of ecological divergence despite ongoing gene flow, a precursor to speciation, have been proposed using statistical inference of historical gene flow for genetic markers (e.g., in *Oryza* and *Pinus*; Wachowiak, et al. 2011; Zheng and Ge 2010). However, examples of sympatric speciation with gene flow are limited to those found in *Howea*, *Metrosideros* and the radiation of *Coprosma* on Lord Howe Island (LHI), where polyploidisation, hybrid speciation and divergence in allopatry could be ruled out with a reasonable degree of confidence (Coyne 2011; Papadopoulos, et al. 2011; Savolainen, et al.

2006a). Despite this lack of proven cases there are many other potential instances of divergence with gene flow in island plants. Estimates of cladogenetic divergence within islands and archipelagos, based on the presence of congeners, reveal surprisingly large numbers of co-occurring species pairs (Table 1; Kisel and Barraclough 2010; Stuessy, et al. 2006). It is important to point out that more detailed analyses have shown in the case of LHI (Papadopoulos, et al. 2011), and the general patterns observed in the data collated here, that these are over-estimates of *in situ* cladogenesis, but the existing phylogenetic evidence shows that there are sister species present on many islands. However, even when sister species are present on an island, polyploidisation, hybrid speciation and geographic isolation are also good candidates for drivers of divergence, particularly in the case of archipelagos and the larger islands (Coyne 2011; Kisel and Barraclough 2010; Papadopoulos, et al. 2011). Alternatively the patterns observed may be driven by extinctions or multiple colonisations. Further identification and investigation of sympatric speciation events in island plants has the potential to greatly advance our understanding of the mechanisms allowing the evolution of reproductive barriers in the face of ongoing genetic exchange and the contribution this kind of speciation has made to patterns of global biodiversity (Coyne 2011).

Homoploid Hybrid Speciation

The process by which new genotypes are generated differs between homoploid hybrid speciation and speciation with gene flow. Despite this, the same obstacle of continuing gene exchange must be overcome in order for hybridisation to progress to speciation. Without allopatric separation of the hybrid population this is believed to occur through ecological speciation in a similar fashion to speciation with gene flow (Buerkle, et al. 2000; Gross and Rieseberg 2005; Mallet 2007; Mavarez, et al. 2006; Rieseberg and Willis 2007). An additional problem also faces new hybrid species; competition with its progenitors. Hybrid genotypes are unlikely to be at a competitive advantage in either of the parent species habitats, for them to persist they must be pre-adapted to, and near the fitness maxima of, an available niche (Coyne and Orr 2004; Mallet 2007; Rieseberg and Willis 2007; Sobel, et al. 2009). This insight has led to the general acceptance that homoploid hybrid speciation is relatively rare, but a number of examples exist in both plants (Rieseberg, et al. 2003; Rieseberg, et al. 1995) and animals (Gompert, et al. 2006; Mavarez, et al. 2006). Alternatively, hybrid genotypes may benefit from improved utilisation of one of the parent species niches leading to its extinction, making the detection of hybrid speciation more problematic (Gross and Rieseberg 2005; Mavárez and Linares 2008).

The genetics and ecology of hybrid speciation is a vibrant area of research, providing insight into adaptive radiations and the role of divergent selection in reproductive isolation. By definition homoploid hybrid speciation occurs in sympatry and, despite plentiful evidence of reticulate evolution and hybridisation in island plants gleaned from phylogenetic studies (e.g., Howarth and Baum 2005; Liu, et al. 2009), it is expected to be rare in very restricted areas where co-occurrence of reproductively compatible species is less likely and the opportunity for a newly formed hybrid lineage to escape competition with its progenitors is significantly diminished. Phylogenetic work on LHI suggests that at least two plant species (*Calystegia affinis* and *Myrsine mcomishii*) are likely to be of hybrid origin, although in both cases it is not clear whether hybridisation did in fact take place on the island (Papadopoulos, et al. 2011).

Although it is expected to be uncommon, surprisingly little research has evaluated whether potential cases can truly be considered as homoploid hybrid speciation (rather than examples of introgressive hybridisation) using population genetic methods (e.g., Brochmann, et al. 2000; Friar, et al. 2006; Remington and Robichaux 2007; Smith, et al. 1996). Our recent population genetic research into the origins of the LHI *Coprosma* species suggests that two species may be the result of hybridisation between species that had previously diverged via speciation with gene flow on the island (Papadopoulos, et al. 2012). The few well studied instances show that homoploid hybrid speciation can occur on islands, but the frequency and mechanisms that govern this process remain heavily understudied.

Polyplloid Speciation

Polyplloid individuals possess three or more complete assemblages of chromosomes. When a new species arises as a result of an increase in the number of chromosome sets this is known as polyplloid speciation (Mallet 2007; Ramsey and Schemske 1998; Rieseberg and Willis 2007). In these cases speciation is expected to be virtually instantaneous because of the incompatibility and sterility associated with crosses between the original species and the polyplloid species (Ramsey and Schemske 1998). Polyplloidisation can develop through three main mechanisms; somatic doubling, meiotic non-reduction and polyspermy (Mallet 2007; Ramsey and Schemske 1998; Rieseberg and Willis 2007). New polyplloid species can either be derived from a single progenitor (autopolyploidy) or by hybridisation of two distinct species (allopolyploidy; Ramsey and Schemske 1998). Estimates of the frequency of polyplloid speciation in nature vary according to the method used and the group assessed (Coyne and Orr 2004). In animals, polyplloid speciation is generally accepted to be exceptionally rare (Otto and Whitton 2000), whereas in vascular plants recent estimates range from 2 - 15% in angiosperms and 7 - 31% in ferns (Otto and Whitton 2000; Wood, et al. 2009). Why is there such a difference in rates of polyplloidisation between plants and animals? There is no conclusive answer to this question, several hypotheses have been proposed and subsequently dismissed (Mable 2004; Orr 1990). Probably the most popular view is that the presence of genetically degenerate sex chromosomes (relatively uncommon in plants) results in incompatibilities in tetraploids, but this too does not satisfactorily explain the variation in rates of polyplloidy between the different groups (Orr 1990).

Allopolyploid speciation is often considered to occur considerably more frequently than autopolyploid speciation, despite the fact that autopolyploidisation has been estimated to occur at higher rates than allopolyploidisation (Coyne and Orr 2004; Mallet 2007; Ramsey and Schemske 1998). The discrepancy between the rates of auto- and allopolyploidy and the frequency of subsequent speciation may be for two reasons: (1) although autopolyploids form more often, they may be less capable of persisting as they are usually ecologically similar to, and thus in competition with, their progenitors. The result is that they frequently go extinct or replace the original diploid species. As with other hybrids, allopolyploids are often morphologically and ecologically divergent from the progenitor species and so may be better suited to co-exist or colonise new habitats (Mallet 2007; Ramsey and Schemske 1998; Rieseberg and Willis 2007; Sobel, et al. 2009). (2) Autopolyploids are often morphologically indistinguishable from their diploid progenitors and as a result they may not be detected. This problem is further compounded by the reticence of taxonomists to describe cryptic polyploids

as new species (Rieseberg and Willis 2007; Soltis, et al. 2007). Although opinions differ over which explanation is more likely, it is generally accepted that allopolyploid speciation is much more common (Mallet 2007).

For similar reasons to homoploid hybrid speciation, allopolyploid speciation and autopolyploid speciation are expected to be rare in island systems; indeed very few of the genera tested to date have shown any chromosome number variation (Table 1). In addition, no angiosperm genera with confirmed sister species present, with the exception of those on the larger island and archipelagos, show signs of polyploidisation; however, very few have actually been tested. The competitive difficulties that face a new polyploid lineage are very similar to those faced by a homoploid hybrid, however, polyploidisation is often considered to cause instant reproductive isolation and, unlike the other mechanisms, one of the major barriers to speciation (i.e., gene flow) is circumvented. Recently, research has shown that not only has gene flow occurred from diploids to tetraploids in natural populations (a widely acknowledged phenomenon; Soltis and Soltis 1999) but is bidirectional in some groups (e.g., *Arabidopsis* and *Dactylorhiza*; Jørgensen, et al. 2011; Ståhlberg 2009). Studying polyploidisation in island plants offers an opportunity to understand how new lineages escape competition with their progenitor/s and whether gene flow between ploidal levels has implications for the ease of sympatric speciation via polyploidisation.

LORD HOWE ISLAND: A MODEL SYSTEM FOR STUDYING SPECIATION

Lord Howe Island Formation and Climate

Lord Howe Island (LHI) is a small island (16 km²; Figures 1 and 2) lying at the southern end of a 1000 km line of volcanic seamounts which extends northwards along the Lord Howe Rise (McDougall, et al. 1981; Pickard 1983). Located 600 km east of Australia (Figure 1), the nearest seamount, Ball's Pyramid, emerges from the sea some 24 km SSE of LHI (Figure 2a). The next link in the chain is Elizabeth reef approximately 170 km north. The island is the eroded remains of a large shield volcano which first became active 6.9 mya and rapidly became dormant, producing the last of the island's basaltic rocks 6.4 mya (McDougall, et al. 1981). The island sits on a shelf that is 24 km wide and 36 km from north to south and is surrounded by a fringing reef that forms a large lagoon on the western side (Woodroffe, et al. 2006). Currently, the landscape is highly heterogeneous, in the south two heavily eroded mountains, Mt. Lidgbird (777 m) and Mt. Gower (875 m), dominate the skyline with numerous cliffs and creeks intersecting the area to produce a patchwork of different habitats (McDougall, et al. 1981; Pickard 1983).

Studies of the geology and bathymetry indicate that the volcanic areas of LHI currently above sea level were rapidly eroded into their current state within 1-2 Myr of the initial eruption and have since been buffered from significant wave erosion by the presence of coral reefs (Brooke 2003; Kennedy 2002; Woodroffe, et al. 2006). LHI and Ball's pyramid lie on either side of the southerly limit to fringing reef formation, known as the Darwin point. The younger Ball's Pyramid possess no reef to provide protection from erosion and has undergone significant (almost complete) planation, whereas LHI has migrated northwards into reef building seas and is protected by a fringing reef (Woodroffe, et al. 2006). These studies also

indicate that neither LHI nor Ball's pyramid have subsided, but changes in sea level during glacial periods would have increased the size of both to the extent of the shelves upon which they sit (Brooke 2003; Kennedy 2002; Woodroffe, et al. 2006). Despite the increased size of the island during brief periods in its history, population genetic analyses for a range of plant species suggest that at no point would geographic distance between individuals of a species have been sufficient to allow neutral differentiation, particularly for wind dispersed plants (Papadopoulos, et al. 2011). In other words allopatric divergence is very unlikely to have played a role during speciation on LHI.

Deposition of calcareous material during periods of low sea level in the late Pliocene and Pleistocene provided the island with the sedimentary Ned's Beach Calcarene which is dominant in the north of the island below 100 m (McDougall *et al.* 1981). The volcanic edifice of Elizabeth reef formed 10.2 mya and rises abruptly from 2000 m below sea level. The Elizabeth reef pedestal is considerably smaller than that of LHI and Ball's Pyramid (10.7 km by 6.2 km) and is capped by an atoll that extends to a depth of 40 m (McDougall, et al. 1981; Woodroffe, et al. 2004). Currently, no plant life is found at Elizabeth reef and only five, highly salt tolerant, plant species occur on Ball's Pyramid, all of which can also be found on LHI (Green 1994; Priddel, et al. 2003). The tropical location, small size and considerable erosion of the Elizabeth reef sea mount suggest that it is unlikely to have been a large and ecologically similar island during the life time of LHI.

LHI's climate is essentially sub-tropical with the humidity ranging between 68 and 73 percent throughout the year (Pickard 1983). The mean number of rain days per month ranges from 11 in summer to 22 in the winter with a similar seasonal pattern seen in the mean monthly precipitation (Gentili 1971; Pickard 1983). Cloud cover is often present, particularly around the mountains which generate their own orographic cloud. As a result rainfall in the south is more common, producing a wetter environment. Mean annual temperature is 19.1°C and temperatures at the summits are calculated to be about 6-8°C lower than the lowlands. There is no record of a frost ever hitting the island. Wind speeds and directions do vary dramatically, but onshore resultants for winds from the North-east and South-east are considerably more dominant and stronger in magnitude (Pickard 1983). These variations in climate generate localised microclimatic changes enhancing the patchiness of the environment.

The Vegetation of LHI and the Impact of Human Settlement

Habitat differentiation on the island is reflected in the vegetative assessment of Pickard (1983), in which he identified nine broad scale vegetative formations which were then divided into 18 plant alliances, these in turn he further classified into finer scale plant associations. Although highly detailed and thoroughly documented, these associations and the map of their distributions do not completely capture the variation that is actually present on the island. The composition of the communities found within these associations can vary dramatically at fine scales and, often, many small patches (<20 m²) of one association are scattered throughout a larger patch of another. Examples include the scattered patches of *Boehmeria* - *Macropiper* throughout the *Dracophyllum* - *Metrosideros* along the east side of Mount Lidgbird or the *Dracophyllum*-*Metrosideros* areas within the *Cleistocalyx* - *chionanthus* region in the Erskine Valley (personal observations). The majority of the island is covered in dense forest of various kinds, however, there are a number of more exposed sites, as well as the offshore islets, where

scrub (e.g., the *Melaleuca-Cassinia* alliance) and grassland (the *Poa* alliance) dominate (personal observations; Pickard 1983).

First discovered in 1788, settlement of the island began in the 1830's, it is now inhabited by around 300 permanent residents with as many as 400 visitors at any one time (Hutton, et al. 2007; Pickard 1983). Although there has been extensive clearing of the lowland areas for settlement, the vegetation of the Northern hills and the southern mountains has remained virtually untouched. Overall, less than 15 percent of the forests have been cleared and less than 20 percent of the vegetation is disturbed (Auld and Hutton 2004; Pickard 1983). The first botanists arrived on the island in 1853, providing an invaluable record of the flora's recent history (Pickard 1984). Although three species of bird have been hunted to extinction, the main cause of declines in native species appears to be the introduction of exotic biota (Auld and Hutton 2004; Hutton, et al. 2007). Consumption by introduced animals is likely to have caused the extinctions of two plant species (Auld and Hutton 2004). Pigs and feral goats had only locally significant impacts on vegetation (Pickard 1976) and these species, as well as feral cats, are now believed to be absent from the island due to eradication programs (Auld and Hutton 2004; Hutton, et al. 2007). The extinctions of further five bird species, and one subspecies, have been attributed to the introduction of ship rats in 1918 (Hutton, et al. 2007). Rats are believed to be the cause of the disappearance of two invertebrate species from the main island and severe declines in population numbers of the only two native lizards (Case 1991) and an endemic worm species (Hutton, et al. 2007). Several plant species suffer from heavy predation of seeds and fruits and others, mainly the palm species, receive stem damage and loss of seedlings, however the impact of this on the ecology of the island is poorly understood (Auld and Hutton 2004).

To date, 230 plant species have been introduced to the island (Hutton, et al. 2007; Pickard 1984), 45 of these appear to be having an impact on native species and the structure of the ecological communities (Auld and Hutton 2004; Hutton, et al. 2007). The secondary damage caused by removing hardy plants to expose more fragile rainforest to strong, salty winds appears to be one of the greatest threats to the islands native wildlife as it both damages the native flora and allows invasive species to get a foothold (Auld and Hutton 2004; Pickard 1983). Despite these problems, the majority of the island's vegetation has not been severely damaged as threats are localised to certain areas of the island or affect particular species (Auld and Hutton 2004; Brown and Baker 2009; Hutton, et al. 2007; Pickard 1983). In recent decades, a great deal of human effort has been invested in preserving the island's unique ecosystem; removal of invasive plant species has been a priority of the local government for some time and efforts are underway to eradicate the rat population (Department of Environment and Climate Change NSW 2007).

Why Study Speciation on LHI?

The Island's wildlife is incredibly diverse and unique; in some groups of the 1600 invertebrate species endemism is as high as 60% (Department of Environment and Climate Change NSW 2007) and the LHI flora comprises 242 vascular plant species of which 90 are endemic (Green 1994). Aside from this diversity and distinctiveness of the LHI flora and fauna there are several other characteristics of LHI that make it an excellent site for research into sympatric speciation. Having formed as a result of a volcanic eruption, the island has never

been connected to other landmasses by land bridges. Combined with the extreme isolation and small size of the island this is believed to make repeated colonisation by the same species, or by close relatives, improbable (Papadopoulos, et al. 2011; Savolainen, et al. 2006a). Furthermore, the small size of LHI reduces the possibility that sister species have ever been completely geographically isolated from each other particularly in organisms with efficient dispersal modes (Kisel and Barraclough 2010; Papadopoulos, et al. 2011). Overall, these properties of the island present an unusually appropriate set of circumstances for the identification of sympatric speciation events (Papadopoulos, et al. 2011; Savolainen, et al. 2006a; Savolainen, et al. 2006b), although this has been disputed (Stuessy 2006b). Due to the young age of the island, any speciation events on LHI are likely to have occurred relatively recently in evolutionary time, facilitating the investigation of factors that may have promoted divergence. Some genera that have speciated on LHI have diversified into similar ecotypes in other island systems (e.g., *Metrosideros polymorpha* and *Coprosma* species in Hawaii), raising the possibility that these island systems may also harbour cases of parallel speciation in plants. The recent discovery of so many sympatric speciation events of the island provides an opportunity to investigate how natural selection has caused different reproductive barriers to evolve in a diverse array of taxa under similar ecological conditions.

CONCLUSION

Although it is still a source of scientific debate, a growing body of evidence suggests that sympatric speciation can occur in island plants via speciation with gene flow as well as through homoploid hybrid speciation and, less controversially, polyploid speciation. The evidence seems to point to a significant role for sympatric speciation in island evolution and we are tantalisingly close to an answer. What has become clear is that all three mechanisms operate to some extent in the island floras that have been examined, however, the basic phylogenetic, population genetic and ploidy level data is often lacking to make complete assessments. Polyploid series are generally rare in the floras assessed and thorough assessment of hybrid speciation has been largely ignored. Repeatedly, the same taxonomic groups have diversified in different island systems suggesting that plants and certain genera in particular may be prone to ecologically driven speciation, in some cases suggesting that parallel speciation may have taken place (a similarly understudied phenomenon in plants; Ostevik, et al. 2012). There is still a great deal we do not know about the mechanisms and processes leading to within island speciation and systems like LHI offer the opportunity to tackle many unanswered questions.

REFERENCES

- Auld TD, Hutton I. 2004. Conservation issues for the vascular flora of Lord Howe Island. *Cunninghamia*. 8: 490-500.
- Babik W, Butlin RK, Baker WJ, Papadopoulos AST, Boulesteix M, Anstett MC, Lexer C, Hutton I, Savolainen V. 2009. How sympatric is speciation in the *Howea* palms of Lord Howe Island? *Mol Ecol*. 18: 3629-3638.

- Baldwin BG. 1997. Adaptive radiation of the Hawaiian silversword alliance: congruence and conflict of phylogenetic evidence from molecular and non-molecular investigations. In: Givnish TJ, Systma KJ, editors. *Molecular Evolution and Adaptive Radiation*. Cambridge: Cambridge University Press. p. 103-128.
- Barluenga M, Stolting KN, Salzburger W, Muschick M, Meyer A. 2006. Sympatric speciation in Nicaraguan crater lake cichlid fish. *Nature*. 439: 719-723.
- Barracough TG, Hogan JE, Vogler AP. 1999. Testing whether ecological factors promote cladogenesis in a group of tiger beetles (Coleoptera: Cicindelidae). *Philos Trans R Soc Lond B*. 266: 1061-1067.
- Barracough TG, Nee S. 2001. Phylogenetics and speciation. *Trends Ecol Evol*. 16: 391-399.
- Barracough TG, Nee S, Harvey PH. 1998a. Sister-group analysis in identifying correlates of diversification. *Evol Ecol Res*. 12: 751-754.
- Barracough TG, Vogler AP. 2000. Detecting the geographical pattern of speciation from species-level phylogenies. *Am Nat*. 155: 419-434.
- Barracough TG, Vogler AP, Harvey PH. 1998b. Revealing the factors that promote speciation. *Philos Trans R Soc Lond B*. 353: 241-249.
- Berlocher SH. 1998. Can sympatric speciation via host or habitat shift be proven from phylogenetic and biogeographic evidence? In: Howard DJ, Berlocher SH, editors. *Endless Forms: Species and Speciation Oxford*: Oxford University Press.
- Bird CE, Fernandez-Silva I, Skillings DJ, Toonen RJ. 2012. Sympatric speciation in the post "modern synthesis" era of evolutionary biology. *Evol Biol*. 39: 158-180.
- Bird CE, Holland BS, Bowen BW, Toonen RJ. 2011. Diversification of sympatric broadcast-spawning limpets (*Cellana* spp.) within the Hawaiian archipelago. *Mol Ecol*. 20: 2128-2141.
- Bolnick DI, Fitzpatrick BM. 2007. Sympatric speciation: models and empirical evidence. *Annu Rev Ecol Evol Syst*. 38: 459-487.
- Bonin A, Taberlet P, Miaud C, Pompanon F. 2006. Explorative genome scan to detect candidate loci for adaptation along a gradient of altitude in the common frog (*Rana temporaria*). *Mol Biol Evol*. 23: 773-783.
- Brochmann C, Borgen L, Stabbetorp OE. 2000. Multiple diploid hybrid speciation of the Canary Island endemic *Argyranthemum sundingii* (Asteraceae). *Plant Syst Evol*. 220: 77-92.
- Brooke B. 2003. Quaternary calcarenite stratigraphy on Lord Howe Island, southwestern Pacific Ocean and the record of coastal carbonate deposition. *Quaternary Sci Rev*. 22: 859-880.
- Brown D, Baker L. 2009. The Lord Howe Island biodiversity management plan: an integrated approach to recovery planning. *Ecol Manage Restor*. 10: S70-S78.
- Buerkle CA, Morris RJ, Asmussen MA, Rieseberg LH. 2000. The likelihood of homoploid hybrid speciation. *Heredity*. 84: 441-451.
- Butlin RK, Galindo J, Grahame JW. 2008. Sympatric, parapatric or allopatric: the most important way to classify speciation? *Philos Trans R Soc Lond B*. 363: 2997-3007.
- Case T. 1991. The role of introduced species in shaping the distribution and abundance of island reptiles. *Evol Ecol*. 5: 272-290.
- Chamberlain NL, Hill RI, Kapan DD, Gilbert LE, Kronforst MR. 2009. Polymorphic butterfly reveals the missing link in ecological speciation. *Science*. 326: 847-850.
- Coyne JA. 1992. Genetics and speciation. *Nature*. 355: 511-515.

- Coyne JA. 2011. Speciation in a small space. *Proc Natl Acad Sci USA*. 108: 12975-12976.
- Coyne JA, Orr HA. 2004. *Speciation*. Massachusetts: Sinauer Associates.
- Coyne JA, Price TD. 2000. Little evidence for sympatric speciation in island birds. *Evolution*. 54: 2166-2171.
- Crow KD, Munehara H, Bernardi G. 2010. Sympatric speciation in a genus of marine reef fishes. *Mol Ecol*. 19: 2089-2105.
- Devaux C, Lande R. 2008. Incipient allochronic speciation due to non-selective assortative mating by flowering time, mutation and genetic drift. *Proc R Soc B*. 275: 2723-2732.
- Dieckmann U, Doebeli M. 1999. On the origin of species by sympatric speciation. *Nature*. 400: 354-357.
- Doebeli M, Dieckmann U. 2003. Speciation along environmental gradients. *Nature*. 421: 259-264.
- Doebeli M, Dieckmann U, Metz JA, Tautz D. 2005. What we have also learned: adaptive speciation is theoretically plausible. *Evolution*. 59: 691-695.
- Elmer KR, Lehtonen TK, Meyer A. 2009. Color assortative mating contributes to sympatric divergence of neotropical cichlid fish. *Evolution*. 63: 2750-2757.
- Elmer KR, Meyer A. 2010. Sympatric speciation without borders? *Mol Ecol*. 19: 1991-1993.
- Endler JA. 1977. *Geographic Variation, Speciation and Clines*. Princeton, NJ: Princeton University Press.
- Felsenstein J. 1985. Phylogenies and the comparative method. *Am Nat*. 125: 1-15.
- Fitzpatrick BM, Fordyce JA, Gavrillets S. 2009. Pattern, process and geographic modes of speciation. *J Evol Biol*. 22: 2342-2347.
- Fitzpatrick BM, Fordyce JA, Gavrillets S. 2008. What, if anything, is sympatric speciation? *J Evol Biol*. 21: 1452-1459.
- Fitzpatrick BM, Turelli M. 2006. The geography of mammalian speciation: mixed signals from phylogenies and range maps. *Evolution*. 60: 601.
- Friar E, Prince L, Roalson E, McGlaughlin M, Cruse Sanders J, De Groot S, Porter JM. 2006. Ecological speciation in the East Maui-endemic *Dubautia* (Asteraceae) species. *Evolution*. 60: 1777-1792.
- Gavrillets S. 2004. *Fitness Landscapes and the Origin of Species*. Princeton, NJ: Princeton University Press.
- Gavrillets S. 2003. Models of speciation: what have we learned in 40 years? *Evolution*. 57: 2197-2215.
- Gavrillets S, Vose A. 2007. Case studies and mathematical models of ecological speciation. 2. Palms on an oceanic island. *Mol Ecol*. 16: 2910-2921.
- Gavrillets S, Vose A, Barluenga M, Salzburger W, Meyer A. 2007. Case studies and mathematical models of ecological speciation. 1. Cichlids in a crater lake. *Mol Ecol*. 16: 2893-2909.
- Gentili J. 1971. *Climates of Australia and New Zealand*. Amsterdam: Elsevier.
- Gompert Z, Fordyce JA, Forister ML, Shapiro AM, Nice CC. 2006. Homoploid hybrid speciation in an extreme habitat. *Science*. 314: 1923-1925.
- Green PS. 1994. *Flora of Australia Volume 49 Oceanic Islands 1*. Canberra: Australian Government Publishing Service.
- Gross BL, Rieseberg LH. 2005. The ecological genetics of homoploid hybrid speciation. *J Hered*. 96: 241-252.

- Hendry AP, Nosil P, Rieseberg LH. 2007. The speed of ecological speciation. *Funct Ecol.* 21: 455-464.
- Higashi M, Takimoto G, Yamamura N. 1999. Sympatric speciation by sexual selection. *Nature.* 402: 523-526.
- Hoekstra HE, Hirschmann RJ, Bunday RA, Insel PA, Crossland JP. 2006. A single amino acid mutation contributes to adaptive beach mouse color pattern. *Science.* 313: 101-104.
- Howarth DG, Baum DA. 2005. Genealogical Evidence of Homoploid Hybrid Speciation in an Adaptive Radiation of *Scaevola* (Goodeniaceae) in the Hawaiian Islands. *Evolution.* 59: 948-961.
- Hutton I, Parkes JP, Sinclair ARE. 2007. Reassembling island ecosystems: the case of Lord Howe Island. *Anim Conserv.* 10: 22-29.
- Jiggins CD, Naisbit RE, Coe RL, Mallet J. 2001. Reproductive isolation caused by colour pattern mimicry. *Nature.* 411: 302-305.
- Jørgensen MH, Ehrich D, Schmickl R, Koch MA, Brysting AK. 2011. Interspecific and interploidal gene flow in Central European *Arabidopsis* (Brassicaceae). *BMC Evol Biol.* 11: 346.
- Kennedy D. 2002. Carbonate sedimentation on subtropical shelves around Lord Howe Island and Balls Pyramid, Southwest Pacific. *Mar Geol.* 188: 333-349.
- Kirkpatrick M, Ravigné V. 2002. Speciation by natural and sexual selection: models and experiments. *Am Nat.* 159: S22-S35.
- Kisel Y, Barraclough TG. 2010. Speciation has a spatial scale that depends on levels of gene flow. *Am Nat.* 175: 316-334.
- Krug PJ. 2011. Patterns of speciation in marine gastropods: a review of the phylogenetic evidence for localized radiations in the sea. *Am Malacol Bull.* 29: 169-186.
- Liu S-C, Lu C-T, Wang J-C. 2009. Reticulate hybridization of *Alpinia* (Zingiberaceae) in Taiwan. *J Plant Res.* 122: 305-316.
- Losos JB. 1994. Integrative approaches to evolutionary ecology: *Anolis* lizards as model systems. *Annu Rev Ecol Syst.* 25: 467-493.
- Losos JB, Glor RE. 2003. Phylogenetic comparative methods and the geography of speciation. *Trends Ecol Evol.* 18: 220-227.
- Losos JB, Miles DB. 1994. Adaptation, constraint, and the comparative method: phylogenetic issues and methods. In: Wainwright PC, Reilly SM, editors. *Ecological Morphology: Integrative Organismal Biology*. Chicago: University of Chicago Press. p. 60-98.
- Losos JB, Schluter D. 2000. Analysis of an evolutionary species-area relationship. *Nature.* 408: 847-850.
- Mable BK. 2004. 'Why polyploidy is rarer in animals than in plants': myths and mechanisms. *Biol J Linn Soc.* 82: 453-466.
- Mallet J. 2007. Hybrid speciation. *Nature.* 446: 279-283.
- Mallet J. 2008. Hybridization, ecological races and the nature of species: empirical evidence for the ease of speciation. *Philos Trans R Soc Lond B.* 363: 2971.
- Mallet J, Meyer A, Nosil P, Feder JL. 2009. Space, sympatry and speciation. *J Evol Biol.* 22: 2332-2341.
- Mavárez J, Linares M. 2008. Homoploid hybrid speciation in animals. *Mol Ecol.* 17: 4181-4185.
- Mavarez J, Salazar CA, Bermingham E, Salcedo C, Jiggins CD, Linares M. 2006. Speciation by hybridization in *Heliconius* butterflies. *Nature.* 441: 868-871.

- Mayr E. 1963. *Animal Species and Evolution*. Cambridge, Massachusetts: Harvard University Press.
- Mayr E. 1982. Speciation and macroevolution. *Evolution*. 36: 1119-1132.
- McDougall I, Embleton BJJ, Stonea DB. 1981. Origin and evolution of Lord Howe Island, southwest Pacific Ocean. *Aust J Earth Sci*. 28: 155-176.
- McKinnon JS, Mori S, Blackman BK, David L, Kingsley DM, Jamieson L, Chou J, Schluter D. 2004. Evidence for ecology's role in speciation. *Nature*. 429: 294-298.
- Nosil P. 2012. *Ecological speciation*. Oxford: Oxford University Press.
- Nosil P. 2008. Speciation with gene flow could be common. *Mol Ecol*. 17: 2103-2106.
- Nosil P, Egan SP, Funk DJ. 2008. Heterogeneous genomic differentiation between walking-stick ecotypes: 'isolation-by-adaptation' and multiple roles for divergent selection. *Evolution*. 62: 316-336.
- Nosil P, Harmon LJ, Seehausen O. 2009. Ecological explanations for (incomplete) speciation. *Trends Ecol Evol*. 24: 145-156.
- Department of Environment and Climate Change NSW. 2007. *Lord Howe Island Biodiversity Management Plan*. Sydney: Department of Environment and Climate Change NSW.
- Orr HA. 1990. "Why Polyploidy is Rarer in Animals Than in Plants" Revisited. *Am Nat*. 136: 759-770.
- Ostevik KL, Moyers BT, Owens GL, Rieseberg LH. 2012. Parallel Ecological Speciation in Plants? *Int J Ecol*. 2012: 17.
- Otto SP, Whitton J. 2000. Polyploid incidence and evolution. *Annu Rev Genet*. 34: 401-437.
- Papadopoulos AST, Baker WJ, Crayn D, Butlin RK, Kynast RG, Hutton I, Savolainen V. 2011. Speciation with gene flow on Lord Howe Island. *Proc Natl Acad Sci USA*. 108: 13188-13193.
- Papadopoulos AST, Price Z, Devaux C, Hipperson H, Smadja CM, Hutton I, Baker WJ, Butlin RK, Savolainen V. 2012. A comparative analysis of the mechanisms underlying speciation on Lord Howe Island. *J Evol Biol*. forthcoming.
- Peccoud J, Ollivier A, Plantegenest M, Simon J-C. 2009. A continuum of genetic divergence from sympatric host races to species in the pea aphid complex. *Proc Natl Acad Sci USA*. 106: 7495-7500.
- Phillimore AB, Orme CDL, Thomas GH, Blackburn TM, Bennett PM, Gaston KJ, Owens IPF. 2008. Sympatric speciation in birds is rare: insights from range data and simulations. *Am Nat*. 171: 646-657.
- Pickard J. 1976. The effect of feral goats (*Capra hircus* L.) on the vegetation of Lord Howe Island. *Aust J Ecol*. 1: 103-114.
- Pickard J. 1984. Exotic plants on Lord Howe Island: distribution in space and time. *J Biogeogr*. 11: 181.
- Pickard J. 1983. Vegetation of Lord Howe Island. *Cunninghamia*. 1: 133-265.
- Priddel D, Carlile N, Humphrey M, Fellenberg S, Hiscox D. 2003. Rediscovery of the 'extinct' Lord Howe Island stick-insect (*Dryococelus australis* (Montrouzier)) (Phasmatodea) and recommendations for its conservation. *Biodivers Conserv*. 12: 1391-1403.
- Quesada H, Posada D, Caballero A, Morán P, Rolán-Alvarez E. 2007. Phylogenetic evidence for multiple sympatric ecological diversification in a marine snail. *Evolution*. 61: 1600-1612.

- Ramsey J, Bradshaw HD, Schemske DW. 2003. Components of reproductive isolation between the monkeyflowers *Mimulus lewisii* and *M. cardinalis* (Phrymaceae). *Evolution*. 57: 1520–1534.
- Ramsey J, Schemske DW. 1998. Pathways, mechanisms, and rates of polyploid formation in flowering plants. *Annu Rev Ecol Syst*. 29: 467–501.
- Remington DL, Robichaux RH. 2007. Influences of gene flow on adaptive speciation in the *Dubautia arborea*--*D. ciliolata* complex. *Mol Ecol*. 16: 4014–4027.
- Rieseberg LH, Raymond O, Rosenthal DM, Lai Z, Livingstone K, Nakazato T, Durphy JL, Schwarzbach AE, Donovan LA, Lexer C. 2003. Major ecological transitions in wild sunflowers facilitated by hybridization. *Science*. 301: 1211–1216.
- Rieseberg LH, van Fossen C, Desrochers AM. 1995. Hybrid speciation accompanied by genomic reorganization in wild sunflowers. *Nature*. 375: 313–316.
- Rieseberg LH, Willis JH. 2007. Plant speciation. *Science*. 317: 910–914.
- Rundle HD, Nagel L, Wenrick Boughman J, Schluter D. 2000. Natural selection and parallel speciation in sympatric sticklebacks. *Science*. 287: 306–308.
- Rundle HD, Nosil P. 2005. Ecological speciation. *Ecol Lett*. 8: 336.
- Savolainen V, Anstett M-C, Lexer C, Hutton I, Clarkson JJ, Norup MV, Powell MP, Springate D, Salamin N, Baker WJ. 2006a. Sympatric speciation in palms on an oceanic island. *Nature*. 441: 210–213.
- Savolainen V, Lexer C, Anstett M-C, Hutton I, Clarkson JJ, Norup MV, Powell MP, Springate D, Salamin N, Baker WJ. 2006b. Sympatric plant speciation in islands? (Reply). *Nature*. 443: E12–E13.
- Schliwen UK, Tautz D, Paabo S. 1994. Sympatric speciation suggested by monophyly of crater lake cichlids. *Nature*. 368: 629–632.
- Schluter D. 2001. Ecology and the origin of species. *Trends Ecol Evol*. 16: 372.
- Schluter D. 2000. *The Ecology of Adaptive Radiation*. Oxford: Oxford University Press.
- Schluter D. 2009. Evidence for ecological speciation and its alternative. *Science*. 323: 737–741.
- Schluter D, Conte GL. 2009. Genetics and ecological speciation. *Proc Natl Acad Sci USA*. 106: 9955–9962.
- Smith JF, Burke CC, Wagner WL. 1996. Interspecific hybridization in natural populations of *Cyrtandra* (Gesneriaceae) on the Hawaiian Islands: Evidence from RAPD markers. *Plant Syst Evol*. 200: 61–77.
- Sobel JM, Chen GF, Watt LR, Schemske DW. 2009. The biology of speciation. *Evolution*. 64: 295–315.
- Soltis DE, Soltis PS. 1999. Polyploidy: recurrent formation and genome evolution. *Trends Ecol Evol*. 14: 348–352.
- Soltis DE, Soltis PS, Schemske DW, Hancock JF, Thompson JN, Husband BC, Judd WS. 2007. Autopolyploidy in angiosperms: have we grossly underestimated the number of species? *Taxon*. 56: 13–30.
- Sorenson MD, Sefc KM, Payne RB. 2003. Speciation by host switch in brood parasitic indigobirds. *Nature*. 424: 928–931.
- Ståhlberg D. 2009. Habitat differentiation, hybridization and gene flow patterns in mixed populations of diploid and autotetraploid *Dactylorhiza maculata* s.l. (Orchidaceae). *Evol Ecol*. 23: 295–328.
- Stuessy TF. 2006a. Evolutionary biology: sympatric plant speciation in islands? *Nature*. 443: E12; discussion E12–13.

- Stuessy TF. 2006b. Sympatric plant speciation in islands? *Nature*. 443: E12.
- Stuessy TF, Jakubowsky G, Gómez RS, Pfosser M, Schlüter PM, Fer T, Sun BY, Kato H. 2006. Anagenetic evolution in island plants. *J Biogeogr*. 33: 1259-1265.
- Szymura JM, Barton NH. 1991. The genetic structure of the hybrid zone between the fire-bellied toads *Bombina bombina* and *B. variegata*: comparisons between transects and between loci. *Evolution*. 45: 237-261.
- Tregenza T, Butlin RK. 1999. Speciation without isolation. *Nature*. 400: 311.
- van Doorn GS, Noest AJ, Hogeweg P. 1998. Sympatric speciation and extinction driven by environment dependent sexual selection. *Proc R Soc B*. 265: 1915-1919.
- Wachowiak W, PalmÉ AE, Savolainen O. 2011. Speciation history of three closely related pines *Pinus mugo* (T.), *P. uliginosa* (N.) and *P. sylvestris* (L.). *Mol Ecol*. 20: 1729-1743.
- Wood TE, Takebayashi N, Barker MS, Mayrose I, Greenspoon PB, Rieseberg LH. 2009. The frequency of polyploid speciation in vascular plants. *Proc Natl Acad Sci USA*. 106: 13875-13879.
- Woodroffe CD, Kennedy DM, Brooke BP, Dickson ME. 2006. Geomorphological evolution of Lord Howe Island and carbonate production at the latitudinal limit to reef growth. *J Coastal Res*. 22: 188-201.
- Woodroffe CD, Kennedy DM, Jones BG, Phipps CVG. 2004. Geomorphology and Late Quaternary development of Middleton and Elizabeth Reefs. *Coral Reefs*. 23: 249-262.
- Zheng X-M, Ge S. 2010. Ecological divergence in the presence of gene flow in two closely related *Oryza* species (*Oryza rufipogon* and *O. nivara*). *Mol Ecol*. 19: 2439-2454.

Chapter 3

SPECIATION OF ARABIAN GAZELLES

**Hannes Lerp¹, Torsten Wronski^{2,3},
Thomas M. Butynski^{2,3} and Martin Plath¹**

¹Evolutionary Ecology Group, J.W. Goethe–University Frankfurt,
Frankfurt am Main, Germany

²Conservation Programmes, The Zoological Society of London,
London, UK

³King Khalid Wildlife Research Centre, Saudi Wildlife Authority,
Riyadh, Kingdom of Saudi Arabia

ABSTRACT

Gazelles are distributed across Africa and Asia and are adapted to arid and semi-arid environments. In this chapter, we discuss potential factors promoting the divergence of lineages within this group (*i.e.*, speciation events). The most recent common ancestor of gazelles is thought to have emerged during the Miocene (12–14 Ma) and to have split into the extant genera *Nanger* and *Eudorcas* (both endemic to Africa), *Antelope* (endemic to Asia), and *Gazella* (present in Africa, the Middle East and Asia). Within *Gazella*, two major clades are thought to have evolved allopatrically: (1) a predominantly Asian Clade (*G. bennettii*, *G. subgutturosa*, *G. marica*, *G. leptoceros*, and *G. cuvieri*) and (2) a predominantly African Clade (*G. dorcas*/ *G. saudiya*, *G. spekei*, *G. gazella*, and *G. arabica*). At present, both clades meet in North Africa and, especially, in Arabia. Other splits in this group are better explained by adaptive speciation in response to divergent ecological selection. In both clades, parallel evolution of sister species pairs (a desert-adapted form and a humid mountain-adapted form) can be inferred; desert-dwelling *G. dorcas* in Africa and *G. saudiya* in Arabia have a sister group relationship with mountain-dwelling *G. gazella* in the Levant and *G. arabica* in Arabia. This relationship exists within Africa between the desert-dwelling slender-horned gazelle (*G. leptoceros*) and the mountain-dwelling Cuvier's gazelle (*G. cuvieri*) of the Atlas Mountains. A third species pair occurs in Asia; desert-dwelling goitred gazelle (*G. subgutturosa*) and mountain-dwelling chinkara (*G. bennettii*). These (ecological) speciation events correlate with ecology and behavior: the mountain forms being browsers, sedentary, territorial, and living in small groups, while the desert forms are grazers, migratory/ nomadic, non-territorial, and living in herds. Furthermore, cryptic sister species (*G. gazella*, *G. arabica*), with strikingly similar phenotypes, exist within presumed '*G. gazella*', alluding to a possible

allopatric origin of this divergence following an isolation of humid mountain regions during hyper-arid phases. On the other hand, phenotypes within *G. arabica* tend to be variable, but are difficult or impossible to distinguish genetically.

INTRODUCTION

The earliest known fossil antelope, found in the Baringo Basin, Kenya, is of early Miocene origin (Thomas 1981), but it is uncertain where gazelline antelopes first emerged: in Africa, as proposed by Kingdon (1988), or in Asia, as suggested by Vrba and Schaller (2000). One of the four extant genera (*Antilope*) that evolved from these ancestors is in Asia, two (*Nanger* and *Eudorcas*) are in Africa, and one (*Gazella*) is on both continents as well as on the Arabian Peninsula. To understand the present distribution of gazelles (*i.e.*, *Antilope*, *Gazella*, *Nanger* and *Eudorcas*) it is important to interpret results from phylogenetic analyses in light of the geological and climatological history of the entire historic ranges of these genera.

The Arabian Peninsula is a prime example of a biogeographic transition zone, as it connects the floral and faunal regions of Africa and Asia (Vincent 2008). The present pattern is predominantly the result of the Afro-Eurasian species interchanges following the joining of the northern edge of the Afro-Arabian continent with Eurasia in the mid-Oligocene (ca. 30 Ma; Tchernov 1988; Bosworth et al. 2005; Vincent 2008). At the beginning of the Neogene (23 Ma), the Tethys acted as a substantial geographic barrier between Eurasian and Afro-Arabian faunas (Bernor 1983), leading to great divergence between these two realms and the evolution of two unique biotas (Tchernov 1988). The first major faunal interchange between Eurasia and Africa took place at the Proboscidean Datum Event (ca. 20 Ma; Madden and Van Couvering 1976), when a new land bridge (*i.e.*, the Gomphotherium Land Bridge; Rögl 1999a) connected Africa and Asia, and the Mediterranean Sea was isolated from the Indian Ocean for the first time.

The following faunal exchange was not continuous though, and was intensified during two main dispersal events in the Miocene at ca. 18–19 Ma and ca. 16–17 Ma (Thomas 1985; Rögl 1999b), interrupted by a re-opening of the seaway between Arabia and South Anatolia (Rögl 1999a). Evidence for faunal exchange during the first phase can be found in the Jibal Hadruk formation in Saudi Arabia (about 19 Ma) which contains fossil representatives of north-Tethyan fauna (Rögl and Steininger 1983). Before the connection to Eurasia was formed, the Arabian Peninsula supported an African fauna. After connection, Palearctic faunal elements appear in Arabia (Tchernov 1988; Delany 1989). During the early Miocene, extensive rifting of the Rift Valley resulted in a dramatic increase in water depth of the Red Sea, thus separating Arabia from Africa (Bosworth et al. 2005). During this period, savannah and steppe ecosystems expanded, leading to a radiation of grasses (Poaceae) followed by the rise of hypsodont ungulates (Strömberg 2011) and a rapid radiation of several tribes of bovids (Matthee and Robinson 1999). Although Africa became seemingly isolated from the northern hemisphere by the Saharo-Arabian arid belt in the late Miocene, faunal exchange of mammals increased once savannah- and desert-adapted forms evolved and the arid belt became a less effective barrier to the dispersal of such species (Thomas 1979; Tchernov 1988).

In the late Miocene/ early Pliocene era, the “savannah-mosaic” assemblages of Mesopotamia were already populated with representatives of the tribe Antilopini (*e.g.*, *Gazella deperdita* and *G. rodleri*) and other ungulates (Bernor 1986). The Miocene/ Pliocene boundary

was characterized by the onset of the Messinian Salinity Crisis (6 Ma), when the Mediterranean Sea became isolated from the Atlantic ocean and water levels regressed dramatically (Krijgsman et al. 1999). This resulted in an accelerated faunal interchange between Africa and Eurasia (*e.g.*, Agusti et al. 2006), especially of savannah-adapted species (Hassanin and Douzery 1999).

Following the Messinian Salinity Crisis the Mediterranean Sea reconnected to the Atlantic Ocean (Hilgen and Langereis 1988). At the same time, there was inflow of marine water into the Red Sea through the Bab el-Mandeb Strait, severing the connection between Arabia and the Horn of Africa (Bosworth et al. 2005). Furthermore, the orogeny of the Zagros Mountains—as part of the Alpine-Himalayan Mountain Belt—hampered biotic exchanges between Arabia and Asia (Tchernov 1988). All these factors led to an increasing isolation of Arabian fauna from Africa and Eurasia.

Moreover, the Afro-Arabian land bridge via the Sinai Peninsula became less permeable to faunal exchange due to a pull-apart basin development along the Aqaba-Levant Transform Fault (Bosworth et al. 2005). It remains uncertain as to whether there was a reconnection of both regions via the Bab el-Mandeb after the Miocene (Wildman et al. 2004; Winney et al. 2004; Fernandes et al. 2006; Bailey et al. 2009; Fernandes 2009). Climatic conditions during this time are thought to have caused a small-scale mosaic of ecosystems in the region (Tchernov 1988). Especially in Africa, faunal and palaeo-climatic records indicate shifts towards increasingly variable (and, on average, drier) conditions during the Plio-/ Pleistocene (2.8 Ma), allowing arid-adapted taxa to become more abundant (Thomas 1979; DeMenocal 2004).

In the Pleistocene the biotic interchange between Arabia and the Sahara was more asymmetric. Asian species, being more adapted to moister (mountainous) conditions, dispersed more easily into Arabia and North Africa along the mountain ridges of Arabia and the Sinai. By contrast, for arid-adapted Saharan species it was more difficult to invade the more humid parts of Asia (Delany 1989). Firstly, Saharan species on their way to Asia needed to cross the Nile Delta, which developed after the Messinian Salinity Crisis (Stanley and Warne 1998). Secondly, only the narrow stretch of sand dunes along the northern Sinai served as a suitable dispersal corridor for species adapted to hyper-arid conditions (Ferguson 1981). In addition, dispersing species would have needed to cross the Aqaba-Levant Transform Fault on their passage from the eastern Mediterranean towards Asia (Tchernov 1988).

The coastal plains of Arabia and the Sinai Peninsula experienced eustatic sea-level fluctuations, and large parts were submerged during inter-glacial periods (Chappell and Shackleton 1986; Shackleton 1987; van Andel 1989; Lambeck and Chappell 2001). During the Holocene (*i.e.*, after the glacial cycles) the geological situation remained more or less stable, and mammalian species in Arabia and surrounding areas—particularly gazelles—were increasingly impaired by human activities. Archeological evidence suggests that hunting by humans in prehistoric times was already having a major impact on populations of gazelles (Legge and Rowley-Conwy 1987; Bar-Oz et al. 2011).

In this section, we have provided a brief overview of the geological and climatological setting in which the evolution of extant gazelle species took place. In the following, we consider the question of how the above-mentioned factors influenced speciation in this group. We concentrate on possible scenarios for the modes of speciation, and discuss evidence for both allopatric and ecological speciation.

Table 1. List of specimens included in the phylogenetic analyses, their collectors/ accession numbers, and source of sequence. Abbreviations: EEZA – Estación Experimental de Zonas Áridas, Almería, Spain; KKWRC – King Khalid Wildlife Research Centre; OCE – Office for Conservation of the Environment, Muscat, Oman; WASWC - Wadi Al-Safa Wildlife Centre, Dubai, United Arab Emirates

Species	Origin	Collector/ accession number	Source	Group
<i>Gazella arabica</i>	Oman, Muscat – Sur	OCE	tissue	African <i>Gazella</i>
<i>G. arabica</i>	Farasan Islands, Saudi Arabia	JN410353	GenBank (Lerp et al. 2011)	African <i>Gazella</i>
<i>G. arabica</i>	Israel, A'rava Valley	KC188759	GenBank (Lerp et al. 2013)	African <i>Gazella</i>
<i>G. bennettii</i>	KKWRC (ancestors from Pakistan)	JN410340	GenBank (Lerp et al. 2011)	Asian <i>Gazella</i>
<i>G. bennettii</i>	KKWRC (ancestors from Iran)	JN410341, JN410357	GenBank (Lerp et al. 2011)	Asian <i>Gazella</i>
<i>G. bennettii</i>	Pakistan	KKWRC	blood, hairs	Asian <i>Gazella</i>
<i>G. cuvieri</i>	EEZA	JN410342, JN410343	GenBank (Lerp et al. 2011)	Asian <i>Gazella</i>
<i>G. dorcas</i>	Israel, A'rava Valley	JN410230	GenBank (Lerp et al. 2011)	African <i>Gazella</i>
<i>G. dorcas</i>	Chad	JN410237	GenBank (Lerp et al. 2011)	African <i>Gazella</i>
<i>G. dorcas</i>	Sudan, Mashail	JN410250	GenBank (Lerp et al. 2011)	African <i>Gazella</i>
<i>G. dorcas</i>	Algeria, Hoggar Mountains	JN410252	GenBank (Lerp et al. 2011)	African <i>Gazella</i>
<i>G. gazella</i>	Israel, Yehuda Mountains	KC188775	GenBank (Lerp et al. 2013)	African <i>Gazella</i>
<i>G. gazella</i>	Israel, Shomeron	KC188774	GenBank (Lerp et al. 2013)	African <i>Gazella</i>
<i>G. leptoceros</i>	Hoggar Mountains, Algeria	JN410259	GenBank (Lerp et al. 2011)	Asian <i>Gazella</i>
<i>G. leptoceros</i>	Tunisia	JN410345	GenBank (Lerp et al. 2011)	Asian <i>Gazella</i>
<i>G. leptoceros</i>	Western Desert, Egypt	JN410346	GenBank (Lerp et al. 2011)	Asian <i>Gazella</i>
<i>G. marica</i>	Syria, Dara Region	K. Habibi	hairs	Asian <i>Gazella</i>
<i>G. marica</i>	Saudi Arabia, Khunfah	KKWRC	tissue	Asian <i>Gazella</i>
<i>G. marica</i>	Saudi Arabia, Uruq Bani ma' Arid	S. Ostrowski	hairs	Asian <i>Gazella</i>
<i>G. spekei</i>	WASWC	D. O'Donovan	hairs	African <i>Gazella</i>
<i>G. subgutturosa</i>	Mongolia, south	D. Maaz	tissue	Asian <i>Gazella</i>
<i>G. subgutturosa</i>	unknown	AF036282	GenBank (Hassanin and Douzery 1999)	Asian <i>Gazella</i>
<i>Antelope cervicapra</i>	unknown	AF022058, AF036283	GenBank (Matthee and Robinson 1999; Hassanin and Douzery 1999)	Larger gazelles
<i>Eudorcas thomsoni</i>	unknown	FJ556559	GenBank (Tungsudjai et al. unpublished)	Larger gazelles
<i>E. rufifrons</i>	Sudan	JN632633, JN632634	GenBank (Hassanin et al. 2012)	Larger gazelles
<i>Nanger dama</i>	unknown	AF025 954	GenBank (Matthee and Robinson 1999)	Larger gazelles
<i>N. granti</i>	unknown	AF034723	GenBank (Hassanin et al. 1998)	Larger gazelles
<i>N. soemmerringii</i>	Egypt, Cairo, Giza Zoo	KC188777	GenBank (Lerp et al. 2013)	Larger gazelles
<i>N. soemmerringii</i>	Saudi Arabia, Jenadriyah, private collection	H. Tatwany	blood	Larger gazelles
<i>Litocranius walleri</i>	unknown	AF249974	GenBank (Matthee and Davis 2001)	outgroup
<i>L. walleri</i>	Somalia	JN632653	GenBank (Hassanin et al. 2012)	outgroup
<i>Antidorcas marsupialis</i>	unknown	AF022054, AF036281	GenBank (Matthee and Robinson 1999; Hassanin and Douzery 1999)	outgroup

MAJOR CLADES OF GAZELLES

Gazelles are members of the tribe Antilopini. Although the other members of this tribe are not part of this review, they are worth mentioning since they are considered to represent highly derived descendants of gazelle-like ancestors (Gentry 1992). Today the tribe Antilopini comprises as many as 13 genera (*Raphiceros*, *Ourebia*, *Madoqua*, *Dorcatragus*, *Saiga*,

Litocranius, *Ammodorcas*, *Antidorcas*, *Procapra*, *Eudorcas*, *Nanger*, *Antilope* and *Gazella*; Effron et al. 1976; Gentry 1992; Rebholz and Harley 1999; Groves 2000; Grubb 2005; Groves and Grubb 2011; Hassanin et al. 2012), four of which are traditionally labeled ‘true gazelles’, i.e., the genera *Gazella*, *Antilope*, *Nanger*, and *Eudorcas* (von Boetticher 1953; Groves 1985, 1988, 2000; Groves and Grubb, 2011).

To infer the phylogeny of gazelles, we investigated sequence variation of the mitochondrial cytochrome *b* gene of 17 taxa (including newly sequenced and already published data, see Table 1) covering all four gazelle genera, as well as the genera *Antidorcas* (springbok) and *Litocranius* (gerenuk). Bayesian analysis was performed in BEAST 1.5.2 (Drummond and Rambaut, 2007); no outgroup was defined beforehand. We used molecular clock data estimates inferred for *Gazella dorcas*. For methodological details see Lerp et al. (2011). jModelTest 0.1.1 (Posada 2008) uncovered HKY + Γ as the best fitting substitution model. We ran a Metropolis coupled Monte Carlo Markov chain (MC3) for 15 million generations with a burn-in phase of 1.5 million generations.

The phylogenetic tree inferred from this analysis is shown in Figure 1. High statistical support [i.e., posterior probability (PP) greater than 0.9] was found for the monophyly of gazelles (i.e., the genera *Antilope*, *Eudorcas*, *Gazella* and *Nanger*), but our analysis could not unambiguously resolve whether *Antidorcas* or *Litocranius* is the extant sister genus to the gazelles. Our findings are congruent with the results from a recent phylogenetic investigation of the order Cetartiodactyla by Hassanin et al. (2012), who analyzed the complete mitochondrial DNA sequence information, but included fewer gazelle taxa. Within the gazelles, all four genera were well supported as forming monophyletic clades, although the exact relationship among those genera could not be resolved. Time estimates for the first emergence of gazelles (95% credibility interval: 10.5–6.3 Ma), based on a molecular clock, were statistically not well supported (PP = 0.68), but are comparable to those provided by Hassanin et al. (2012), who estimated 8.5 ± 1.3 Ma (mean $\pm$ SD) for the corresponding phylogenetic split. During this time (i.e., in the late Miocene) savannah and steppe ecosystems with xerophytic shrub-land expanded into eastern and northern Africa and onto Arabia (Pound et al. 2011). This expansion of grasslands, together with the subsequent diversification of grasses (Strömberg 2011), probably facilitated the remarkable diversification (i.e., radiation) of antelopes at this time.

In contrast to paleontological studies describing the earliest fossil Antilopini from the middle Miocene in Africa (14 Ma; Vrba 1985) our molecular estimates for the first appearance of gazelles are considerable younger (10.5–6.3 Ma). How can these contrasting findings be reconciled? First of all, phylogenetic analyses through the analysis of sequence variation are based on extant taxa only, so extinct clades typically go undetected unless analyses of ancient DNA are feasible. Also, inference of time estimates from molecular phylogenetic approaches—as was done in this study—depend on the settings (i.e., substitution model and rates) for the molecular clock. Here, we used no fossil calibration points as constraints for our analysis (see below), but found similar time estimates as described by Hassanin et al. (2012), who used six calibration points from the fossil record for estimating the diversification of the entire order Cetartiodactyla. We are, therefore, confident that the settings of the molecular clock used in this study were realistic. Secondly, the classification of fossils is based on morphological measurements, especially with respect to skull and horn morphology. Gazelles show character state combinations that are likely plesiomorphic for the entire subfamily Antilopinae or, perhaps, even for the entire family Bovidae, which first appeared in the early Miocene (Gentry

1992; Vrba and Schaller 2000). Such morphological parallelisms and the incomplete fossil record render taxonomy within the Bovidae difficult (Vrba 1985; Gentry 1992). In addition, some bovid fossils showing this plesiomorphic character state combination are likely misclassified and falsely described as belonging in the vicinity of the genus *Gazella*.

The divergence of gazelles (PP = 1; 95% credibility interval: 8.0–4.8 Ma)—ultimately leading to the four extant genera in a relatively short period of time (Fig. 1)—could have been promoted by climate change following the Messinian Salinity Crisis (~6 Ma). Conditions were generally dryer (DeMenocal 2004), and new and larger areas became inhabitable for arid-adapted antelopes. The ancestors of the genus *Antilope* seem to have reached Asia by this time (Khan et al. 2006). The occurrence of blackbuck (*Antilope cervicapra*)—the only extant species of this genus—is still restricted to the Indian subcontinent and might be a descendant of this first expansion wave. Today, the descendants of *Eudorcas* and *Nanger* occur exclusively in Africa, and it remains doubtful if these genera ever occurred outside Africa.

The situation within genus *Gazella*, however, is more complex, because extant species occur both in Africa and Asia, as well as in Arabia (Kingdon 1988; Gentry 1992). Two major clades, with a well-supported monophyly, are inferred by our present study; their split is estimated at 3.9–2.3 Ma, i.e., in the Pliocene (Fig. 1). The ‘African Clade’, comprises more species, endemic to Africa, whereas the ‘Asian Clade’ is predominantly in Asia. Both clades, however, comprise taxa that occur on the “opposite” continent (Fig. 1).

The African Clade contains Speke’s gazelle (*G. spekei*), which is endemic to the Horn of Africa in Somalia (East 1999), dorcas (*G. dorcas*), mountain (*G. gazella*) and Arabian gazelles (*G. arabica*; Effron et al. 1976; Rebholz and Harley 1999; Wronski et al. 2010; Bärmann et al. 2012; this study). The diversification of the African Clade started 2.8–1.6 Ma ago (early Pleistocene; Fig. 1). By far the widest distribution range within this clade is that of *G. dorcas*, which includes large parts of northern Africa and, once, much of Arabia (where described as Saudi gazelle *G. saudiya*; Carruthers and Schwarz 1935; Rebholz et al. 1991; Rebholz and Harley 1997; Hammond et al. 2001). Together with *G. gazella* and *G. arabica*—which also inhabit the Arabian Peninsula—this is the most eastern extent of the range of the African Clade. We hypothesize that *G. dorcas* represents the ancestral character state combination of the African Clade because cytogenetic and morphological data showed *G. dorcas* to be basal to several species within *Gazella* (Lowenstein 1986; Gentry 1992; Vassart et al. 1995b). Moreover, it is suggested that the Antilopini evolved as grazers in the open, semi-desert and desert habitats of Africa (Kingdon 1988; Hassanin et al. 2012) and that the dispersal into mountainous and more humid habitats represents a shift associated with speciation events. At the edges of its distribution range, *G. dorcas* seems to have split rapidly into *G. spekei* and *G. gazella*, leaving sister group relations of these three species unresolved. This diversification was probably the result of ‘ecological speciation’ (see below). Lerp et al. (2011) found support for the idea that *G. dorcas* colonized Arabia via the Sinai and not via the Bab el-Mandeb. Thus, great distance and the Red Seas likely separated the ancestors of today’s *G. arabica*, *G. gazella* and ‘*G. saudiya*’ of Arabia from Africa’s *G. spekei* and *G. dorcas*.

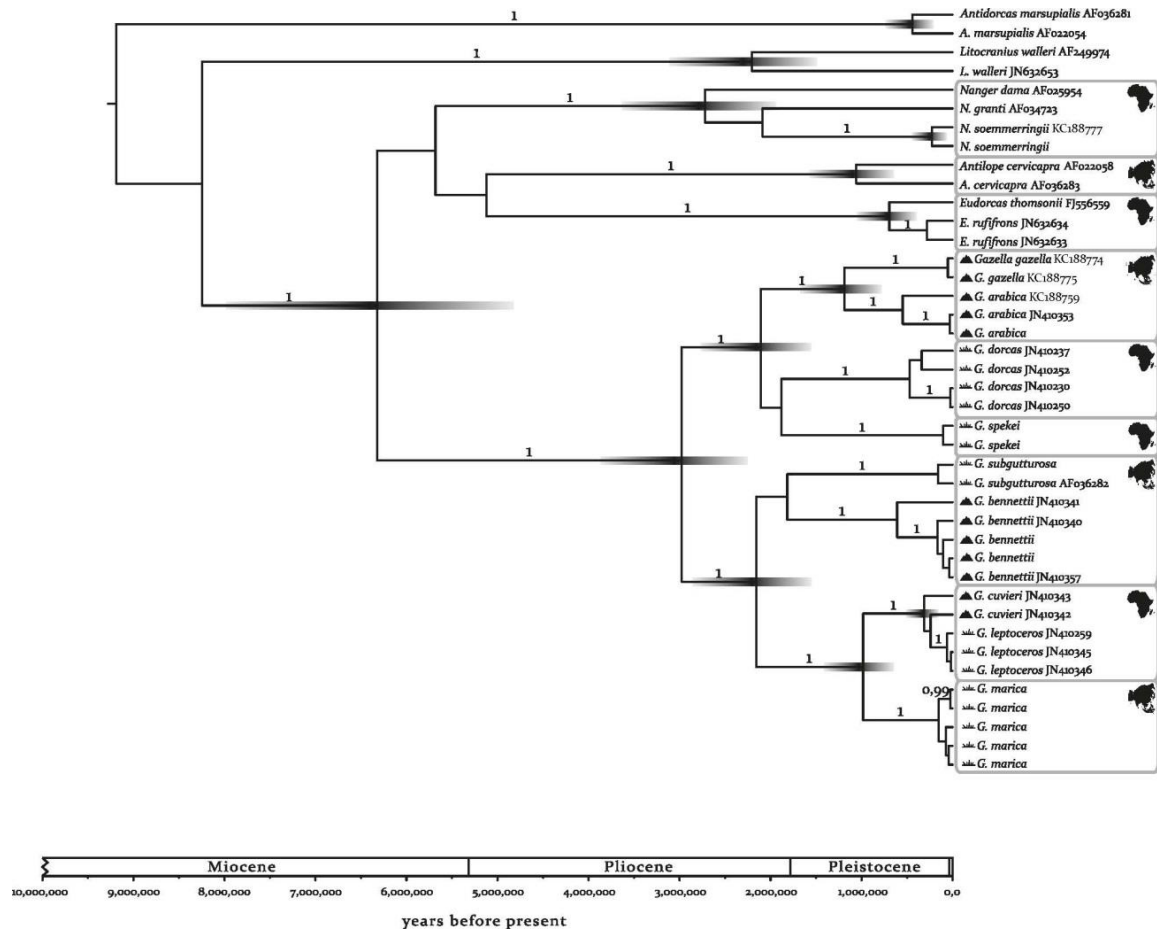


Figure 1. Phylogeny based on the alignment of the complete sequences of the cytochrome *b* gene. Bayesian analysis was performed with 41 sequences with the HKY + Γ substitution model. Only posterior probability values larger than 0.9 are reported. Node bars represent the 95% credibility intervals of the divergence times of statistically supported phylogenetic splits. Symbols of Africa and Asia indicate the occurrence of single species on that continent. Within the genus *Gazella*, a mountain symbol indicates a more humid and/or mountainous habitat, whereas a grass symbol indicates open savannah/ desert habitat.

Within the Asian Clade the majority of species are distributed on the Asian continent. The divergence time of this clade is estimated as 2.9–1.6 Ma ago and is comparable with the diversification of the African Clade of the genus *Gazella*. Therefore, the early Pleistocene is when most of today's species of *Gazella* emerged. We hypothesize that the diversification of the Asian Clade occurred in central Asia after the first (smaller) gazelles appeared, probably in the late Pliocene. The Asian Clade consists of *G. subgutturosa* and *G. bennettii*, both forming a reciprocally monophyletic clade in our present phylogeny. Both species occur in central Asia and India. Other members of the Asian Clade include *G. marica* and the African *G. cuvieri* and *G. leptoceros*, which together form a highly supported monophylum (Hammond et al. 2001; Wacher et al. 2010; Fig. 1). Changing climatic and geological conditions at the beginning of the Pleistocene could have enabled the ancestors of *G. marica* to cross the Zagros Mountains and invade the Middle East, where they occurred sympatrically with gazelles from the African Clade. Pliocene and early Pleistocene fossils of gazelles found in Turkey support this hypothesis, because they are distinct from fossils of *G. gazella* from the same period (Sickenberg 1975). Later (1.4–0.7 Ma ago) members of the Asian Clade crossed the Sinai Peninsula and Nile River to enter Africa and evolve into *G. leptoceros* which, today, occupies a habitat type similar to that of *G. marica* (i.e., the sand dunes and gravel plains of northern Africa; Harrison 1968; Devillers et al. 2005).

EXTANT SPECIES OF SMALLER GAZELLES

Before we elaborate on the mechanisms of speciation within the group of smaller gazelles (genus *Gazella*), we provide a brief overview of the historical and current distribution patterns as well as the current threats to the survival of the nine extant species in this group.

Dorcas gazelles (*G. dorcas*) were originally distributed from Morocco and Mauretania eastwards to the Horn of Africa, Sinai Peninsula, the Levant (Yom-Tov et al. 1995; East 1999; Hammond et al. 2001) to east of the Hejaz and Asir Mountains of western Arabia. This species was extirpated from Arabia about 30–40 years ago (Vesey-Fitzgerald 1952; Thouless et al. 1991; Habibi and Williamson 1997). With the exception of Israel and Ethiopia, numbers are decreasing rapidly and populations are increasingly fragmented (Smith 1999; Mallon and Kingswood 2001; Lafontaine et al. 2005). This decline is estimated at >30% over three generations, with less than 25% of the remaining animals living in protected areas, resulting in the IUCN status 'Vulnerable' (Mallon and Kingswood 2001; IUCN/SSC Antelope Specialist Group 2008a).

Mountain gazelles (*G. gazella*) are distributed from the eastern Turkey and Lebanon, through Palestine, Golan and western Jordan. Previously lumped with the Arabian gazelle (which we refer to as *G. arabica*—see below), which ranged over the Arava Valley in southern Israel, western Saudi Arabia, Yemen, Oman and United Arab Emirates. The number of *G. arabica* has declined dramatically during the past 50 years (Thouless and Al Bassri 1991; Magin and Greth, 1994; Mallon and Kingswood, 2001). Extensive hunting, habitat loss, and population fragmentation are principal causes of decline (Thouless et al., 1991; Magin and Greth 1994; Mallon and Kingswood 2001). The IUCN category is 'Vulnerable' based on *G. gazella* plus *G. arabica* (Mallon and Kingswood 2001; IUCN/SSC Antelope Specialist Group

2008b). The situation for *G. gazella* from northern and central Israel is less critical (Clark and Frankenberg 2001).

Speke's gazelle (*G. spekei*) is endemic to the Horn of Africa, occurring in Somalia from the Indian Ocean westwards to the Gulis Range (Heckel et al. 2008). Although traditionally not hunted by people, numbers have collapsed over the last 20 years due to uncontrolled hunting by soldiers (Heckel et al. 2008). Probably the species is extirpated from Ethiopia. No effective protection is in place for *G. spekei*. The IUCN status is 'Endangered' (Heckel et al. 2008).

Goitred gazelle (*G. subgutturosa*) occurs east of the Tigris/ Euphrates Basin, north into the Caucasus and across Iran into Turkmenistan. Following the steppes of central Asia, *G. subgutturosa* inhabits the Takla Makan, Tarim Basin and Sianking of China and extends farther eastwards to central Mongolia, where it is replaced by the Mongolian gazelle (*Procapra gutturosa*; Groves 1985; Kingswood and Blank 1996; Mallon and Kingswood 2001; Zachos et al. 2010). Large populations occurred over a vast area until recently with ca. 100,000 individuals in the early 1990's (Mallon and Kingswood 2001). Hunting and habitat loss have caused a decline of >30% over the last ten years in many populations, resulting in the IUCN status 'vulnerable' (Mallon 2008a). The example of Mongolia should be highlighted, since a substantial proportion of the global population of *G. subgutturosa* once lived there, but heavy poaching, following collapse of the communist regime, has eliminated most of the large herds, resulting in a population decline >50% (Mallon 2008a).

The chinkara (*G. bennettii*) occurs in western and central India (especially in the Thar Desert), in the arid regions of Baluchistan and Sindh Provinces in Pakistan, south-western Afghanistan and north-central Iran (Rahmani 1990, 2001; Habibi 2001; Karami et al. 2002; Mallon 2008b). Scattered populations are also found in the sub-mountainous tracts of Punjab (Roberts 1977; Habibi 2001). Although numbers in Pakistan and Iran are decreasing due to overhunting (Mallon 2008b), the population in India is >100,000 (Rahmani 2001). Despite the large number of people in India, antelope populations there are relatively stable. This is mainly the result of an extensive network of protected areas coupled with low hunting pressure (Mallon and Kingswood 2001). *G. bennettii*, in particular, is secure in the Thar Desert with 80,000 individuals (Rahmani 2001) and, furthermore, is protected in reserves or by local people (Mallon and Kingswood 2001). For these reasons, *G. bennettii* is the only *Gazella* sp. that is not threatened (IUCN status 'least concern'; Mallon 2008b).

The sand gazelle (*G. marica*) is found in open habitats of the Middle East from the Tigris/ Euphrates Basin in Iraq, through Jordan and Syria into southern Turkey, and southwards through much of Arabia (Wacher et al. 2010). Current distribution is limited to a few (protected) areas in the United Arab Emirates, Oman, Syria, Turkey, probably in Jordan and perhaps western Iraq (Kasperek 1986; Mallon and Kingswood 2001; Massolo et al. 2008). In Saudi Arabia, *G. marica* is probably extinct outside of two protected areas Mahazat as-Sayd and Uruq Bani Ma'arid, both of which harbor reintroduced populations (Cunningham and Wacher 2009). There are probably <10,000 mature individuals and the population trend is downwards (IUCN/SSC Antelope Specialist Group 2008c). A good number of *G. marica* occur in captivity and are available for re-introductions (Cunningham and Wacher 2009). The current IUCN status is 'Vulnerable' (IUCN/SSC Antelope Specialist Group 2008c).

Slender-horned gazelle (*G. leptoceros*) is endemic to the sand dunes (ergs) of the Sahara, west of the Nile River (Devillers et al. 2005). Until recently, two subspecies were distinguished, i.e., *G. l. loderi* from the sand deserts of Tunisia, Algeria and Libya, and *G. l. leptoceros* from the Western Desert in Egypt (Devillers et al. 2005). However, phylogeographic analyses for

validating these subspecies are lacking (Mallon et al. 2008). Numbers of *G. leptoceros* have decreased severely in the past decade due to hunting, especially in Egypt (Saleh 1987; Mostafa 2005), but also to habitat loss (Devillers et al. 2005). The conservation status of *G. leptoceros* in Mali, Niger, Chad and Libya is not known, but numbers are probably low (Devillers et al. 1999, 2005). All known populations are small to very small. The IUCN degree of threat status is ‘Endangered’ (Mallon et al. 2008).

Finally, Cuvier’s gazelle (*G. cuvieri*) is endemic to the Atlas Mountains and neighboring ranges in Morocco (including the lowlands in the west), Algeria and Tunisia (Lafontaine et al. 1999; Beudels-Jamar et al., 2005). As for most *Gazella* spp., hunting is the major threat to the species and has caused a sharp population decline since the 1930’s (Lafontaine et al. 1999; Beudels-Jamar et al. 2005). Habitat loss and degradation have also contributed to this decline (Sellami et al. 1990; de Smet 1991, 1994; Beudels-Jamar et al. 2005). There are currently ca. 2,500 individuals in several fragmented populations. The IUCN status is ‘Endangered’ (Mallon and Cuzin 2008). Some populations recently reported to be stable or even increasing (Mallon and Kingswood 2001; Mallon and Cuzin 2008).

PARALLEL, ADAPTIVE SPECIATION OF SPECIES PAIRS

Within both clades of genus *Gazella*, species pairs exist that exhibit parallel specializations in trophic ecology and social organization: On-the-one-hand, there are species more adapted to open, hot dry deserts. These species likely represent the ancestral character state combination. These species tend to be grazers, form herds and migrate. On-the-other-hand, species adapted to a more humid climate, are browsers that live in small groups and are sedentary and territorial. Our phylogenetic analysis infers three such species pairs (i.e., *G. dorcas* vs. *G. gazella* plus *G. arabica*; *G. subgutturosa* vs. *G. bennettii*, and *G. leptoceros* vs. *G. cuvieri*), where three lineages of desert-adapted forms independently diverged into a browsing, mountain-dwelling form, and a grazing, desert- or savannah-dwelling form. Even though we lack a plausible explanation as to how adaptation to different habitat types promoted reproductive isolation in gazelles, we argue that these three splits represent ecological speciation events.

Schluter and Nagel (1995) presented three—rather strict—prerequisites for parallel ecological speciation to occur; “(1) separate populations in similar environments must be phylogenetically independent [...], (2) ancestral and descendant populations [...] must be reproductively isolated, and (3) separate descendant populations inhabiting similar environments must not be reproductively isolated from one another”. This concept is particularly useful when considering contemporary parallel speciation within the same species, as indicated by the third criterion. When trying to apply the concept of parallel speciation to the phylogeny of gazelles it needs to be interpreted in a slightly broader sense. Point (3) of Schluter and Nagel’s (1995) definition is not met, as speciation in response to adaptation to a more humid climate occurred, independently, three times, and at different times, in different geographical regions, and from different ancestral species.

The oldest split of an ecologically divergent species pair inferred from our phylogenetic analysis is between *G. dorcas* and *G. gazella*/*G. arabica*. This split occurred 2.8–1.6 Ma (late Pliocene; Fig. 1). *G. dorcas* are grazers that inhabit Sahelian savannahs as well as semi-arid gravel and sand deserts, while avoiding hyper-arid areas and the upper elevations of the central-

Saharan massifs (Yom-Tov et al. 1995; Wachter et al. 2004). This species usually forms small family groups of 5–12 individuals (Yom-Tov et al. 1995), but during migration form herds of more than 100 individuals (Haltenorth and Diller 1977). *G. gazella* and *G. arabica*, by contrast, are sedentary, live in very small groups (two to maximal 20 individuals), live in upland areas of broken terrain on the Arabian Peninsula and the Levant, and adult males defend territories (Walther et al. 1983; Mendelssohn et al. 1995; Martin 2000; Wronski and Plath 2010). *G. dorcas* can cope without surface water by relying on hygroscopic food and respiratory water (Yom-Tov et al. 1995), whereas *G. gazella* and *G. arabica* prefer to drink on a regular basis (Mendelssohn et al. 1995). *G. dorcas* are reproductively isolated from *G. gazella*, and their hybrids are sterile or at least sub-fertile (Mendelssohn et al. 1995).

The second ecologically diverged species pair is *G. subgutturosa* and *G. bennettii*. Divergence probably occurred ca. 2.4–1.3 Ma ago (late Pliocene/ early Pleistocene) but statistical support for this date is weak (PP=0.89). *G. bennettii* are adapted to sand dune areas, regolith plains and hilly regions up to 1,500 m above sea level. This species avoids flat and steep terrain, and is typically on the edge of deserts (Roberts 1977; Sharma 1977; Rahmani 1990; Karami et al. 2002). *G. bennettii* are sedentary and live in groups of one to three individuals, but sometimes in larger herds (Rahmani 1990; Bagchi et al. 2008). Males form territories that they defend vigorously (Walther et al. 1983). The species is typically a browser, but during the rainy season they also graze (Sharma 1977; Habibi 2001). Compared to *G. subgutturosa*—which meet their water needs entirely from hygroscopic food plants—*G. bennettii* are independent of surface water only in winter. In the hotter months, when temperatures are >40°C, they have to drink regularly (Habibi 2001). *G. subgutturosa* are grazers that can also browse on xerophytic bushes (Roberts 1977; Kingswood and Blank 1996; Karami et al. 2002). This species is semi-nomadic with males forming territories only during the rut (i.e., October–December; Kingswood and Blank 1996; Blank 1998; Bekenov et al. 2001).

The youngest split of an ecologically divergent species pair is between *G. leptoceros* and *G. cuvieri* and dates to 420,000–110,000 years ago (middle Pleistocene). The young age of this split (i.e., the small genetic divergence between them) raises doubt concerning their species status (Hassanin et al. 2012). Nonetheless, both species are morphologically readily distinguished (Gentry 1964; Groves 1969; Groves and Grubb 2011). *G. leptoceros* are desert-dwelling grazers (Louys et al. 2011; Smith et al. 2001), that occasionally browse on *Acacia* (Saleh 2001). The species is nomadic, crossing vast areas of flat, open desert in search of sparse, ephemeral grasses (Kingdon 1997; Saleh 2001; Smith et al. 2001). The typical group size is <15 individuals (Smith et al. 2001). In contrast, *G. cuvieri* inhabit dry forests and maquis of the semi-arid Mediterranean type (Sellami and Bouredjli 1991; Beudels-Jamar et al. 2005), browse on acorns and young leaves of legumes, but also graze (Kingdon 1997; Smith et al. 2001). They live up to 2,600 m above sea level where they are limited by snow in winter (Aulagnier et al. 2001; Beudels-Jamar et al. 2005). *G. cuvieri* need to drink on a regular basis (Smith et al. 2001; Beudels-Jamar et al. 2005). This species lives in groups of 5–8 individuals, but solitary individuals are common (Sellami and Bouredjli 1991; Kingdon 1997; Beudels-Jamar et al. 2005). Males are territorial during the rut (in winter; Sellami and Bouredjli 1991; Kingdon 1997; Smith et al. 2001).

A TAXONOMIC REVIEW OF THE GENUS *GAZELLA*

It has been repeatedly emphasized that the taxonomy of gazelles is one of the least resolved among mammals (Groves and Harrison 1967; Groves, 1969). No other genus of large mammals creates such problems with regards to its classification based on skull morphometry, phenotypic appearance and genetic information, as does *Gazella*. As such, many taxonomic revisions of this genus have been put forth (Lydekker and Blaine 1914; Ellerman and Morrison-Scott 1951; von Boetticher 1953; Gentry 1964; Groves and Harrison 1967; Groves 1969, 1985, 1988; Lange 1972; Rostron 1972).

While the taxonomy of *Antelope* and *Nanger* has not changed substantially in recent decades (von Boetticher 1953; Gentry 1964; Lange 1972; Groves and Grubb 2011) the taxonomy within *Eudorcas* and *Gazella* remains uncertain, and with recent molecular findings casting doubt on earlier classifications that are based on morphological and cytogenetic traits. Our phylogenetic analysis of *Gazella* supports the existence of nine species (*G. gazella*, *G. arabica*, *G. dorcas*, *G. spekei*, *G. bennettii*, *G. subgutturosa*, *G. marica*, *G. leptoceros* and *G. cuvieri*), most of which require further taxonomic clarification.

Gazella marica (Thomas 1897), was subsumed within *Gazella leptoceros* by Ellerman and Morrison-Scott (1951). Subsequently, *G. marica* was considered a subspecies of *G. subgutturosa* based on morphological and karyological similarity (Groves and Harrison 1967; Kingswood et al. 1996, 1997). In more recent studies the phylogenetic relationships between *G. subgutturosa* from east of the Euphrates/ Tigris Basin and from the Arabia (*G. marica*) were reanalyzed based on molecular genetic information (Hammond et al. 2001; Wachter et al. 2010) and supported *G. marica* as a species. This conflicted with the grouping pattern inferred from skull structure and horn conformation (Groves and Harrison 1967). *G. marica* appears to be most closely related to the North African species *G. leptoceros* and *G. cuvieri* (Hammond et al. 2001; Wachter et al. 2010; see above).

In case of *G. subgutturosa*, Vassart et al. (1995b) state that *Gazella* will be paraphyletic when this species is included, because *G. subgutturosa* could be a sister taxon of *Antelope*. Both taxa share two unique centric fusions in their chromosomes causing the need to revive the genus *Trachelocele* (Ellerman and Morrison-Scott 1951; Groves 1969). Other studies investigating morphology or mitochondrial sequence variation placed *G. subgutturosa* within *Gazella* and refute *Trachelocele* (Grubb 2005; Groves and Grubb 2011; Hassanin et al. 2012; this study). Due to morphological variation within *G. subgutturosa*, up to three species are proposed by Groves and Grubb (2011), but there are no empirical data to support this position.

Early classifications place *G. bennettii* as either a subspecies of *G. gazella* (Haltenorth and Diller 1977; Roberts 1977) or as a subspecies of *G. dorcas* (Gentry 1964; Groves 1969; Lange 1972). Karyological data, however, found *G. bennettii* to be unrelated to *G. gazella* (Furley et al. 1988; Kumamoto et al. 1995). Within *G. bennettii*, up to six species are proposed on the basis of morphological divergence (Hemami and Groves 2001; Groves and Grubb 2011), but, again, evidence justifying this division is lacking. In this study—where two of the proposed *G. bennettii* taxa were included—there was no indication of more than one species. Nevertheless, a phylogeographic study with individuals from the entire distribution range is highly warranted.

In the cases of *G. cuvieri* and *G. leptoceros* the taxonomic classification remains confusing. Lange (1972) classified *G. cuvieri* under *G. gazella*, while *G. leptoceros* was considered a subspecies of *G. subgutturosa*. Later, a karyological study showed that *G. cuvieri* is unrelated

to *G. gazella* (Kumamoto and Bogart 1984). Furthermore, a division of *G. leptoceros* into two subspecies (*G. l. loderi* and *G. l. leptoceros*) was suggested based on differences in distribution ranges and ecology (Devillers et al. 2005). In contrast, *G. marica* and *G. leptoceros* are recently proposed to be subspecies of *G. cuvieri* because of their relatively low mitochondrial sequence divergence (Hassanin et al. 2012).

Within *G. dorcas*, several subspecies are described on the basis of phenotypic variation, such as coat coloration and horn shape and length (Groves 1969, 1981; Alados 1987; Yom-Tov et al. 1995; Groves and Grubb 2011). A phylogeographic study based on sequence variation of the mitochondrial cytochrome *b* gene and control region recently indicates that *G. dorcas*—including ‘*G. saudiya*’ and ‘*G. pelzelni*’—represent a reciprocally monophyletic group with a sister-group relationship to *G. gazella* and *G. arabica* (Lerp et al. 2011). No statistically significant support was found for any geographic structure within the distribution range of *G. dorcas*. Nevertheless, keeping *G. dorcas*, ‘*G. saudiya*’ and ‘*G. pelzelni*’ separated at captive breeding centers is warranted as low genetic divergence at neutral markers does not preclude the potential existence of local adaptations (Hammond et al. 2001; Lerp et al. 2011).

Confusion over taxonomy and nomenclature at the species level has reached a maximum in *G. gazella* and *G. arabica* (Groves and Harrison 1967; Harrison 1968; Groves 1969, 1983, 1989, 1996, 1997; Lange 1972; Groves and Lay 1985; Vassart et al. 1995a; Greth et al. 1996; Vassart et al. 1996; Kingswood et al. 1997; Rebholz and Harley 1999; Wronski et al. 2010). At least four species (*G. gazella*, *G. bilkis*, *G. arabica*, and *G. erlangeri*) and eight subspecies have been named (Groves 1996, 1997; Grubb 2005; Groves and Grubb 2011). Based on the analysis of cytochrome *b* sequences of five *G. gazella* in the context of a phylogeny of the Antilopinae, Rebholz and Harley (1999) suggested that two genetically distinct lineages might exist: one from the Levant (Galilee to Turkey) and one from Negev and Arabia. Those findings have been confirmed in an analysis comprising more individuals from a larger area and more mitochondrial and microsatellite markers (Wronski et al. 2010; Lerp et al. 2013). This supports recognition of two ‘cryptic’ species in this clade, which may have evolved due to prolonged isolation or local adaptations to divergent environments (Wronski et al., 2010; Lerp et al. 2013). The nominate *G. gazella* was originally described as *Antilope gazella* (Buffon 1764) from the Levant. This raises the question of which species name to assign to the populations in Arabia. Recent molecular analyses of the cytochrome *b* gene from the type *G. arabica* (described as *Antilope arabica* Lichtenstein 1827) indicate that this taxon is invalid, because skin and skull of the type specimen of *G. arabica* did not form a separate lineage, but clustered with *G. gazella* (skin) and with *G. arabica* (skull; Bärmann et al. 2012). Following the rules of precedence (priority rule, International Code of Zoological Nomenclature, ICZN) the name *G. arabica* is available for gazelles in Arabia.

Within *G. arabica*, however, much taxonomic uncertainty remains. One of the most challenging questions is the status of *G. erlangeri*. Neumann (1906) described specimens from Lahadsch (Lahej), north of Aden, as a greyer form of *G. arabica*. He introduced a new subspecies name to account for this difference and cited the illustration labeled *G. arabica* in Sclater and Thomas (1898) as an accurate representation of what he was describing. Due to its putative sympatric distribution with *G. arabica*, Groves (1996) suggested full species status for *G. erlangeri*. Gazelles currently kept in captivity at King Khalid Wildlife Research Centre in Saudi Arabia and at Al Wabra Wildlife Preservation in Qatar show the described combination of diagnostic features and thus, were considered to represent *G. erlangeri* (Groves 1996)—even though the provenance of these gazelles is not known. Phylogenetic studies (using

mitochondrial markers) on these putative *G. erlangeri* cluster them amongst other *G. arabica* from all over Arabia (Hammond et al. 2000; Blacket et al. 2001; Hundertmark and Omer 2004; Wronski et al. 2010). In summary, it remains unsolved whether Neumann's (1906) *G. erlangeri* is a distinct taxon and how it relates to other gazelles.

Finally, the most enigmatic gazelle described from Arabia should be mentioned: the Queen of Sheba's gazelle (*Gazella bilkis*). Specimens shot in the Taizz Mountains of southern Yemen in 1951 (now stored at Chicago FMNH) were originally identified as *G. arabica erlangeri* by the collector Hoogstraal. They were, however, re-evaluated retrospectively based on skull morphology and described as *Gazella bilkis* (Groves and Lay 1985; Groves and Grubb 2011). Even though the taxonomic status of these gazelles remains unclear, there is no doubt that *G. bilkis* is extinct (Mallon and Al-Safadi 2001).

CONCLUSION

Gazelles comprise four monophyletic genera (*Antilope*, *Nanger*, *Eudorcas* and *Gazella*) and emerged in the early Miocene (10.5–6.3 Ma). While three genera are restricted to the continent on which they probably evolved (*Antilope* to Asia, *Nanger* and *Eudorcas* to Africa), the situation in *Gazella* is more complex, with extant species in Africa, the Middle East, and Asia. Different modes of speciation are apparent within *Gazella*: (1) allopatric speciation in two major clades, with one predominantly Asian Clade and the other a predominantly African Clade; (2) parallel, adaptive speciation of three species pairs in parapatry, with one representative being a grazing, desert- or savannah-dwelling, (semi-)nomadic form, and the other being a browsing, mountain-dwelling and mostly sedentary form; and (3) cryptic speciation following phases of geographic isolation, where two genetically distinct forms with similar phenotypes can be seen (*G. gazella* and *G. arabica*). In general, gazelles are characterized by pronounced phenotypic variability that is not always mirrored by molecular sequence divergence, and a part of this variation may be due to phenotypic plasticity. This led to taxonomical incongruence plainly expressed in the number of described species that reached a maximum in a recent book by Groves and Grubb (2011), with 36 extant gazelle species (including 1 species in the genus *Antilope*, 5 in *Nanger*, 6 in *Eudorcas* and even 24 in *Gazella*) being listed. In terms of conservation this situation is unfortunate. The taxonomical incongruence hampers conservation efforts regarding captive breeding or re-introduction programs, as it remains confusing which gazelles should be bred separately to preserve natural biodiversity. Further investigations using nuclear DNA markers of the extant taxa will be helpful to clarify the situation for critical taxa.

ACKNOWLEDGMENTS

We thank His Highness, Prince Bandar bin Saud bin Mohammed al Saud (The President, Saudi Wildlife Authority) for his permission to conduct research on wildlife in the Kingdom. We are particularly grateful to the many colleagues and institutions who have contributed to the large number of samples collected from many sites in Africa, the Middle East and Asia. In this regard, we thank Timothy Wachter, Robert L. Hammond, Tamar Ron, Saeed Mubarak,

Khalid Alageel, Mohammed Shobrak, Osama B. Mohammed, Iyad Nader, Andrew Spalton, Roni King, Kris Hundertmark, Nick Lindsay, Doug Williamson, Jamal M. Mansoor, Denny Maaz, the National History Museum, London, the Harrison Museum, the Powell-Cotton Museum, the Field Museum, Chicago, the Steinhardt National Collection of Natural History, Zoological Museum at Tel-Aviv University, the Office for Conservation of the Environment, Muscat, Oman, the Estación Experimental de Zonas Áridas, Almeria, Spain and Al Wabra Wildlife Preservation, Al Wabra, Qatar. Furthermore, we thank the two Laboratory technicians at King Khalid Wildlife Research Centre (William Macasero and Benito Flores) for their considerable efforts in extracting DNA from most of our samples.

REFERENCES

- Agusti J, Garces M, Krijgsman W. 2006. Evidence for African–Iberian exchanges during the Messinian in the Spanish mammalian record. *Palaeogeography, Palaeoclimatology, Palaeoecology* 238:5–14.
- Alados CL. 1987. A cladistic approach to the taxonomy of the dorcas gazelles. *Israel Journal of Zoology* 34:33–49.
- Aulagnier S, Cuzin F, Loggers C, Thévenot M. 2001. Morocco. In: Mallon DP, Kingswood SC, editors. *Antelopes. Part 4: North Africa, the Middle East, and Asia*. Gland, Switzerland and Cambridge, UK: IUCN/SSC Antelope Specialist Group. p. 13–21.
- Bagchi S, Prakash Goyal S, Shankar K. 2008. Social organisation and population structure of ungulates in a dry tropical forest in western India (Mammalia, Artiodactyla). *Mammalia* 72:44–49.
- Bailey G, Al Sharekh A, Flemming N, Lambeck K, Momber G, Sinclair A, Vita-Vinzi C. 2009. Coastal prehistory in the southern Red Sea Bassin, underwater archaeology, and the Farasan Islands. *Proceedings of the Seminar for Arabian Studies* 37:1–16.
- Bärmann EV, Börner S, Erpenbeck D, Rössner GE, Hebel C, Wörheide G. 2012. The curious case of *Gazella arabica*. *Mammalian Biology*. In press.
- Bar-Oz G, Zeder M, Hole F. 2011. Role of mass-kill hunting strategies in the extirpation of Persian gazelle (*Gazella subgutturosa*) in the northern Levant. *Proceedings of the National Academy of Sciences of the United States of America* 108: 7345–7350.
- Bekenov AB, Blank DA, Grachev YA, Plakhov KN. 2001. Kazakhstan. In: Mallon DP, Kingswood SC, editors. *Antelopes. Part 4: North Africa, the Middle East, and Asia*. Gland, Switzerland and Cambridge, UK: IUCN/SSC Antelope Specialist Group. p. 134–140.
- Bernor RL. 1983. Geochronology and zoogeographic relationships of Miocene Hominoids. In: Ciochon RL, Corruccini R, editors. *New Interpretations of Ape and Human Ancestry*. New York: Plenum Press. p. 21–64.
- Bernor RL. 1986. Mammalian biostratigraphy, geochronology, and zoogeographic relationships of the late Miocene Maragheh fauna, Iran. *Journal of Vertebrate Paleontology* 6:76–95.
- Beudels-Jamar RC, Lafontaine R-M, Devillers P. 2005. *Gazella cuvieri*. In: Beudels-Jamar Rosaline C., Devillers P, Lafontaine R-M, Devillers-Terschuren J, Beudels MO, editors. *Sahelo-Saharan Antelopes. Status and Perspectives*. Report on the conservation status of

- the six Sahelo-Saharan Antelopes. CMS SSA Concerted Action. 2nd edition. Bonn, Germany: CMS Technical Series Publication N°11, UNEP/CMS Secretariat. p. 83–92.
- Blacket MJ, Macasero W, Flores B, Mohammed OB. 2001. *Report on mountain gazelle samples from Al Wabra*, Qatar examined at King Khalid Wildlife Research Centre, Thumamah.
- Blank DA. 1998. Mating behavior of the Persian Gazelle *Gazella subgutturosa* Gldenstaedt, 1780. *Mammalia* 62:499–520.
- von Boetticher H. 1953. Gedanken ber eine natrliche systematische Gruppierung der Gazellen (*Gazella*). [Thoughts about a natural systematic grouping of gazelles]. *Zeitschrift fr Sugetierkunde* 17:83–92.
- Bosworth W, Huchon P, McClay K. 2005. The Red Sea and Gulf of Aden Basins. *Journal of African Earth Sciences* 43:334–378.
- Buffon GLL. 1764. *Natural history, general and particular*. London: W. Strahan and T. Cadell
- Carruthers D, Schwarz E. 1935. On a new gazelle from central Arabia. *Proceedings of the Zoological Society of London* 1935:155–156.
- Chappell J, Shackleton NJ. 1986. Oxygen isotopes and sea level. *Nature* 324:137–140.
- Clark B, Frankenberg E. 2001. Israel (including West Bank and Gaza). In: Mallon DP, Kingswood SC, editors. *Antelopes. Part 4: North Africa, the Middle East, and Asia*. Gland, Switzerland and Cambridge, UK: IUCN/SSC Antelope Specialist Group. p. 107–111.
- Cunningham PL, Wachter T. 2009. Changes in the distribution, abundance and status of Arabian Sand Gazelle (*Gazella subgutturosa marica*) in Saudi Arabia: A review. *Mammalia* 73:203–210.
- Delany MJ. 1989. The zoogeography of the mammal fauna of southern Arabia. *Mammal Review* 19:133–152.
- DeMenocal PB. 2004. African climate change and faunal evolution during the Pliocene–Pleistocene. *Earth and Planetary Science Letters* 220:3–24.
- Devillers P, Beudels-Jamar RC, Lafontaine R-M, Devillers-Terschuren J. 2005. *Gazella leptoceros*. In: Beudels-Jamar R.C., Devillers P., Lafontaine RM, Devillers-Terschuren J., Beudels MO, editors. *Sahelo-Saharan Antelopes. Status and Perspectives*. Report on the conservation status of the six Sahelo-Saharan Antelopes. CMS SSA Concerted Action. 2nd edition. Bonn, Germany: CMS Technical Series Publication N°11, UNEP/CMS Secretariat. p. 71–82.
- Devillers P, Lafontaine R-M, Beudels-Jamar RC, Devillers-Terschuren J. 1999. Report on the status and perspectives of a species: *Gazella leptoceros*. In: UNEP/CMS, editor. *Conservation Measures for Sahelo-Saharan Antelopes*. Action Plan and Status Reports. Bonn, Germany: CMS Technical Series Publication N°4, UNEP/CMS Secretariat. p. 163–174.
- Drummond AJ, Rambaut A. 2007. BEAST: Bayesian evolutionary analysis by sampling trees. *BMC Evolutionary Biology* 7:214.
- East R. 1999. *African Antelope Database 1998*. Gland, Switzerland: IUCN/SSC Antelope Specialist Group
- Effron M, Bogart MH, Kumanoto AT, Benirschke K. 1976. Chromosom studies in the mammalian subfamily Antilopinae. *Genetica* 46:419–444.
- Ellerman JR, Morrison-Scott TCS. 1951. Checklist of palearctic and Indian mammals, 1758 to 1946. In: *Trustees of the British Museum*. London.

- Ferguson WW. 1981. The systematic position of *Gazella dorcas* (Artiodactyla: Bovidae) in Israel and Sinai. *Mammalia* 45:453–457.
- Fernandes CA, Rohling EJ, Siddall M. 2006. Absence of post-Miocene Red Sea land bridges: biogeographic implications. *Journal of Biogeography* 33:961–966.
- Fernandes CA. 2009. Bayesian coalescent inference from mitochondrial DNA variation of the colonization time of Arabia by the hamadryas baboon (*Papio hamadryas hamadryas*). In: Petraglia MD, Rose JJ, editors. The evolution of human populations in Arabia: Palaeoenvironments, *Prehistory and Genetics*. Dordrecht: Springer. p. 89–100.
- Furley CW, Tichy H, Uerpmann H-P. 1988. Systematics and chromosomes of the Indian gazelle, *Gazella bennetti* (Sykes, 1831). *Zeitschrift für Säugetierkunde* 53:48–54.
- Gentry AW. 1964. Skull characteristics of African gazelles. *Annales and Magazine of Natural History* 13:353–382.
- Gentry AW. 1992. The subfamilies and tribes of the family Bovidae. *Mammal Review* 22:1–32.
- Greth A, Schwede G, Vassart M, Granjon L. 1996. The controversial subject of gazelle subspecies in Saudi Arabia and its significance for reintroduction programmes. In: Greth A, Magin C, Ancrenaz M, editors. *Conservation of Arabian gazelles*. Riyadh: National Commission for Wildlife Conservation and Development. p. 51–68.
- Groves CP, Grubb P. 2011. *Ungulate taxonomy*. Baltimore: John Hopkins University Press
- Groves CP, Harrison DL. 1967. The taxonomy of the gazelles (genus *Gazella*) of Arabia. *Journal of Zoology*, London 152:381–387.
- Groves CP, Lay DM. 1985. A new species of the genus *Gazella* (Mammalia: Artiodactyla: Bovidae) from the Arabian peninsula. *Mammalia* 49:27–36.
- Groves CP. 1969. On the smaller gazelles of the Genus *Gazella* de Blainville, 1816. *Zeitschrift für Säugetierkunde* 37:38–60.
- Groves CP. 1981. Notes on the gazelles III: the dorcas gazelles of North Africa. *Annali del Museo Civico di Storia Naturale di Genova* 83:455–471.
- Groves CP. 1983. Notes on the gazelles IV: the Arabian gazelles collected by Hemprich and Ehrenberg. *Zeitschrift für Säugetierkunde* 48:371–381.
- Groves CP. 1985. *An introduction to the gazelles*. Chinkara 1:4–16.
- Groves CP. 1988. A catalogue of the genus *Gazella*. In: Dixon A, Jones D, editors. *Conservation and biology of desert antelopes*. London: Christopher Helm. p. 193–198.
- Groves CP. 1989. Gazelles of the Arabian peninsula. In: Abu-Zinada AH, Goriup PD, Nader IA, editors. *Proceedings of the First Symposium, Wildlife conservation and development in Saudi Arabia*, Riyadh, February 1987. Riyadh: National Commission for Wildlife Conservation and Development. p. 37–248.
- Groves CP. 1996. Taxonomic diversity in Arabian gazelles: the state of the art. In: Greth A, Magin C, Ancrenaz M, editors. *Conservation of Arabian gazelles*. Riyadh: National Commission for Wildlife Conservation and Development. p. 8–39.
- Groves CP. 1997. Taxonomy of Arabian gazelles. In: Habibi K, Abu-Zinada AH, Nader IA, editors. *The Gazelles of Arabia*. Riyadh: National Commission for Wildlife Conservation and Development. p. 24–51.
- Groves CP. 2000. Phylogenetic relationships within recent Antilopinae. In: Vrba ES, Schaller GB, editors. *Antelopes, deer and relatives – fossil record, behavioural ecology, systematics and conservation*. Yale University Press. p. 223–233.

- Grubb P. 2005. Order Artiodactyla. In: Wilson DE, Reeder DM, editors. *Mammal species of the world*, Vol. 1. Baltimore: John Hopkins University Press. p. 629–722.
- Habibi K, Williamson D. 1997. Conservation and management of Arabian gazelles. In: Habibi K, Abu-Zinada AH, Nader IA, editors. *The Gazelles of Arabia*. Riyadh: National Commission for Wildlife Conservation and Development. p. 231–241.
- Habibi K. 2001. Pakistan. In: Mallon DP, Kingswood SC, editors. *Antelopes. Part 4: North Africa, the Middle East, and Asia*. Gland, Switzerland and Cambridge, UK: IUCN/SSC Antelope Specialist Group. p. 122–128.
- Haltenorth T, Diller H. 1977. *A field guide to the mammals of Africa including Madagascar*. London: Collins.
- Hammond RL, Al Khaldy A, Macasero W, Flores B, Mohammed OB, Al Sagga M, Wachter TW, Bruford MW, Winney B. 2000. *The application of molecular genetics to the conservation of Arabian gazelles*. KKWRC report to King Abdulaziz City for Science and Technology, Riyadh.
- Hammond RL, Macasero W, Flores B, Mohammed OB, Wachter T, Bruford MW. 2001. Phylogenetic Reanalysis of the Saudi Gazelle and Its Implications for Conservation. *Conservation Biology* 15:1123–1133.
- Harrison DL. 1968. Genus *Gazella* in Arabia. In: Harrison DL, editor. *The mammals of Arabia: Carnivora, Hyracoidea, Artiodactyla*. Berlin: Ernest Benn. p. 349–380.
- Hassanin A, Delsuc F, Ropiquet A, et al. (11 co-authors) 2012. Pattern and timing of diversification of Cetartiodactyla (Mammalia, Laurasiatheria), as revealed by a comprehensive analysis of mitochondrial genomes. *Comptes Rendus Biologies* 335:32–50.
- Hassanin A, Douzery EJ. 1999. The tribal radiation of the family Bovidae (Artiodactyla) and the evolution of the mitochondrial cytochrome b gene. *Molecular Phylogenetics and Evolution* 13: 227–243.
- Hassanin A, Lecointre G, Tillier S. 1998. The “evolutionary signal” of homoplasy in protein-coding gene sequences and its consequences for a priori weighting in phylogeny. *Comptes Rendus De L Academie Des Sciences Serie Iii-Sciences De La Vie-Life Sciences* 321:611–620. [*Proceedings of the Academy of Sciences Series II-Life Sciences-Life Sciences* 321: 611–620.].
- Heckel J-O, Amir OG, Kaariye XY, Wilhelmi F. 2008. *Gazella spekei*. IUCN Red List of Threatened Species. Version 2011.2.
- Hemami MR, Groves CP. 2001. Iran. In: Mallon DP, Kingswood SC, editors. *Antelopes. Part 4: North Africa, the Middle East, and Asia*. Gland, Switzerland and Cambridge, UK: IUCN/SSC Antelope Specialist Group. p. 114–118.
- Hilgen F, Langereis C. 1988. The age of the Miocene-Pliocene boundary in the Capo Rossello area (Sicily). *Earth and Planetary Science Letters* 91:214–222.
- Hundertmark K, Omer S. 2004. *Genetic identification of gazelles from Al Wabra Wildlife Preservation, Doha, Qatar*. Riyadh.
- IUCN/SSC Antelope Specialist Group. 2008a. *Gazella dorcas*. IUCN Red List of Threatened Species. Version 2011.2.
- IUCN/SSC Antelope Specialist Group. 2008b. *Gazella gazella*. IUCN Red List of Threatened Species. Version 2011.2.
- IUCN/SSC Antelope Specialist Group. 2008c. *Gazella subgutturosa ssp. marica*. IUCN Red List of Threatened Species. Version 2011.2.

- Karami M, Hemami MR, Groves CP. 2002. Taxonomic, distributional and ecological data on gazelles in Iran. *Zoology in the Middle East* 26:29–36.
- Kasperek M. 1986. On the historical distribution and present situation of gazelles, *Gazella* spp., in Turkey. *Zoology in the Middle East* 1:11–15.
- Khan MA, Ghaffar A, Farooq U, Akhtar M. 2006. Ruminant Fauna from the Tertiary Hills (Neogene) of the Siwaliks of Pakistan. *Journal of Applied Sciences* 6:131–137.
- Kingdon J. 1988. East African Mammals: *An Atlas of Evolution in Africa*, Volume 3, Part D: Bovids. Chicago: University of Chicago Press.
- Kingdon J. 1997. *The Kingdon field guide to African mammals*. London: Academic Press.
- Kingswood SC, Blank DA. 1996. *Gazella subgutturosa*. *Mammalian Species* 518:1–10.
- Kingswood SC, Rebholz WER, Vassart M, Kumamoto AT. 1997. Cytogenetics and molecular genetics of Arabian gazelles: chromosomes, proteins and nucleic acids. In: Habibi K, Abu-Zinada AH, Nader IA, editors. *The Gazelles of Arabia*. Riyadh: National Commission for Wildlife Conservation and Development. p. 159–191.
- Kingswood SC, Vassart M, Williamson D. 1996. Systematics of *Gazella subgutturosa*: implications for conservation in the Arabian Peninsula. In: Greth A, Magin C, Ancrenaz M, editors. *Conservation of Arabian gazelles*. Riyadh: National Commission for Wildlife Conservation and Development. p. 104–111.
- Krijgsman W, Hilgen F, Raffi I, Sierro F, Wilsonk D. 1999. Chronology, causes and progression of the Messinian salinity crisis. *Nature* 400:652–655.
- Kumamoto AT, Bogart MH. 1984. The chromosomes of Cuvier's gazelle. In: Ryder OA, Byrd ML, editors. *One medicine: a tribute to Kurt Benirschke*. New York: Springer. p. 101–108.
- Kumamoto AT, Kingswood SC, Rebholz WER, Houck ML. 1995. The chromosomes of *Gazella bennetti* and *Gazella saudiya*. *Zeitschrift für Säugetierkunde* 60:159–169.
- Lafontaine R-M, Beudels-Jamar RC, Devillers P. 1999. Report on the status and perspectives of a species: *Gazella cuvieri*. In: UNEP/CMS, editor. *Conservation Measures for Sahelo-Saharan Antelopes*. Action Plan and Status Reports. Bonn, Germany: CMS Technical Series Publication N°4, UNEP/CMS Secretariat. p. 175–184.
- Lafontaine RM, Beudels-Jamar RC, Devillers P, Wachter T. 2005. *Gazella dorcas*. In: Beudels-Jamar RC, Devillers P, Lafontaine RM, Devillers-Terschuren J, Beudels MO, editors. *Sahelo-Saharan Antelopes*. Status and Perspectives. Report on the conservation status of the six Sahelo-Saharan Antelopes. Bonn: UNEP/CMS Secretariat. p. 93–108.
- Lambeck K, Chappell J. 2001. Sea level change through the last glacial cycle. *Science* 292:679–686.
- Lange J. 1972. Studien an Gazellenschädeln. Ein Beitrag zur Systematik der kleinen Gazellen, *Gazella* (de Blainville, 1816). [Studies on gazelle skulls. A contribution to the systematics of the small Gazelles, *Gazella* (de Blainville, 1816)]. *Säugetierkundliche Mitteilungen* 20:193–249.
- Legge AJ, Rowley-Conwy PA. 1987. Gazelle Killing in Stone-Age Syria. *Scientific American* 257:88–95.
- Lerp H, Wronski T, Pfenninger M, Plath M. 2011. A phylogeographic framework for the conservation of Saharan and Arabian Dorcas gazelles (Artiodactyla: Bovidae). *Organisms Diversity and Evolution* 11:317–329.
- Lerp H, Wronski T, Plath M, Schröter A, Pfenninger M. 2013. Phylogenetic and population genetic analyses confirm the existence of a distinct species boundary between Mountain

- (*Gazella gazella*) and Arabian Gazelles (*G. arabica*) in the Levant. *Mammalian Biology*. In press.
- Lichtenstein H. 1827. Darstellung neuer oder wenig bekannter Säugethiere. [Representation of new or little-known mammals]. In: Lüderitz CG, editor. *Abbildung und Beschreibung von fünfundsechzig Arten auf fünfzig colorierten Steindrucktafeln nach den Originalen des Zoologischen Museums der Universität zu Berlin*. Berlin: Humboldt Universität zu Berlin. [Illustration and description of sixty-five species on fifty colored lithographic plates according to the originals of the Zoological Museum of the University of Berlin. Berlin: Humboldt University of Berlin].
- Louys J, Meloro C, Elton S, Ditchfield P, Bishop LC. 2011. Mesowear as a means of determining diets in African antelopes. *Journal of Archaeological Science* 38:1485–1495.
- Lowenstein J. 1986. Molecular phylogenetics. *Annual Review of Earth and Planetary Sciences* 14:71–83.
- Lydekker R, Blaine G. 1914. Artiodactyla, Family Bovidae, Subfamilies Aepycerotinae to Tragelaphinae (Plea, Saiga, gazelles, Oryx group, bushbucks, kudus, elands, etc.), Antilocaprinae (prongbuck), and Giraffidae (giraffes and okapi). In: Vol. 3 of *Catalogue of the ungulate mammals in the British Museum*. London: British Museum of Natural History.
- Madden CT, Van Couvering JA. 1976. The Proboscidean Datum Event: Early Miocene migration from Africa. *Geological Society of America Abstracts with Programs* 8:992–993.
- Magin CD, Greth A. 1994. Distribution, status and proposals for the conservation of mountain gazelles *Gazella gazella cora* in the southwest of Saudi Arabia. *Biological Conservation* 70:69–75.
- Mallon DP, Al-Safadi M. 2001. Yemen. In: Mallon DP, Kingswood SC, editors. *Antelopes. Part 4: North Africa, the Middle East, and Asia*. Gland, Switzerland and Cambridge, UK: IUCN/SSC Antelope Specialist Group. p. 63–68.
- Mallon DP, Cuzin F, De Smet K, Hoffmann M. 2008. *Gazella leptoceros*. *IUCN Red List of Threatened Species*. Version 2011.2.
- Mallon DP, Cuzin F. 2008. *Gazella cuvieri*. *IUCN Red List of Threatened Species*. Version 2011.2.
- Mallon DP, Kingswood SC. 2001. *Antelopes. Part 4: North Africa, the Middle East, and Asia*. Global Survey and Regional Action Plans. Gland, Switzerland and Cambridge, UK: IUCN/SSC Antelope Specialist Group.
- Mallon DP. 2008a. *Gazella subgutturosa*. *IUCN Red List of Threatened Species*. Version 2011.2.
- Mallon DP. 2008b. *Gazella bennettii*. *IUCN Red List of Threatened Species*. Version 2011.2.
- Martin L. 2000. Gazelle (*Gazella* spp.) behavioural ecology: predicting animal behaviour for prehistoric environments in south-west Asia. *Journal of Zoology* 250:13–30.
- Massolo A, Spalton A, Al Lamki F. 2008. Notes on the status and conservation of rheem gazelle *Gazella subgutturosa marica* in the Sultanate of Oman. *Italian Journal of Zoology* 75:305–309.
- Matthee CA, Davis SK. 2001. Molecular insights into the evolution of the family Bovidae: a nuclear DNA perspective. *Molecular Biology and Evolution* 18:1220–30.
- Matthee CA, Robinson TJ. 1999. Cytochrome b phylogeny of the family Bovidae: resolution within the Alcelaphini, Antilopini, Neotragini, and Tragelaphini. *Molecular Phylogenetics and Evolution* 12: 31–46.

- Mendelssohn H, Yom-Tov Y, Groves CP. 1995. *Gazella gazella*. *Mammalian Species* 490:1–7.
- Mostafa H. 2005. Death of a Species? *Gnusletter* 23:12–16.
- Neumann O. 1906. *Über einige Gazellen und Kuh-Antilopen*. Sitzungsberichte der Gesellschaft Naturforschender Freunde zu Berlin 1906. [About some gazelles and cow antelopes. Meeting reports of the Society of Friends of Nature for Berlin 1906].
- Posada D. 2008. jModelTest: phylogenetic model averaging. *Molecular Biology and Evolution* 25:1253–1256.
- Pound MJ, Haywood AM, Salzmann U, Riding JB, Lunt DJ, Hunter SJ. 2011. A Tortonian (Late Miocene, 11.61–7.25Ma) global vegetation reconstruction. *Palaeogeography, Palaeoclimatology, Palaeoecology* 300:29–45.
- Rahmani AR. 1990. Distribution, density, group size and conservation of the Indian gazelle or chinkara *Gazella bennetti* (Sykes 1831) in Rajasthan, India. *Biological Conservation* 51:177–189.
- Rahmani AR. 2001. India. In: Mallon DP, Kingswood SC, editors. *Antelopes. Part 4: North Africa, the Middle East, and Asia*. Gland, Switzerland and Cambridge, UK: IUCN/SSC Antelope Specialist Group. p. 178–187.
- Rebholz WER, Harley EH. 1997. Cytochrome *b* Sequences from the Endangered Saudi Gazelle (*Gazella saudiya*) Suggest Hybridization with Chinkara (*G. bennetti*). *Conservation Biology* 11:251–255.
- Rebholz WER, Harley EH. 1999. Phylogenetic relationships in the bovid subfamily Antilopinae based on mitochondrial DNA sequences. *Molecular Phylogenetics and Evolution* 12:87–94.
- Rebholz WER, Williamson D, Rietkerk F. 1991. Saudi Gazelle (*Gazella saudiya*) Is Not a Subspecies of Dorcas Gazelle. *Zoo Biology* 10:485–489.
- Roberts TJ. 1977. *The mammals of Pakistan*. London: Ernest Benn
- Rögl F, Steininger FF. 1983. Vom Zerfall der Tethys zu Mediterran und Paratethys. [From the decay of the Tethys to Mediterranean and Paratethys]. *Annalen des Naturhistorischen Museums Wien Serie A* 85:135–163.
- Rögl F. 1999a. Short Note Mediterranean and Paratethys. Facts and Hypotheses of an Oligocene to Miocene Paleogeography (Short Overview). *Geologica Carpathica* 50:339–349.
- Rögl F. 1999b. Circum-Mediterranean Miocene paleogeography. In: Rössner GE, Heissig K, editors. *The Miocene land mammals of Europe*. München: F. Pfeil. p. 39–48.
- Rostron J. 1972. A multivariate study of skull measurements of five taxa of gazelles. *Biological Journal of the Linnean Society* 4:1–14.
- Saleh MA. 1987. The decline of gazelles in Egypt. *Biological Conservation* 39:83–95.
- Saleh MA. 2001. Egypt. In: Mallon DP, Kingswood SC, editors. *Antelopes. Part 4: North Africa, the Middle East, and Asia*. Gland, Switzerland and Cambridge, UK: IUCN/SSC Antelope Specialist Group. p. 48–54.
- Schluter D, Nagel LM. 1995. Parallel speciation by natural selection. *The American Naturalist* 146:292–301.
- Sclater PL, Thomas O. 1898. *The book of antelopes*, Volume 3, Part 12. London: R. H. Porter
- Sellami M, Bouredjli HA, Chapuis JL. 1990. *Repartition de la gazelle de Cuvier (Gazella cuvieri Ogilby, 1841) en Algérie*. *Vie Milieu* 40:234–237. [Distribution of the Cuvier's Gazelle (*Gazella cuvieri* Ogilby, 1841) in Algeria. *Life Middle* 40: 234–237].

- Sellami M, Bouredjli HA. 1991. *Preliminary data about the social structure of the Cuvier's gazelle, Gazella cuvieri* (Ogilby, 1841) of the reserve of Mergueb (Algeria). *Ongules/Ungulates* 91:357–360.
- Shackleton NJ. 1987. Oxygen isotops, ice volume and sea level. *Quaternary Science Reviews* 6:183–190.
- Sharma IK. 1977. Ecological study of habitat, feeding and survival of Indian gazelle *Gazella gazella* Pallas. *Journal of the Bombay Natural History Society* 74:347–350.
- Sickenberg O. 1975. Die Gliederung des höheren Jungtertiärs und Altquartärs in der Türkei nach Vertebraten und ihre Bedeutung für die internationale Neogen-Gliederung. [The structure of the higher Tertiary and Altquartärs in Turkey after Vertebraten and their importance for the international Neogen outline]. *Geologisches Jahrbuch* B15:1–167.
- de Smet K. 1991. *Cuvier's gazelle in Algeria*. *Oryx* 25:99–104.
- de Smet K. 1994. *Algeria and Tunisia*. *Gnusletter* 13:10.
- Smith TR, Mallon DP, De Smet K. 2001. Tunisia. In: Mallon DP, Kingswood SC, editors. *Antelopes. Part 4: North Africa, the Middle East, and Asia*. Gland, Switzerland and Cambridge, UK: IUCN/SSC Antelope Specialist Group. p. 30–40.
- Smith TR. 1999. Report on the status and perspectives of a species: *Gazella dorcas*. In: UNEP/CMS, editor. *Conservation Measures for Sahelo-Saharan Antelopes. Action Plan and Status Reports*. Bonn, Germany: CMS Technical Series Publication N°4, UNEP/CMS Secretariat. p. 185–201.
- Stanley D, Warne A. 1998. Nile Delta in its destruction phase. *Journal of Coastal Research* 14:794–825.
- Strömberg CAE. 2011. Evolution of Grasses and Grassland Ecosystems. *Annual Review of Earth and Planetary Sciences* 39:517–544.
- Tchernov E. 1988. The biogeographical history of the southern Levant. In: Yom-Tov Y, Tchernov E, editors. *The Zoogeography of Israel. The Distribution and Abundance at a Zoogeographical Crossroad*. Dordrecht, Boston, Lancaster: Dr W. Junk Publishers.
- Thomas H. 1979. Le rôle de barrière écologique de la ceinture Saharo-Arabique au Miocène; arguments paléontologiques. *Bulletin Muséum National d'Histoire Naturelle in Paris* 4:Sec. C, 2, 127–135.
- Thomas H. 1981. Les Bovidés miocène de la formation de Ngorora du Bassin de Baringo (Kenya). [The Miocene Bovids of the Ngorora Formation of the Baringo Basin (Kenya)]. *Proceedings Koninklijke Nederlandse Akademie van Wetenschappen* B84:335–409.
- Thomas H. 1985. The Early and Middle Miocene land connection of the Afro-Arabian plateau and Asia: a major event for hominoid dispersal? In: Delson E, editor. *Ancestors: The Hard Evidence*. New York: Alan R. Liss. p. 42–50.
- Thomas O. 1897. On a new gazelle from Central Arabia. *Annals and Magazine of Natural History* 6 19:162–163.
- Thouless CR, Al Bassri K. 1991. Taxonomic status of the Farasan island gazelle. *Journal of Zoology*, London 223:151–159.
- Thouless CR, Grainger JG, Shobrak M, Habibi K. 1991. Conservation Status of Gazelles in Saudi Arabia. *Biological Conservation* 58: 85–98.
- van Andel T. 1989. Late Quaternary sea level changes and archaeology. *Antiquity* 63:733–745.
- Vassart M, Granjon L, Greth A. 1995a. Genetic study of *Gazella gazella*: chromosomal and allozymic data. *Comptes Rendus De L Academie Des Sciences Serie Iii-Sciences De La Vie-Life Sciences* 318:27–33.

- Vassart M, Seguela A, Hayes H. 1995b. Chromosomal evolution in gazelles. *Journal of Heredity* 86: 216-227.
- Vassart M, Granjon L, Greth A. 1996. Cytogenetic and electrophoretic investigations of *Gazella gazella*. In: Greth A, Magin C, Ancrenaz M, editors. *Conservation of Arabian gazelles*. Riyadh: National Commission for Wildlife Conservation and Development. p. 94-103.
- Vesey-Fitzgerald LDEF. 1952. *Wild Life in Arabia*. Oryx 1:232-235.
- Vincent P. 2008. Saudi Arabia - *An Environmental Overview*. London: Taylor and Francis.
- Vrba ES, Schaller GB. 2000. Phylogeny based on behavior, gland, skull and Postcrania. In: Vrba ES, Schaller GB, editors. *Antelopes, deers and relatives*. Yale University Press. p. 203-222.
- Vrba ES. 1985. African Bovidae: evolutionary events since the Miocene. *South African Journal of Science* 81:263-266.
- Wacher T, Newby J, Monfort S, Tubiana J, Moksia D, Houston W, Dixon A. 2004. *Sahelo-Saharan interest group antelope update Chad 2001 and Niger 2002*.
- Wacher T, Wronski T, Hammond RL, et al. (12 co-authors) 2010. Phylogenetic analysis of mitochondrial DNA sequences reveals polyphyly in the goitred gazelle (*Gazella subgutturosa*). *Conservation Genetics* 12:827-831.
- Walther FR, Mungall EC, Grau GA. 1983. *Gazelles and their relatives. A study in territorial behavior*. New Jersey: Noyes Publication.
- Wildman DE, Bergman TJ, al-Aghbari A, Sterner KN, Newman TK, Phillips-Conroy JE, Jolly CJ, Disotell TR. 2004. Mitochondrial evidence for the origin of hamadryas baboons. *Molecular Phylogenetics and Evolution* 32: 287-296.
- Winney BJ, Hammond RL, Macasero W, Flores B, Boug A, Biquand V, Biquand S, Bruford MW. 2004. Crossing the Red Sea: phylogeography of the hamadryas baboon, *Papio hamadryas hamadryas*. *Molecular Ecology* 13: 2819-2827.
- Wronski T, Plath M. 2010. Characterization of the spatial distribution of latrines in reintroduced mountain gazelles (*Gazella gazella*): do latrines demarcate female group home ranges? *Journal of Zoology* (London) 280:92-101.
- Wronski T, Wacher T, Hammond RL, et al. (13 co-authors). 2010. Two reciprocally monophyletic mtDNA lineages elucidate the taxonomic status of Mountain gazelles (*Gazella gazella*). *Systematics and Biodiversity* 8: 119-129.
- Yom-Tov Y, Mendelssohn H, Groves CP. 1995. *Gazella dorcas*. *Mammalian Species* 491:1-6.
- Zachos FE, Karami M, Ibenouazi Z, Hartl GB, Eckert I, Kirschning J. 2010. First genetic analysis of a free-living population of the threatened goitered gazelle (*Gazella subgutturosa*). *Mammalian Biology* 5:277-282.

Chapter 4

CHROMOSOME PLASTICITY, ADAPTATION AND SPECIATION IN MALARIA MOSQUITOES

Igor V. Sharakhov*

Department of Entomology, 203 Fralin Life Science Institute,
Virginia Tech, Blacksburg, VA, US

ABSTRACT

Nonrandom distribution of rearrangements is a common feature of eukaryotic chromosomes that is not well understood in terms of genome organization and evolution. In malaria mosquitoes, chromosomal inversions—genome rearrangements that flip chromosomal segments by 180°—are often highly nonuniformly distributed among five chromosomal arms. These rearrangements are associated with epidemiologically important adaptations and, possibly, speciation in mosquitoes. A fundamental question is whether the genomic content of the chromosomal arms is associated with inversion polymorphism and fixation rates. This chapter highlights important differences in evolutionary dynamics of the sex chromosome and autosomes and reviews data about association between characteristics of the genome landscape and rates of chromosomal evolution. Recent studies suggest that a unique combination of various classes of genes and repetitive DNA in each arm, rather than a single type of repetitive element, is likely responsible for arm-specific rates of rearrangements. Additional factors, such as spatial constraints imposed by the nuclear architecture, may be responsible for the nonuniform distribution of rearrangements. Another important question is whether polymorphic inversions on homologous chromosomal arms of distantly related mosquito species nonrandomly share similar sets of genes. The available data indicate that natural selection favors specific gene combinations within polymorphic inversions when distant species are exposed to similar environmental pressures. This knowledge could be useful for the discovery of genes responsible for an association of inversion polymorphisms with phenotypic variations in multiple species. In this chapter, I also review the literature about a possible role of heterochromatin in speciation of malaria mosquitoes. The existing data demonstrate the elevated evolutionary plasticity of the heterochromatic portion of the mosquito genome.

* Corresponding Author's Email: igor@vt.edu (Department of Entomology, 203 Fralin Life Science Institute, West Campus Drive, MC 0346, Virginia Tech, Blacksburg, VA 24061; Office phone: (540) 231-7316; Lab phone: (540) 231-0731; Fax: (540) 231-7126).

Finally, I discuss the importance of high-quality genome assemblies for reconstructing a gene order-based phylogeny and studying mosquito evolution.

INTRODUCTION

Malaria mosquitoes belong to genus *Anopheles*, which is the largest of the three genera, *Anopheles*, *Bironella*, and *Chagasia*. These genera make up subfamily Anophelinae, which belongs to Culicidae, Diptera. Genus *Anopheles* is further subdivided into subgenera, *Anopheles*, *Cellia*, *Kerteszia*, *Lophopodomyia*, *Nyssorhynchus*, and *Stethomyia*. More than 500 recognized species of genus *Anopheles* inhabit every continent except Antarctica. Taxonomic and population complexity is a common feature of malaria mosquitoes (Krzywinski, Besansky 2003). The rich biodiversity of malaria mosquitoes has the direct epidemiological implications. Of the ~500 anopheline species, no more than 30 significantly contribute to the malaria transmission. It is still not completely clear why some species or populations are efficient malaria vectors while others are of no medical importance. Therefore, understanding the adaptation and speciation in malaria mosquitoes has not only a theoretical interest for evolutionary biology but also practical applications for vector control. Comparative genomic analyses of vector competence and other epidemiologically important traits will be informative if performed within a phylogenetic framework. Inferring ancestral and derived genomic features in anophelines is crucial for identifying the evolutionary changes associated with the origin and loss of human blood choice, ecological and behavioral adaptations, and association with human habitats. Traditionally, reconstructions of the anopheline phylogeny have been done using morphological and molecular markers (Krzywinski, Wilkerson, Besansky 2001; Krzywinski, Besansky 2003; Harbach 2004). Available data support the monophyly of the six *Anopheles* subgenera, a sister-group relationship between subgenera *Nyssorhynchus* and *Kerteszia*, and a sister-group relationship between subgenera *Cellia* and *Anopheles*. If *Plasmodium falciparum* was transferred to the Americas by European colonialists in post-Colombian times, then the contact between *Nyssorhynchus* and *P. falciparum* (which is normally transmitted by *Cellia*) is very recent (Moreno et al. 2010). This supports the hypothesis that the susceptibility of mosquitoes to the malaria parasite is an ancestral trait, while the refractoriness to this parasite in nonmalaria vector species is a derived trait. Gene-based phylogenies have limited resolutions because DNA sequences are prone to homoplasy, introgression, and ancestral polymorphism. Therefore, they must be corroborated by an independent approach based on phylogenomics analysis of entire genomes. The rearrangement-based phylogeny reconstruction has been proven to be an effective and informative way to understand evolutionary relationships among various taxa (Bourque, Pevzner, Tesler 2004; Bhutkar, Gelbart, Smith 2007; Alekseyev, Pevzner 2009). Given the abundance and uniqueness of fixed inversions in the evolution of *Anopheles* (Sharakhova et al. 2010a; Xia et al. 2010), this approach should produce a reliable phylogenetic framework for comparative genomics studies of vectorial capacity (Xia, Sharakhova, Sharakhov 2008). The focus of this chapter is to review a role of chromosomal changes in adaptation and speciation of mosquitoes from the subgenus *Cellia*, which is the largest subgenus within the genus *Anopheles* and is restricted to the Old World. This subgenus includes some of the most important malaria vectors, such as *An. gambiae*, *An. nili*, *An. funestus*, and *An. stephensi*. The *An. gambiae* and *An. funestus* lineages diverged from a common ancestor at least 36 MY ago (Krzywinski, Grushko, Besansky 2006).

1. A ROLE OF CHROMOSOME POLYMORPHISM IN ECOLOGICAL ADAPTATIONS OF MALARIA MOSQUITOES

Data obtained on various organisms suggest that chromosomal polymorphism is a mechanism species use to rapidly adapt to climate changes. Environment-dependent changes in the frequencies of chromosomal inversions can be seen within a human lifetime (Krimbas, Powell 1992; Gordeev, Ejov 2004; Hoffmann, Daborn 2007). Development of “slash-and-burn” agricultural techniques allowed humans to leave the wet tropical forest and spread into dry areas in the tropics and sub-tropics (Coluzzi et al. 2002; Ayala, Coluzzi 2005). Likewise, adaptation to aridity was a key ecological adaptation that allowed malaria mosquitoes to occupy large regions and survive during dry seasons in Africa and Asia. What is the evidence for a role of chromosomal inversions in ecological adaptation of species? Several observations suggest that inversion polymorphism is likely the major mechanism by which malaria mosquitoes adapt to aridity. The polytene chromosome complement of *Anopheles* females consists of the X chromosome and four autosomal arms: 2R, 2L, 3R, and 3L. Of the seven members of the *An. gambiae* complex, *An. arabiensis* and *An. gambiae* s.s. are the only species that have a highly polymorphic chromosome 2 and a continent-wide distribution in arid sub-Saharan Africa. Mosquito species with little or no chromosomal polymorphisms tend to occupy smaller and wetter geographic regions (Coluzzi et al. 2002). However, the casual relationships between the level of polymorphism and the geographic distribution have not been tested experimentally. The 2Rb, 2Rbc, 2Rcu, 2Ru, 2Rd, and 2La inversions of *An. gambiae* are frequent in arid Sahel Savanna mosquitoes and almost absent in those in humid equatorial Africa, strongly suggesting that these inversions confer adaptive fitness to the drier environment (Coluzzi et al. 1979; Toure et al. 1998; Powell et al. 1999; Coluzzi et al. 2002). The nonrandom pattern of distribution of adaptive inversions suggests that these rearrangements are the product of selection. Natural selection has been implemented in fixation of the 2Rj inversion during ecotypic speciation in *An. gambiae* (Manoukis et al. 2008). Finally, the variations in rates of water loss and in thermotolerance are associated with alternative arrangements of the 2La inversion in *An. gambiae* (Gray et al. 2009; Rocca et al. 2009).

Anopheles gambiae is highly polymorphic for the two 2La arrangements, which are non-randomly distributed temporally and spatially with respect to degree of humidity in East and West Africa (Petrarca et al. 2000). This pattern is most apparent in West Africa (Coluzzi et al. 1979), where strong north-south clines in the frequency of the 2La inversion range from fixation in the arid northern Sahel to absence in the humid southern rainforests. At sites along the cline where neither arrangement is fixed, seasonal fluctuations occur in which 2La cycles from low to high frequency between wet and dry seasons. The adaptive flexibility provided by this chromosomal polymorphism has probably allowed *An. gambiae* to exploit a very broad range of climatic conditions, an important factor underlying the wide distribution and abundance of this species across Africa as well as its status as primary malaria vector. In addition, the inverted arrangements have been preferentially associated with indoor biting and resting behaviors (Coluzzi et al. 1979; Powell et al. 1999). Thus, chromosomal inversions could influence epidemiologically important traits of *An. gambiae*, such as its geographic distribution, its probability of vector-human contact, and the likelihood of vector exposure to insecticide-treated walls and bed nets.

Ecological adaptations have played a central role in the radiation of *An. gambiae*. It is believed that dry seasons and zones without artificial irrigation were the ecological limits to the spatial and temporal distribution of *An. gambiae* in the past. However, today, man-made modifications of the environment, such as large irrigation schemes, are perfectly exploited all year-round by a genetically distinct taxonomic unit of *An. gambiae sensu stricto* named the M form (Pombi 2005; Gimonneau et al. 2012). This adaptation represents a clear shift from breeding in natural, sunlit, temporary rain-dependent pools characteristic of the other taxonomic unit, the S form. These two forms were first described on the basis of the virtual absence of hybrid genotypes between form-specific haplotypes found in ribosomal DNA-linked markers (Gentile et al. 2001). The process of ecological divergence is probably driven by expansion of the original larval niche of the S form, which is considered to be ancestral (Lehmann, Diabate 2008; Costantini et al. 2009). However, sunlit, temporary, freshwater pools, such as road traces, are now created by human activity as well. The ecological factors and genetic mechanisms used by the M and S forms to partition their larval environment are not known. Differential oviposition behavior, possibly affected by the chemical composition of water, the ability of larvae to escape predators, and the availability of nutrients could be among the ecological factors (Edillo et al. 2006; Diabate et al. 2008; Gimonneau et al. 2010). A likely genetic mechanism of this ecological divergence is the capture and protection of adaptive allele combinations by chromosomal inversions, as suggested by studies on adult populations of *An. gambiae* (Manoukis et al. 2008; White et al. 2009).

2. ECOLOGICAL HETEROGENEITY OF MALARIA MOSQUITOES IS A MAJOR CHALLENGE TO VECTOR CONTROL

Malaria has a devastating impact on public health and welfare on the African continent and vector control is seen as a cornerstone of the malaria control strategy worldwide. Because of economic and practical reasons, vector control in Africa mainly relies on the use of synthetic insecticides (Takken, Knols 2009). However, this strategy is inefficient if all vector species and populations are not targeted and is further jeopardized by the rapid spread of insecticide multi-resistance in major mosquito vector species. The most efficient African malaria vector *An. gambiae* consists of populations that adapt to different environments, exhibit alternative behaviors, and vary in their vectorial capacity (Lehmann, Diabate 2008; Riehle et al. 2011). Some malaria control initiatives have been unsuccessful because they targeted the wrong species or population (Coluzzi 1992; Van Bortel et al. 2001). For example, the famous \$6-million Garki malaria control program in Nigeria failed because the massive indoor residual spraying of insecticides against *An. gambiae* did not impact the chromosomally distinct outdoor population (Coluzzi et al. 1977).

Malaria in tropical, humid savannah and forest environments is quite stable with entomological inoculation rates (EIR: number of infective bites per person per year) varying between 50 and 300 (Fontenille, Simard 2004). Five major vector species are responsible for malaria transmission in these areas: *An. gambiae*, *An. arabiensis*, *An. funestus*, *An. moucheti*, and *An. nili* (Fontenille, Simard 2004). These species are responsible for >95% of the total malaria transmission on the African continent (Mouchet et al. 2004). *Anopheles nili* has as wide a geographic distribution as *An. gambiae*, *An. arabiensis*, and *An. funestus*, spreading across

most of West, Central, and East Africa, mainly in humid savannah and degraded rainforest areas (Antonio-Nkondjio et al. 2009; Ayala et al. 2009). However, unlike other major vectors, *An. nili* breeds in slow-moving streams and large lotic rivers exposed to light and containing vegetation or debris (Fontenille, Simard 2004; Antonio-Nkondjio et al. 2009). A recent study of the ecological niche profile of major malaria vectors in Cameroon demonstrated that the habitats of *An. gambiae*, *An. arabiensis*, and *An. funestus* have more overlap with each other than with the habitats of *An. moucheti* and *An. nili* (Ayala et al. 2009). This results in a much more unusual geographic distribution of *An. moucheti* and *An. nili*, revealing their crucial role in malaria transmission in degraded and equatorial forests in Cameroon (Antonio-Nkondjio et al. 2009).

Key features that render *Anopheles* species very efficient malaria vectors are host-seeking and resting behavior, ecological adaptations, reproductive biology, longevity, and vector competence. Unique aspects of ecological adaptation and behavior of each malaria vector can, in part, explain the sustained malaria transmission in Africa (Table 1).

At least some of the epidemiologically important features are associated with inversion polymorphisms. For example, the 2La inversion in *An. gambiae* has been associated with a tolerance to aridity and reduced susceptibility to *Plasmodium falciparum* (Coluzzi et al. 1979; Petrarca, Beier 1992; Gray et al. 2009). The two major malaria vectors *An. arabiensis* and *An. gambiae* had been sympatric species in most of their distribution range, allowing for introgressive hybridization between them. Available data support the hypothesis of introgression of the 2La arrangement from *An. arabiensis* into *An. gambiae* (Besansky et al. 2003; White et al. 2009; Neafsey et al. 2010). The frequency of chromosomal inversion 2La is correlated with microclimatic differences in humidity that impact mosquito behavior: this arrangement is more common in mosquitoes found resting indoors where a nocturnal saturation deficit exists (Powell et al. 1999). Such population heterogeneity has important epidemiological and ecological consequences. Indoor residual spraying of insecticides against *An. gambiae* did not impact the population uniformly. It has major implications now because WHO is recommending the use of indoor residual spraying throughout Africa. Therefore, understanding and targeting the heterogeneity and complexity of *An. gambiae* is necessary for effective vector control and malaria elimination.

Table 1. Epidemiologically important adaptations and behaviors of four major malaria vectors in sub-Saharan Africa

	<i>An. gambiae</i>	<i>An. arabiensis</i>	<i>An. funestus</i>	<i>An. nili</i>
Preferential resting (Fontenille, Simard 2004)	Indoor	Indoor/ Outdoor	Indoor	Indoor/Outdoor
Preferential biting (Dia et al. 2003; Fontenille, Simard 2004; Antonio-Nkondjio et al. 2006)	Indoor	Indoor/ Outdoor	Indoor	Indoor/Outdoor
Biting time peak (Hanney 1960; Lemasson et al. 1997; Taye et al. 2006; Keraf-Hinzoumbe et al. 2009)	12 am – 5 am	11 pm – 5 am	11 pm – 6 am	10 pm – 1 am
Host preference (Lemasson et al. 1997; Dia et al. 2003; Antonio-Nkondjio et al. 2006)	Human	Human/ Animal	Human	Human/Animal
Peak of seasonal abundance (Lemasson et al. 1997; Antonio-Nkondjio et al. 2002; Dia et al. 2003; Keraf-Hinzoumbe et al. 2009)	July–October	July–October	October–November	October–December
Larval habitats (Fontenille, Simard 2004; Antonio-Nkondjio et al. 2009)	Rain- or irrigation-dependent pools	Grassy swamps		Lotic rivers, exposed to light

A chromosomal study of indoor-collected adult females in Burkina Faso found non-random spatial distribution of inversion variants with reference to environmental variables and habitat quality. The same karyotypes have been shown to be spatially distributed in a parallel pattern in both M and S molecular forms in response to gross environmental gradients of climate, vegetation, and land cover (Costantini et al. 2009). A likely reason for the lack of association between inversion polymorphisms and types of larval habitats in this study is that indoor-resting adult females represent only a fraction of the natural population of *An. gambiae*. Indeed, a recent study of larval genotypes in the Goundry village of Burkina Faso discovered an unexpected genetic diversity within *An. gambiae* (Riehle et al. 2011). The study suggests the presence of a previously unidentified genetically distinct subgroup of *An. gambiae*. The new form is presumably exophilic and abundant, lacks differentiation into M and S molecular form genetic markers, and is highly susceptible to infection with wild *Plasmodium falciparum*. Interestingly, 2La inversion alleles are significantly different in the larval pool (where the 2L⁺ arrangement occurs at a frequency of 68%) and the indoor-resting mosquito collections, which are nearly fixed for the inverted 2La allele (98%). Thus, indoor captures likely provide a biased description and estimate of *An. gambiae* genetic diversity. However, collecting adult mosquitoes outdoors is notoriously difficult because of the lack of effective sampling methods and ethical problems with using humans as bait. Aquatic larvae are relatively easier to collect, and they are more likely to represent the full genetic diversity of the species. Detailed chromosomal analyses of larval populations may reveal new rare inversions or combinations thereof that are not represented in indoor adult populations. In addition, these unprecedented investigations may reveal new functions of the common aridity-linked ‘adult’ chromosomal arrangements 2Rb, 2Rc, 2Ru, 2Rd, and 2La at the larval stage, at which a mosquito grows and spends a significant portion of its active lifetime.

If niche partitioning is associated with specific polymorphic inversions, as is postulated with indirect evidence from adult sampling (Toure et al. 1994), then the higher diversity of available types of water reservoirs will contribute to the higher genetic diversity of the adult population. Detailed knowledge of the extent of chromosomal complexity and heterogeneity of *An. gambiae* populations will, thus, be useful for vector control. A possible introgression of inversion arrangements could impact the dynamics of mosquito adaptation and malaria transmission. Therefore, understanding the association between inversion polymorphism and ecological adaptations may lead to identification of phenotype-causing alleles that could be targeted in the future.

3. CHROMOSOME PLASTICITY IN ADAPTATION AND EVOLUTION OF MALARIA MOSQUITOES

A growing number of studies suggest that chromosome rearrangements play an important role in species evolution and adaptation (Noor et al. 2001; Rieseberg 2001; Feder et al. 2003; Schaeffer et al. 2003; Anderson et al. 2005; Ayala, Coluzzi 2005; Kirkpatrick, Barton 2006; Baack, Rieseberg 2007; Hoffmann, Daborn 2007; Noor et al. 2007; Manoukis et al. 2008). An intriguing fact is that the rates of rearrangements are taxon and chromosome sensitive (Gonzalez, Ranz, Ruiz 2002; Sharakhov et al. 2002; Coghlan et al. 2005; Ranz et al. 2007). Comparison of vertebrate genomes demonstrated a slow rate of chromosomal evolution in fish

and chicken and an accelerated rate of genome rearrangements in mammals (Postlethwait et al. 2000; 2004). Within mammals, rodent lineages have undergone 3.2–3.5 chromosome rearrangements per million years (MY) while primates have accumulated only 1.6 rearrangements per MY since the two lineages diverged. Within carnivores, the rate of chromosome evolution in *Canidae* is much higher than in other lineages (Yang et al. 1999). Comparison of genomic sequences of 12 species of *Drosophila* revealed that inversions have been fixed at different rates in different lineages (Clark et al. 2007). For example, 29 fixed inversions are located between *D. melanogaster* and *D. yakuba*. All but one of these inversions occurred in the *D. yakuba* lineage (Ranz et al. 2007). Likewise, the distribution of polymorphic rearrangements varies dramatically among lineages. More than 500 polymorphic inversions are known for *D. melanogaster* while only 14 inversions have been described for its close relative, *D. simulans* (Aulard et al. 2004a). Cytogenetic studies performed on malaria mosquitoes provided some of the most obvious examples of the nonuniform inversion distribution both among species and chromosomal arms (Coluzzi et al. 2002; Sharakhov et al. 2002; Pombi et al. 2008; Xia et al. 2010). The availability of the *An. gambiae* genome sequence (Holt et al. 2002) and the physical maps for *An. funestus* (Sharakhov et al. 2002; Sharakhov et al. 2004) and *An. stephensi* (Xia et al. 2010) enabled a fresh perspective to be gained on the relationships between the genomic landscape and evolutionary rates. Taxonomically, these species belong to the different series of the subgenus *Cellia* diverged: *Pyretophorus* (*An. gambiae*), *Myzomyia* (*An. funestus*), and *Neocellia* (*An. stephensi*) (Green, Hunt 1980). Comparative mapping between *An. gambiae* and *An. funestus* established arm homologies among these species, found no evidence for inter-arm transposition events, pericentric inversions, or partial-arm translocations, and confirmed that whole-arm translocations and paracentric inversions are the common rearrangements among species in subgenus *Cellia* (Green, Hunt 1980; Sharakhov et al. 2002) (Xia et al. 2010). Given that *An. gambiae* and *An. funestus* diverged from each other at least 36 million years ago (Krzywinski, Grushko, Besansky 2006), the rate of genome rearrangement in the subgenus *Cellia* is 0.003-0.005 inversions per 1 megabase (Mb) per million years per lineage.

3.1. Chromosome Arms Evolve at Different Rates

The availability of polytene chromosomes in anopheline mosquitoes allows for the development of comparative maps that can be used to determine sizes of syntenic regions. A recent study mapped 231 uniquely located markers to the *An. stephensi* chromosomes for a comparative study with *An. gambiae* and *An. funestus* (Xia et al. 2010). The number of inversions between *An. gambiae* and *An. stephensi* using the Nadeau and Taylor method (Nadeau, Taylor 1984) and the Genome Rearrangements In Man and Mouse (GRIMM) program (Tesler 2002) has been calculated. Both Nadeau-Taylor and GRIMM analyses revealed of inversion that the X chromosome had the highest rate fixation and that the 2R arm evolved faster than other autosomes (Xia et al. 2010) (Figure 2). Another study demonstrated a striking contrast among chromosome arms in the length of conserved segments: small conserved blocks (< 1 Mb) are located on arm 2R, and large conserved blocks (up to 6-8 Mb) are located on arms 3R and 3L (Sharakhova et al. 2011).

Of 10 inversions fixed among species of the *An. gambiae* complex, 5 have been found on the X chromosome and 3 on the 2R arm (Coluzzi et al. 2002). In contrast to the polymorphic

inversions, the majority of species-specific inversions (5 of 10) were found on the X (sex) chromosome in the *An. gambiae* complex (Coluzzi et al. 2002). The contrasting pattern of inversion polymorphism and inversion fixation on the X chromosome suggests that different forces govern sex chromosome and autosome evolution. The excess of fixed inversions and the deficit of polymorphic inversions on the X chromosome has been explained by the underdominance (homozygote advantage) model of speciation (Charlesworth, Coyne, Barton 1987). It has been proposed that genes responsible for species reproductive isolation could be located on the X chromosome (Ayala, Coluzzi 2005). Indeed, the X chromosome has a disproportionately large effect on male and female hybrid sterility in *An. gambiae* and *An. arabiensis* (Slotman, Della Torre, Powell 2004; Slotman, Della Torre, Powell 2005). The smaller effective population size of the X chromosome, as compared to autosomes, suggests that drift might have played a bigger role in the rapid fixation of the inversion on the X chromosome. The rapid evolution of sterility and inviability genes captured by polymorphic inversions on the X chromosome may cause a selection against inversion heterozygotes. From a vector control point of view, if heterozygote inversions on the X chromosome have a deleterious effect on viability and reproduction of mosquitoes, then they could be introduced artificially into the vector population to reduce its size. A study of gene ontology (GO) term distribution suggests that the X chromosome is enriched in genes that may be involved in premating isolation, such as genes encoding for proteins with molecular and signal transduction activity. Signal transduction is a crucial component of olfaction that plays a major role in mate recognition. For example, X-linked genes encoding for signal transduction proteins were differentially expressed between virgin females of two incipient species of *An. gambiae* that differ in swarming behavior (Cassone et al. 2008). Rapid generation and fixation of inversions on the X chromosome may facilitate speciation in *Anopheles* by differentiating alleles inside of the inverted regions as has been shown in *Drosophila* (Machado, Haselkorn, Noor 2007). Unlike the X chromosome in insects, the eutherian X chromosome had its gene order conserved during 105 million years of evolution, probably reflecting strong selective constraints posed by the X inactivation system in mammals (Rodriguez Delgado et al. 2009).

A study of the opossum genome revealed that the evolution of the X chromosome inactivation was associated with suppression of large-scale rearrangements in eutherians (Mikkelsen et al. 2007). Conversely, rapidly evolving sex chromosomes in insects have a dosage compensation system. Because the X chromosome in *Drosophila* males recruits fewer histones and possesses an “open” chromatin (Corona et al. 2007), it may be more sensitive to breakage (Fisher et al. 2005) and, thus, more prone to rearrangements.

In contrast to the X chromosome, the 2R and 2L arms of *An. gambiae* and their homologous arms in *An. stephensi* and *An. funestus* harbor polymorphic inversions associated with ecological adaptations (Mahmood, Sakai 1984; Costantini et al. 1999; Coluzzi et al. 2002). Natural selection has been implicated in fixation of the 2Rj inversion during ecotypic speciation in *An. gambiae* (Manoukis et al. 2008). Adaptive alleles or allelic combinations can be maintained within a polymorphic inversion by suppressing recombination between the loci (Kirkpatrick, Barton 2006; Hoffmann, Rieseberg 2008). It has been demonstrated that adaptive inversions are less frequent at shorter lengths (Caceres, Barbadilla, Ruiz 1997; Pombi et al. 2008), reflecting a smaller selective advantage when an inversion captures fewer genes (Krimbas, Powell 1992). Therefore, it has been predicted that chromosomal arms rich in polymorphic inversions (2R, 2L) would have higher gene densities. This prediction was met; moreover, the polymorphic inversion-poor X chromosome had the lowest gene density (Xia et

al. 2010). Similarly, the polymorphic inversion-rich chromosomal elements C and E have higher gene densities than the rest of the genome in *Drosophila* (Gonzalez, Ranz, Ruiz 2002). These observations highlight the fundamental differences between the evolutionary dynamics of the sex chromosome and autosomes. The high rate of sex chromosome evolution is being achieved by the rapid generation and fixation of inversions without maintenance of a stable inversion polymorphism. In contrast, the high rate of the autosomal evolution results from the high level of inversion polymorphism maintained by selection acting on gene-rich chromosomal arms. The increase of gene density in rearrangement-rich regions of autosomes was also found in vertebrates (Murphy et al. 2005; Gordon et al. 2007; Larkin et al. 2009) suggesting the general applicability of the principle “from polymorphism to fixation” to autosomal evolution.

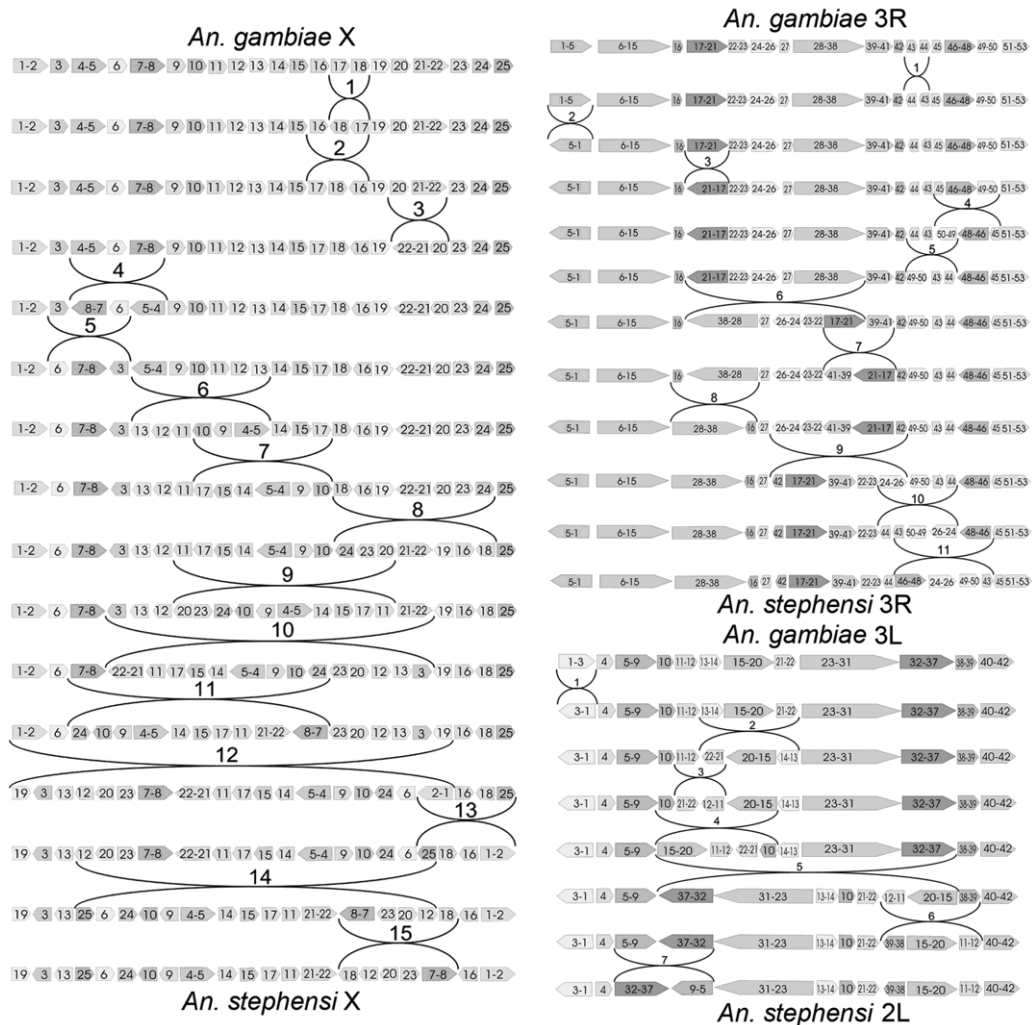


Figure 2. The gene order transformation between *An. gambiae* and *An. stephensi*. Relative position and orientation of the conserved syntenic blocks are shown by blocks. Numbers within the blocks indicate markers physically mapped to polytene chromosomes. Numbers over brackets show inversion steps. The telomere ends are on the left (Xia et al. 2010).

The polymorphic inversions 2Rb, 2Rbc, 2Rcu, 2Ru, 2Rd, and 2La of *An. gambiae* are associated with adaptation of mosquitoes to the dry environment (Coluzzi et al. 2002). Cuticle seems to play a major role in desiccation resistance of embryo and adult mosquitoes (Goltsev et al. 2009; Gray et al. 2009). These observations suggest an exciting possibility that genes involved in the cuticle development may be disproportionally clustered on the 2R and 2L arms. A study of GO terms provides evidence that 2L is indeed enriched with genes involved in the structural integrity of a cuticle while the 2R arm has overrepresentation of genes involved in cellular response to stress (e.g., temperature, humidity) and in building membrane parts (Xia et al. 2010). These data support the role of natural selection in maintaining polymorphic inversions associated with ecological adaptations.

3.2. Genomic Landscapes and Nonuniform Rates of Inversions Fixation

Evolution of gene order likely has species- and chromosome-specific facilitators or inhibitors. The major consequences of unequal rates of karyotype evolution are differential plasticity of species and an increased role of certain chromosomes in adaptation and evolution (Coluzzi et al. 2002; Hoffmann, Sgro, Weeks 2004). What are the factors that constrain or promote chromosomal rearrangements? The common assumption about the major role of transposable elements (TEs) in generating inversions (Caceres et al. 1999; Mathiopoulos et al. 1999; Evgen'ev et al. 2000) was not supported by other studies (Matzkin et al. 2005; Richards et al. 2005; Ranz et al. 2007). Moreover, the TE density in the *An. gambiae* genome was found to be lowest on the 2R arm (Holt et al. 2002); thus, it is not clear whether the molecular content could be associated with inversion polymorphism and fixation rates. A recent study has shown that fragility of certain regions rather than functional constraints plays the main role in nonuniform distribution of inversions in *Drosophila* chromosomes (von Grotthuss, Ashburner, Ranz 2010). However, the molecular determinants of the fragile breakage have not yet been determined. If nonrandom origin of inversions can be attributed to unequal density of repetitive DNA among chromosome arms, higher densities of break-causing elements on faster evolving arms can be predicted.

A study of the distribution of 82 rare polymorphic inversions in *An. gambiae* s.s. has found no inversions on the X chromosome, 67 inversions on the 2R arm, and only 15 inversions on the 2L, 3R, and 3L arms together (Pombi et al. 2008) (Figure 3). The clustering of neutral chromosomal polymorphisms and cytological co-localization of multiple breakpoints on 2R indicates that this arm is especially prone to breakages (Coluzzi et al. 2002; Pombi et al. 2008). The high rate of rearrangements could be explained by 2R-biased distribution of repetitive DNA capable of generating inversions. However, the transposable element density in the *An. gambiae* genome was found to be lowest on the 2R arm (Holt et al. 2002) (Xia et al. 2010) and, thus, it is not clear whether the molecular content could be associated with inversion fixation rates. Although, the role of transposable elements in the origin of individual inversions was demonstrated earlier (Mathiopoulos et al. 1998; Caceres et al. 1999), their presence at the breakpoints has not been confirmed by genome-wide analyses in *Drosophila* (Richards et al. 2005; Ranz et al. 2007; Bhutkar et al. 2008). Segmental duplications have been found in the opposite orientation at the breakpoints of the 2Rj inversion of *An. gambiae* and have been implicated in generating this rearrangement (Coulibaly et al. 2007). Simple repeats have been shown to play a role in the formation of hairpin and cruciform structures, which can cause

double-strand DNA breaks and rearrangements (Lobachev, Rattray, Narayanan 2007). In *Drosophila*, the microsatellite density of the chromosomes parallels their evolution rates (Gonzalez, Ranz, Ruiz 2002).

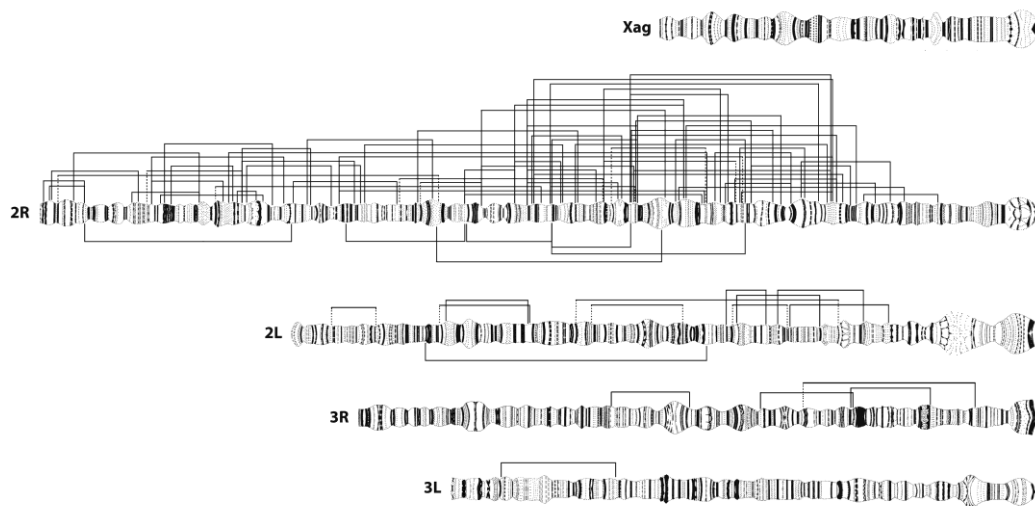


Figure 3. Distribution of polymorphic inversions among chromosomal arms X, 2R, 2L, 3R, and 3L in *An. gambiae*. Brackets above chromosomes indicate rare polymorphic inversions, and brackets below chromosomes show common polymorphic inversions (Pombi et al. 2008).

A recent study applied a Bayesian statistical model and procedure for discerning differences between arms in molecular features, such as DNA-mediated TEs (DNA TEs), RNA-mediated TEs (RNA TEs), segmental duplications (SDs), micro- and minisatellites, satellites, MARs, and genes (Xia et al. 2010) (Figure 4). The X chromosome had the highest density of TEs and the highest coverage of microsatellites, minisatellites, and satellites. The 2R arm had the highest density of genes and regions involved in SDs but had the lowest densities of TEs and the lowest coverage of minisatellites and MARs. This study detected the highest density of regions with SDs in the proximal half of the 2R arm where the breakpoint-rich area is located (Pombi et al. 2008) (Figure 3). The correlation coefficient between the densities of breakpoints and regions involved in SDs in 5-Mb intervals within 50 Mb of the euchromatic part of 2R was 0.9, suggesting an arm-specific involvement of SDs in inversion formation rather than a genome-wide impact.

Simple repeats have been shown to play a role in the formation of hairpin and cruciform structures, which can cause double-strand DNA breaks and rearrangements (Lobachev, Rattray, Narayanan 2007). In *Drosophila*, the fastest evolving X chromosome has the highest densities of microsatellites and TEs (Gonzalez, Ranz, Ruiz 2002; Fontanillas, Hartl, Reuter 2007). Although, the role of TEs in the origin of individual inversions was demonstrated earlier (Lyttle, Haymer 1992; Mathiopoulos et al. 1998; Caceres et al. 1999; Mathiopoulos et al. 1999; Aulard et al. 2004b), the more recent sequencing of breakpoints discovered alternative mechanisms of inversion generation (Richards et al. 2005; Sharakhov et al. 2006; Ranz et al. 2007; Bhutkar et al. 2008). SDs have been implicated in inversion generation in mosquitoes and mammals (Goidts et al. 2004; Coulibaly et al. 2007) and are considered as a marker of genome fragility (Bailey et al. 2004).

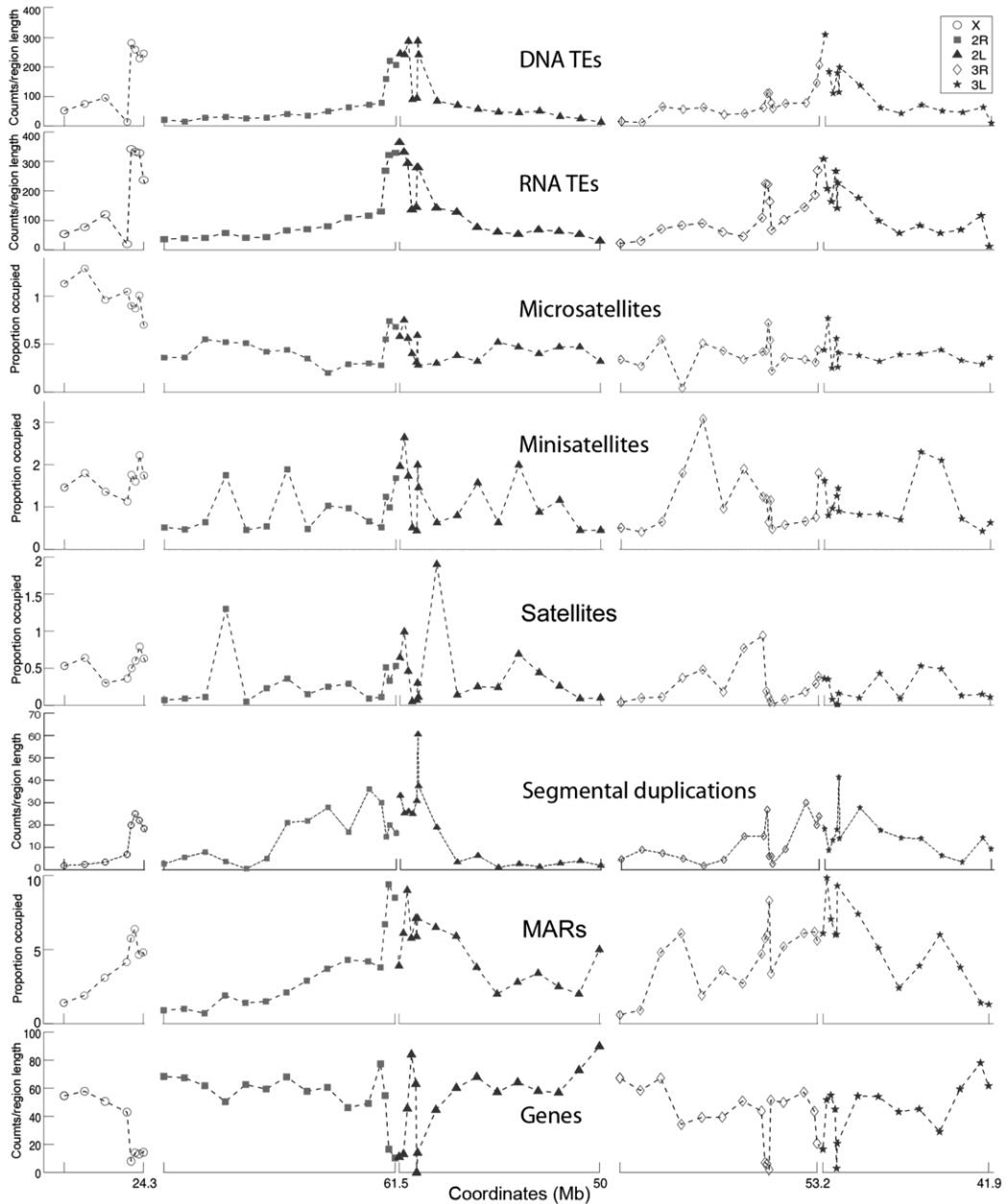


Figure 4. Genome landscapes of the *An. gambiae* chromosomal arms. Median counts per 1 Mb are given for regions involved in SDs (Xia et al. 2010). The coordinates and orientation of each arm are the following: X: 0 Mb—telomere, 24.3 Mb—centromere; 2R: 0 Mb—telomere, 61.5 Mb—centromere; 2L: 0—centromere, 50 Mb—telomere; 3R: 0 Mb—telomere, 53.2 Mb—centromere; 3L: 0 Mb—centromere, 41.9 Mb—telomere.

The *An. gambiae* X chromosome and 2R arm have the highest G+C content. GC-rich regions have been implemented in forming hotspots for chromosome rearrangements (Fisher et al. 2005; Gordon et al. 2007) because of their propensity to form Z-DNA, hairpin loops, and other unstable structures that are capable of generating double-strand breaks (Wang, Christensen, Vasquez 2006). Interestingly, a GO term analysis demonstrated that the X

chromosome is enriched with nucleobase, nucleoside, and nucleotide metabolic processes and that the 2R arm has overrepresented gene clusters involved in DNA damage repair. It is possible that these GO term enrichments have evolved in response to high rates of DNA breakage on the X and 2R chromosomes. Because of the paucity of pericentric inversions and partial-arm translocations in mosquito evolution, the genome landscapes and evolutionary histories of individual arms are different. There is a strong association between the genome landscape characteristics and the rates of chromosomal evolution. A unique combination of various classes of genes and repetitive DNA in each arm, rather than a single type of repetitive element, is likely responsible for arm-specific rates of rearrangements. These findings call for a reevaluation of the genomic analyses, which must be performed on an arm-by-arm basis using sequences physically mapped to the chromosomes.

3.3. Polymorphic Inversions and Parallel Adaptations

As was mentioned previously, polymorphic chromosomal inversions are highly non-uniformly distributed among five chromosomal arms of malaria mosquitoes. In *An. gambiae* s.l., 18 of the 31 common polymorphic inversions (58%) are on chromosome 2R, which represents less than 30% of the polytene complement. Of the eight common polymorphic inversions described in *An. gambiae* s.s., seven occur on chromosome 2R and one on chromosome 2L (Coluzzi et al. 2002). The central part of chromosomal arm 2R appears to play a major role in both chromosomal changes between species and in the polymorphism within species of the *An. gambiae* complex. Similarly, in distantly related malaria vectors, *An. gambiae* s.s., *An. funestus*, and *An. stephensi*, the majority of successful polymorphic inversions are located on chromosomal arm 2R (Mahmood, Sakai 1984; Coluzzi et al. 2002; Sharakhov et al. 2004). Is the concentration of adaptive rearrangements on just one of five chromosomal arms in the distant mosquito species coincidental or does it reflect a common genetic mechanism of adaptation to a dry climate? The location of ecologically important inversions on chromosome arm 2R in different species may reflect the importance of specific (homologous) genomic regions in the evolution and adaptation of mosquitoes. If the phenotype associated with inversion polymorphisms is the same in *An. gambiae*, *An. funestus*, and *An. stephensi*, the common genes, gene families, and clusters are likely to be responsible for the ecological adaptation of these species. Natural selection seems to play a role in maintaining inversion polymorphisms on 2R and 2L of *An. gambiae* and their homologous arms in *An. stephensi* and *An. funestus* (Mahmood, Sakai 1984; Costantini et al. 1999; Coluzzi et al. 2002).

A previous study has shown that chromosomal arms rich in polymorphic inversions (2R, 2L) have higher gene densities (Xia et al. 2010). This observation confirmed the assumption that an inversion with fewer genes would have a smaller selective advantage (Krimbas, Powell 1992). Moreover, the study of gene ontology terms has provided evidence that 2L is enriched with genes involved in the structural integrity of a cuticle, while the 2R arm has an overrepresentation of genes involved in cellular response to stress (e.g., temperature, humidity) (Xia et al. 2010). These data strongly support the role of natural selection in maintaining polymorphic inversions associated with adaptation of mosquitoes to the dry environment (Coluzzi et al. 2002). A study of larval ecology demonstrated that sympatric species *An. gambiae* and *An. funestus* inhabit a wide range of the same ecological settings in Cameroon (Ayala et al. 2009). If polymorphic inversions in the distant species confer the same adaptations

then similar sets of genes can be present within inversions of these species. The presence of similar sets of genes within independent inversions would imply the role of natural selection acting on similar genetic content of homologous chromosomal arms and creating parallel phenotypes of the evolutionary distant species. A theoretical model suggests that the probability of parallel evolution under natural selection is about two times bigger than that under neutrality (Orr 2005). Results of a recent study demonstrated that inversions on 2L of *An. gambiae*, 3L of *An. stephensi*, and 3R of *An. funestus* have almost random sets of genes (Sharakhova et al. 2011). This study found that the several 2R inversions in *An. gambiae*, *An. stephensi*, and *An. funestus* do share common genes. The 2Rf inversion of *An. stephensi* has an increased frequency in the urban environment (Mahmood, Sakai 1984) and nonrandomly share common genes with overlapping inversions 2Rc, 2Rd, 2Rbk, and 2Ru of *An. gambiae*. Another “urban” inversion in *An. stephensi*, 2Rb, had a gene homology to the inversion 2Rc of *An. gambiae*. In contrast, the 2Re inversion of *An. stephensi* has an increased frequency in the rural environment (Mahmood, Sakai 1984) and nonrandomly shares common genes with inversion 2Rb of *An. gambiae* and overlapping inversions 2Rd/2Rh of *An. funestus*. This nonrandom distribution of markers is not only the result of preservation of ancestral gene order. In fact, cases with extensively reshuffled gene orders have been found within independently originated polymorphic inversions.

The gene shuffling is common in malaria mosquitoes (Xia et al. 2010), and these species are phylogenetically distant enough from each other (Green, Hunt 1980; Krzywinski, Grushko, Besansky 2006) to have independently originated polymorphic inversions, which differ in chromosomal positions and size. The nonrandom presence of homologous genes within inversion 2Rb of *An. gambiae* and inversion 2Rh of *An. funestus* is especially interesting in the light of ecological adaptations associated with these inversions. The high frequency of the 2Rb inversion of *An. gambiae* has been found strongly associated with increased degree of aridity. In contrast, the low frequencies of this inversion have been recorded in humid areas (Coluzzi et al. 1979). The 2Rh inversion of *An. funestus* has a similar (although reverse) pattern of association with aridity. Correlation of the 2Rh inversion with the higher vapor pressure has been demonstrated the strongest among all studied inversions of *An. funestus*. In contrast, the standard (2Rh+) arrangement has been found associated with the lower vapor pressure (Ayala et al. 2011). Thus, it is likely that natural selection favors adaptive gene combinations within polymorphic inversions on 2R when distantly related species are exposed to similar environmental pressures. The availability of these gene complexes would support long-term maintenance of polymorphic inversions. This knowledge could be useful for the discovery of genes responsible for an association of inversion polymorphisms with phenotypic variations in multiple species. If candidate genes were identified within a polymorphic inversion in one species, the orthologous genes in another species likely play a similar role in adaptation if they are captured by a polymorphic inversion involved in the parallel adaptation. Future studies should identify specific alleles associated with parallel adaptation of species of subgenus *Cellia*.

3.4. Nuclear Architecture and Chromosome Plasticity

Polymorphic inversions tend to cluster on the chromosomal arms 2R and 2L but not on X, 3R and 3L in *An. gambiae* and homologous arms in other species. However, it is unclear if the evolutionary breakage of inversion-poor chromosomal arms is under constraints. A recent study

has detected the chromosomal arm specificity in rates of gene order disruption during mosquito evolution (Sharakhova et al. 2011). It concluded that the distribution of breakpoint regions is evolutionary conserved on slowly evolving arms and tends to be lineage-specific on rapidly evolving arms. It has been hypothesized that the arm-specific tolerance to chromosomal breakage could be responsible for the nonuniform distribution of inversions in autosomes. The comparative analysis of conserved and disrupted gene blocks in chromosomal arms across the three species provided evidence that 2R is more tolerant to disrupting gene orders and generating new evolutionary breakpoints than other arms. The study observed that if a block on 2R was conserved between two mosquito species it was likely disrupted in the third species. In contrast, all identified gene blocks remain preserved on the 3R arm of *An. gambiae* and the homologous arms of the other species suggesting the existence of arm-specific constraints to breakage. These constraints could be controlled by negative selection acting against disruption of certain gene combinations. It is possible that slowly and rapidly evolving chromosomes may differ in sizes or abundance of coregulated gene clusters. Accumulating evidence suggests that gene order in eukaryotic genomes randomly aggregates in clusters with similar expression levels (Michalak 2008). Purifying selection against genomic rearrangements may preserve physical colocalization of coexpression clusters (Hurst, Pal, Lercher 2004). For example, clusters of genes deregulated in *trx* mutant *D. melanogaster* larvae are not uniformly distributed along the genome; 60% of them are located on chromosome 3L (Blanco et al. 2008), which is a slowly-evolving arm in *Drosophila* (Ranz et al. 2007; Bhutkar et al. 2008). Physically clustered genes may have shared regulatory regions, common expression pattern, and chromatin-level regulation, or they may represent clusters of essential genes (Ng, Wu, Zhang 2009). Additionally, conservation of gene order in certain regions of the genome has been explained by long-range gene regulation (Mongin, Dewar, Blanchette 2009). If the 3R arm of *An. gambiae* is enriched in large functional gene clusters, then generating inversions on this arm or inserting a transgene into 3R will likely have a negative effect on mosquito fitness. Alternatively, the location of breakpoint clusters in only specific chromosomal regions of the 3R arm of *An. gambiae* and the homologous arms of the other species could be responsible for the preservation of the gene order in evolution. Indeed, a recent study has shown that fragility of certain regions rather than functional constraints plays the main role in nonuniform distribution of inversions in *Drosophila* chromosomes (von Grotthuss, Ashburner, Ranz 2010). If this is the case in mosquitoes, then the distribution of breakpoints is lineage-specific on 2R and evolutionary conserved on the 3R arm of *An. gambiae* and the homologous arms of the other species. A previous study has found that the 2R arm has the highest density of regions involved in segmental duplications that clustered in the breakpoint-rich zone of the arm. Future analyses of the genome sequence in different species will shed light on the exact mechanism of breakage in each individual chromosomal arm. Regardless of the mechanism, it is clear that new rearrangement breaks are more easily allowable on 2R than on other arms in different mosquito lineages, thus, contributing to the arm-specific differences in rates of chromosomal evolution in *Anopheles*.

An interesting observation is that specific genome rearrangements are often associated with the genesis of certain types of tumors in humans (Therman, Susman, Denniston 1989; Kozubek et al. 1999; Neves et al. 1999; Marshall 2002; Bartova et al. 2005; Gandhi et al. 2006; Folle 2008). Analyses of the architecture of these nuclei revealed a general principle: because of nonrandom nuclear organization, certain loci are nonrandomly close together in the nucleus and have increased opportunities to interact and generate rearrangements. A recent study of 3D

organization of the yeast genome revealed that interchromosomal contacts colocalize with fragile sites where chromosomal breakpoints occur in evolution (Duan et al. 2010). Additionally, other interactions may be inhibitory. Matrix-associated regions (MARs) of DNA can bind directly to lamin—a major protein of the nuclear envelope—and can potentially increase chromosome stability in the cell nucleus (Baricheva et al. 1996; Dechat et al. 2008). Can evolutionary changes in nuclear architecture explain the taxon and chromosome specificity in rates of rearrangements? As demonstrated in insects, spatial organization of polytene chromosomes has species-specific characteristics. Closely related species within the *An. maculipennis* complex and the *D. melanogaster* subgroup can be discriminated on the basis of the spatial localization and morphology of the chromosomal regions to which the nuclear envelope is attached in germ-line cells (Stegnii 1987; Stegnii, Vasserlauf 1994). F1 hybrids between members of the *An. gambiae* complex demonstrate spatial separation of homologous polytene chromosomes in nuclear envelope-attachment regions suggesting reorganization of nuclear architecture during speciation (Coluzzi, Sabatini 1969). Although positioning within the nucleus of particular chromosomes is evolutionarily preserved between humans and higher primates (Tanabe et al. 2002; Neusser et al. 2007), mouse chromosomes are positioned at different nuclear locations within mouse nuclei when compared to the nuclear position of human chromosomes (Foster, Bridger 2005; Meaburn, Parris, Bridger 2005). Moreover, parental genomes are spatially separated in F1 hybrids between different species of mice (Mayer, Fundele, Haaf 2000; Mayer et al. 2000). Thus, drastic reorganization of nuclear architecture may occur during evolution. Comprehensive research efforts should specifically target nuclear architecture dynamics in evolution and its role in genome rearrangements.

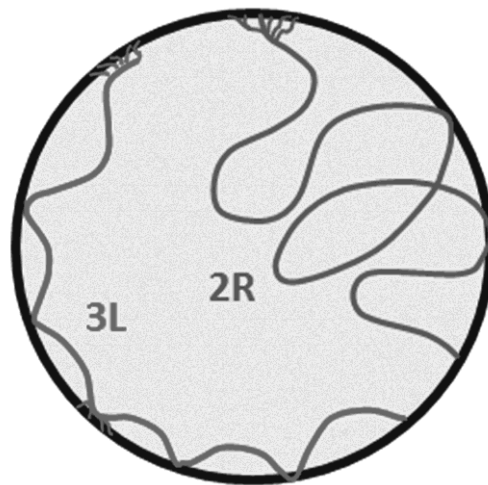


Figure 5. A model of interaction of the 2R and 3L arms with the nuclear envelope (Xia et al. 2010).

To better understand the evolutionary dynamics of chromosomal inversions, a genomic study developed and deployed novel Bayesian statistical models to analyze genome landscapes in individual chromosomal arms *An. gambiae* (Xia et al. 2010). This study has shown that, unlike all other repeats, MARs were concentrated in arms 2L, 3R, and 3L (Figure 4). The study found a negative correlation between the rates of fixed inversions and MARs coverage ($r = -0.766$), suggesting a role for nuclear architecture in controlling the rearrangements. It has been proposed that multiple attachments of 2L, 3R, and 3L to the nuclear envelope make rejoining

different breaks and forming inversions more difficult despite the abundance of TEs and simple repeats in these arms. In contrast, the lower coverage of MARs on 2R makes fewer nuclear envelope—chromosome contacts and allows more interaction between loci (Xia et al. 2010) (Figure 5).

In the nuclei of most cell types and organisms, heterochromatin contacts with the periphery and plays an important role in maintaining spatial and functional organization of chromosomes (Joffe, Leonhardt, Solovei 2010). Pericentric heterochromatin of polytene chromosomes in fruit flies and mosquitoes is represented by two morphological types—compact and diffuse (Heitz 1934; Gall, Cohen, Polan 1971; Zhimulev 1998; Sharakhova et al. 2000; Sharakhov et al. 2001). A number of studies have demonstrated direct associations between diffuse but not compact heterochromatin and the nuclear envelope in fruit flies and mosquitoes (Hochstrasser, Sedat 1987b; Hochstrasser, Sedat 1987a; Baricheva et al. 1996; Sharakhov et al. 2001; Singh, Georgatos 2002; Akhtar, Gasser 2007; Scherthan 2007). It is possible that the compact and diffuse types of heterochromatin, because of their distinct molecular compositions, have different functions in the nucleus. Specialized nuclear envelope-associated chromosomal regions have been described in *Drosophila*, human, and mouse (Pickersgill et al. 2006; Guelen et al. 2008; Peric-Hupkes et al. 2010). They represent large genomic regions (from 40 kb to 15 Mb) that tend to reside in close (<50 nm) proximity to the nuclear envelope. However, the actual DNA sequence motifs responsible for these associations have yet to be discovered.

4. HETEROCHROMATIN AND SPECIATION IN MALARIA MOSQUITOES

It is now widely recognized that the functional characterization of any genome, in order to be meaningful, must include an understanding of the epigenome. One major focus of epigenomics is to study heterochromatin and the chromosomal proteins and histone modifications involved in DNA replication, gene expression, gene silencing, and inheritance. Epigenetic modifications are being extensively characterized in genomes of model organisms (<http://www.modencode.org>), humans (www.epigenome.org), and *Plasmodium* (Salcedo-Amaya et al. 2009; Ponts et al. 2010). *Anopheles* is the only member of the “malaria triad” that lacks epigenomic studies. The mosquito heterochromatin contains protein-coding genes and accumulates essential genes important for establishing, maintaining, and modifying chromatin structure (Grushko et al. 2009; Sharakhova et al. 2010b). The chromatin state largely determines the position of a locus with respect to the transcriptionally active nuclear interior and the nuclear periphery, which is a repressive environment with respect to transcription (Pickersgill et al. 2006; Fillion et al. 2010). The packaging of DNA into euchromatin and heterochromatin of *An. gambiae* could have major implications for understanding how genome sequences function and respond to *Plasmodium* infection. A recent study has shown that gene loci that are upregulated after infection with *Schistosoma* have become spatially repositioned in the nuclei of the infected cells of *Biomphalaria glabrata* snail into a significantly more interior location (Knight et al. 2011). The mechanism through which chromatin behavior and dynamics respond to parasitic infection could be a target to control infection. Therefore, characterization of heterochromatin domains will open a new venue for studying the *Anopheles*—*Plasmodium* interaction.

To determine the extent of heterochromatin within the *An. gambiae* genome, genes were physically mapped to the euchromatin-heterochromatin transition zone of polytene chromosomes. The study revealed that a minimum of 232 genes reside in 16.6 Mb of mapped heterochromatin (Sharakhova et al. 2010b). Gene ontology analysis revealed that heterochromatin is enriched in genes with DNA-binding and regulatory activities. Sequencing of the genome of the major African malaria vector *An. gambiae* (Holt et al. 2002) provides an opportunity to analyze the molecular structure of the heterochromatin and to study genomic determinants of heterochromatin formation, maintenance, and function. In malaria mosquitoes, the heterochromatin size and morphology of sex chromosomes vary significantly among species and within species (Gatti et al. 1977; Sharakhova, Stegnii, Braginets 1997; Sharakhova, Stegnii, Timofeeva 1997), possibly affecting mating behavior and fertility (Bonaccorsi et al. 1980). A genome-wide microsatellite study of members of the *An. gambiae* complex has determined a high level of genetic introgression among species (Wang-Sattler et al. 2007). However, the *An. gambiae* microsatellites at six loci of X, 3L, and 3R could not be amplified in all sibling species, indicating significant sequence divergence from the major malaria vector. These loci were identified as heterochromatic in another study (Sharakhova et al. 2010b).

In the *An. gambiae* complex, one of the species, *An. gambiae sensu stricto*, is subdivided into two subtaxa: the M and S molecular forms (della Torre et al. 1997). These two partially isolated subtaxa predominantly breed within their own form and differ in behavior and environmental adaptations (Lehmann, Diabate 2008). Earlier cytological studies showed the presence of significant intra- and interspecific differences in amount and location of heterochromatin in the *An. gambiae* complex (Gatti et al. 1977; Bonaccorsi et al. 1980). The major regions of genomic differentiation between M and S forms of *An. gambiae* have been found in pericentric regions of all chromosomes (White et al. 2010). The study found three islands of genomic divergence: a ~4-Mb region on the X chromosome, a ~2.5-Mb region on the 2L arm, and a 1.7-Mb region on the 3L arm. However, it is not clear if the pericentric islands of genomic divergence are located within heterochromatin or mostly overlap with euchromatin of *An. gambiae*. A recent analysis showed that the positions of islands of genomic divergence mostly correspond to the positions of physically mapped regions of pericentric heterochromatin (Sharakhova et al. 2010b) (Figure 6). The sizes of the pericentric heterochromatin were the following: 4.4 Mb of the X chromosome, 2.4 Mb of the 2L arm, and 1.8 Mb of the 3L arm. Thus, the overlaps with islands of genomic divergence are 91% in the X chromosome, 97% in the 2L arm, and 94% in the 3L arm. The high levels of genomic divergence in pericentric heterochromatin could be due to low recombination or selection or a combination of both factors. Although, more recent whole-genome sequencing showed the genome-wide pattern of divergence between the two forms, the pericentric heterochromatin of X and 2L as well as intercalary heterochromatin of 3R were among the highly diverged regions between the M and S forms (Lawniczak et al. 2010). Moreover, the pericentric heterochromatin of all three chromosomes exhibited the strongest signals of selective sweeps for the M and S forms, suggesting that the extensive divergence observed in these regions has been driven by selection (Neafsey et al. 2010). Of 536 genes experienced strong and recent selective sweep 161 genes are heterochromatic (9.7 gene per Mb), 153 genes are in pericentric heterochromatin (12.6 gene/Mb) and 375 genes are euchromatic (1.5 gene/Mb). It has been argued that variation in divergence between M and S forms across the genome is maintained in the islands of genomic divergence despite a gene flow in other regions rather than they retained the ancestral

divergence with a minimal gene flow (Weetman et al. 2012). These observations indicate that heterochromatic sequences diverge rapidly during speciation of malaria mosquitoes.

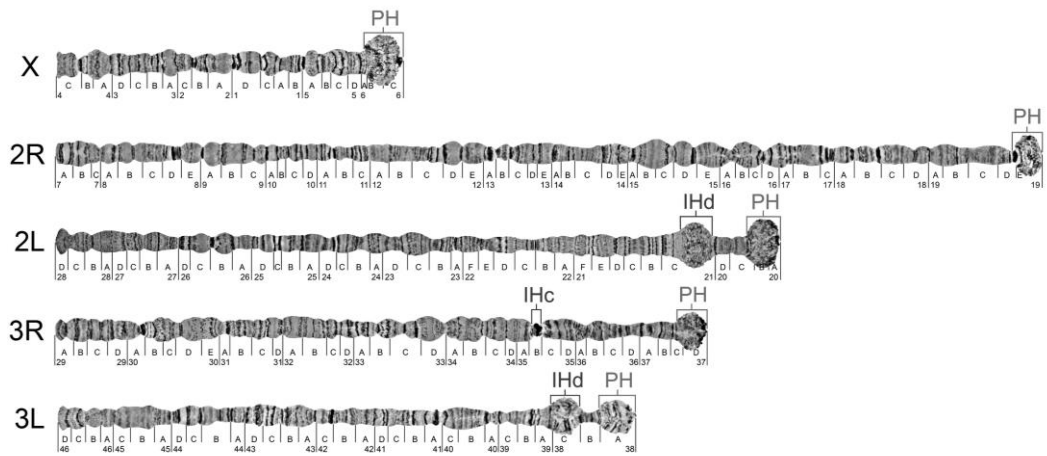


Figure 6. The pericentric and intercalary heterochromatin of polytene chromosomes shown on a standard cytogenetic map of *An. gambiae*. PH—pericentric heterochromatin, IHC—compact intercalary heterochromatin, IHD—diffuse intercalary heterochromatin (Sharakhova et al. 2010b).

Differences in the sequence and amount of heterochromatin, together with differences in the composition of heterochromatic proteins could contribute to chromosomal change leading to postzygotic isolation (Brown, O'Neill 2010). Fast changes in heterochromatic DNA can be accompanied by the rapid evolution of heterochromatic proteins. Although HP1 is an evolutionarily conserved protein, other heterochromatin- and centromere-associated proteins demonstrate rapid adaptive evolution (Talbert, Bryson, Henikoff 2004; Vermaak, Henikoff, Malik 2005). For example, an LHR protein encoded by *lhr* (*Lethal hybrid rescue*) colocalizes with HP1 in heterochromatic regions and has diverged extensively in sequence between *D. melanogaster* and *D. simulans* species in a manner consistent with positive selection. Interestingly, F1 hybrids between these species demonstrate altered chromatin structure, probably attributable to the effects of species-specific differences in TEs and other repetitive DNAs (Brideau et al. 2006), suggesting a role for heterochromatin in speciation. Because pericentric islands of genomic divergence between the M and S forms of *An. gambiae* are almost completely heterochromatic, it is possible that the mosquito heterochromatin has the elevated evolutionary plasticity and a role in speciation. One of the important tasks is to test if the X chromosome heterochromatin patterns are significantly differentiated among the *An. gambiae* complex species and the M and S forms.

Studies on *Drosophila* demonstrated the role of heterochromatin in controlling repeated sequences. For example, the COM locus, the centre organisateur de mobilisation, is a site in *Drosophila* that is responsible for the control of multiple retroTEs. It is located within the 20A region of the X heterochromatin, near the centromere (Desset et al. 2003). This locus is also known as the flamenco (flam) locus (Mevel-Ninio et al. 2007). The COM/flam locus has been found to control three retrotransposons in particular, ZAM, gypsy, and Idefix. In a normal strain of *D. melanogaster*, the COM locus is present and intact, thus leading to the silencing of these retrotransposons. In certain mutant strains – Rev and its derivatives, the COM locus is disrupted and leads to a high mobilization of these retrotransposons (Desset et al. 2008). The COM/flam

locus has been described as a master locus for the production of PIWI interacting small (24 to 30 nt long) RNAs (piRNAs) (Brennecke et al. 2007). piRNAs have been directly involved in the silencing of retrotransposons (Saito, Siomi 2010). Because inter-specific hybridization is often associated with transcriptional and epigenetics changes, and sometimes with mobilization of TEs (Michalak 2009), studying of piRNAs in hybrids may shed light on the mechanism of incompatibilities.

5. SIGNIFICANCE OF HIGH-QUALITY GENOME ASSEMBLIES FOR STUDYING MOSQUITO EVOLUTION

Draft genome assemblies of 16 *Anopheles* species (https://agcc.vectorbase.org/index.php/Main_Page) (Besansky 2008) require physical mapping to make them adequate for comprehensive genomic analyses related to studying various aspects of vectorial capacity. Assembling the genomes of other anophelines without a physical map, even with the help of the *An. gambiae* genome, will be extremely difficult given the extensive rearrangements of gene order among species (Cornel, Collins 2000; Sharakhov et al. 2002). Therefore, physical mapping of the polytene chromosomes will complement a sequencing project by facilitating the genome assembly. The quality of genome annotation of any organism depends highly on the completeness of the genome assembly. To perform full genome annotation for *An. gambiae*, one needs to identify all sequences that have a biological role in this malaria vector. Knowledge of the full complement of mosquito genes, regulatory elements, and repetitive elements is incomplete without information about what lies in those missing sequences. Creating physical genome maps is a crucial step in identifying the assembly gaps. The presence of haplotypes from cryptic species or isolated populations complicates genome assemblies. Although the *An. gambiae* PEST genome assembly has been improved by additional physical mapping of polytene chromosomes (Sharakhova et al. 2007), the abundance of misassembled M and S haplotype scaffolds still poses a serious problem for accurate annotation and functional characterization of the genome. The unmapped portion of the AgamP3 *An. gambiae* genome assembly comprises 42 Mb (Lawson et al. 2009; Megy et al. 2012) (15.6% of the genome) and includes both haplotype scaffolds and heterochromatin sequences (Sharakhova et al. 2010b). The importance of chromosome-based physical mapping for comparative genomics was emphasized in the article titled “Every genome sequence needs a good map.” The authors suggested looking “back in the future” for developing high-resolution physical maps as an important framework for the genome annotation and evolutionary analysis (Lewin et al. 2009). Physical maps and genome sequences allowed researchers to reconstruct ancestral genomes and determine the patterns and mechanisms of chromosomal evolution in mammalian and insect species (Murphy et al. 2005; Alekseyev, Pevzner 2007; Ranz et al. 2007; Bhutkar et al. 2008). Unmapped genome assemblies of anopheline mosquitoes will not be useful for reconstructing a gene order-based phylogeny and studying chromosome evolution in mosquitoes.

Anopheles gambiae belongs to a complex of seven sibling species, most of which are of less or no importance as malaria vectors. It is possible to determine an ancestral karyotype if outgroup arrangements are known (Green 1982; Lemeunier, Ashburner 1984). This argument is based on the idea that the chances of inversion re-use are very small. Moreover, phylogenies

based on inversion data are highly congruent with phylogenies based on nucleotide sequence data (O'Grady et al. 2001). A reconstruction of the *An. gambiae* complex phylogeny using fixed inversions and polytene chromosome maps of outgroup species has been attempted (Pape 1992). Although the sister group relationships of some species have been confirmed, the identification of the ancestral arrangements has failed because cytogenetic maps provide insufficient information on gene order and breakpoint locations. Therefore, higher resolution of gene order arrangements of an outgroup species and information on molecular position of inversion breakpoints are needed to reconstruct the inversion history in the *An. gambiae* complex. In a recent study, the breakpoint sequences of fixed overlapping inversions 2Ro and 2Rp in the *An. merus*—*An. gambiae* clade and homologous sequences in *An. stephensi*, *Aedes aegypti*, and *Culex quinquefasciatus* were obtained and analyzed (Kamali et al. 2012). The study demonstrated that all studied outgroup species had the gene arrangement identical to that in the 2Ro breakpoints of *An. merus* and in the 2R⁺ breakpoints of in *An. gambiae*. Because 2Ro and 2R⁺ uniquely characterize the *An. gambiae*—*An. merus* clade (Figure 1), these two species have the least chromosomal differences from the ancestral species of the complex as compared with other members. *Anopheles gambiae* and *An. merus* are vectors of human malaria and a more derived species, *An. quadriannulatus*, is not an effective vector. Thus, this study supports the repeated origin of vectorial capacity in the complex (Kamali et al. 2012).

The presence of readable polytene chromosomes in *Anopheles* species provides an uncommon opportunity for creating highly finished reference genome assemblies. The fruit fly research community made a special effort to create polytene chromosome maps for 12 species of genus *Drosophila* (Clark et al. 2007; Schaeffer, Group 2008). These maps allowed high-quality whole-genome assemblies with correct ordering of genic and nongenic DNA segments. A timely investment in producing reliable genome assemblies for major vectors will enable researchers to perform comprehensive comparative genomic analyses and will maximize investments in whole-genome resequencing of individual genomes from natural populations. The whole-genome resequencing approach has already been used to investigate patterns of the *An. gambiae* population divergence along a cline a latitudinal cline in aridity in Cameroon (Cheng et al. 2012).

CONCLUSION

Future mosquito genome sequencing projects, together with high-resolution physical mapping, will provide a better understanding of mechanisms of chromosomal rearrangements and roles of inversions and heterochromatin variations in creating biodiversity. The unequal chances of chromosome arms and lineages in generating and fixing inversions, as well as in heterochromatin modifications, may have serious consequences for differential adaptive and evolutionary plasticity of organisms. Therefore, it is important to understand the forces that govern the structural malleability of genomes. Although chromosomal inversions are abundant in malaria mosquitoes, their patterns and mechanisms remain a “black box” in vector biology. Research efforts to understand evolutionary mechanisms of chromosome evolution in vector species can be beneficial if conducted before new genomics strategies for vector control (e.g., using transgenic mosquitoes) are attempted. The following studies will significantly advance our understanding of chromosomal mechanisms of mosquito evolution.

A recent comparative study of the physical maps of *An. gambiae*, *An. funestus*, and *An. stephensi* demonstrated an excess of fixed inversions, as compared to a deficit of polymorphic inversions on the X chromosome (Xia et al. 2010). This phenomenon, if confirmed by whole-genome analyses of multiple species, could indicate that polymorphic inversions on the X chromosome are underdominant, as was theoretically predicted earlier (Charlesworth, Coyne, Barton 1987). Underdominant systems for gene drive could be used for replacement of defined target vector populations (Sinkins, Gould 2006). A rapid fixation of underdominant rearrangements could be an excellent tool for driving transgenes into natural populations of mosquitoes, if the heterozygotes are close to inviable or infertile. Molecular analysis of inversion breakpoints will inform about a mechanism for creating artificial rearrangements.

- [1] A comparative physical mapping has shown that natural selection favors adaptive gene combinations within polymorphic inversions on the 2R chromosome arm when three distantly related *Anopheles* species are exposed to similar environmental pressures (Sharakhova et al. 2011). This knowledge, obtained from multispecies genome sequence comparison, can be useful in the discovery of genes responsible for an association of inversion polymorphisms with phenotypic variations. If candidate genes are identified within a polymorphic inversion in one species, the orthologous genes in another species likely play a similar role in adaptation if they are captured by a polymorphic inversion involved in the parallel adaptation. Next generation sequencing analysis of transcriptome can be used to quantify expression differences between alternative inversion arrangements for different life stages and sexes of mosquitoes.
- [2] In malaria mosquitoes, the heterochromatin size and morphology vary significantly among species and within species (Gatti et al. 1977; Sharakhova, Stegnii, Braginets 1997; Sharakhova, Stegnii, Timofeeva 1997), possibly affecting mating behavior and fertility (Bonaccorsi et al. 1980). The mechanisms that constrain or promote evolutionary reorganizations of heterochromatin and nuclear architecture are largely unknown. What role do heterochromatin, noncoding RNAs, and nuclear architecture play in the speciation process? Can evolutionary changes in nuclear architecture explain the taxon and chromosome specificity in rates of rearrangements? Is organization and regulation of heterochromatic genes conserved in evolution? Instead of looking only at the DNA sequence (study in 1D), research efforts should take into account the chromatin structure (study in 2D) and the nuclear architecture (study in 3D).

The knowledge of the structure, function, and evolution of mosquito chromosomes will facilitate development of novel genome-based approaches for vector control. Full understanding of the mechanisms of genome rearrangement is crucial for developing novel approaches to manipulate chromosome structure and creating an artificial insect chromosome. The ecologically sound approaches to genetic modifications of insects using an artificial chromosome will significantly impact agriculture and public health.

ACKNOWLEDGMENTS

This chapter reproduced figures from the following open-access publications. Figure 1 is from (Kamali et al. 2012). Figure 3 is from (Pombi et al. 2008). Figures 2, 4, and 5 are from (Xia et al. 2010). Figure 6 is from (Sharakhova et al. 2010b).

REFERENCES

2004. Sequence and comparative analysis of the chicken genome provide unique perspectives on vertebrate evolution. *Nature* 432:695-716.
- Akhtar, A, SM Gasser. 2007. The nuclear envelope and transcriptional control. *Nat Rev Genet* 8:507-517.
- Alekseyev, MA, PA Pevzner. 2007. Are there rearrangement hotspots in the human genome? *Plos Computational Biology* 3:2111-2121.
- Alekseyev, MA, PA Pevzner. 2009. Breakpoint graphs and ancestral genome reconstructions. *Genome Res* 19:943-957.
- Anderson, AR, AA Hoffmann, SW McKechnie, PA Umina, AR Weeks. 2005. The latitudinal cline in the In(3R)Payne inversion polymorphism has shifted in the last 20 years in Australian *Drosophila melanogaster* populations. *Mol Ecol* 14:851-858.
- Antonio-Nkondjio, C, P Awono-Ambene, JC Toto, JY Meunier, S Zebaze-Kemleu, R Nyambam, CS Wondji, T Tchuinkam, D Fontenille. 2002. High malaria transmission intensity in a village close to Yaounde, the capital city of Cameroon. *J Med Entomol* 39:350-355.
- Antonio-Nkondjio, C, CH Kera, F Simard, P Awono-Ambene, M Chouaibou, T Tchuinkam, D Fontenille. 2006. Complexity of the malaria vectorial system in Cameroon: contribution of secondary vectors to malaria transmission. *J Med Entomol* 43:1215-1221.
- Antonio-Nkondjio, C, C Ndo, C Costantini, P Awono-Ambene, D Fontenille, F Simard. 2009. Distribution and larval habitat characterization of *Anopheles moucheti*, *Anopheles nili*, and other malaria vectors in river networks of southern Cameroon. *Acta Trop* 112:270-276.
- Aulard, S, L Monti, N Chaminade, F Lemeunier. 2004a. Mitotic and polytene chromosomes: comparisons between *Drosophila melanogaster* and *Drosophila simulans*. *Genetica* 120:137-150.
- Aulard, S, P Vaudin, V Ladeveze, N Chaminade, G Periquet, F Lemeunier. 2004b. Maintenance of a large pericentric inversion generated by the hobo transposable element in a transgenic line of *Drosophila melanogaster*. *Heredity* 92:151-155.
- Ayala, D, C Costantini, K Ose, GC Kamdem, C Antonio-Nkondjio, JP Agbor, P Awono-Ambene, D Fontenille, F Simard. 2009. Habitat suitability and ecological niche profile of major malaria vectors in Cameroon. *Malar J* 8:307.
- Ayala, D, MC Fontaine, A Cohuet, D Fontenille, R Vitalis, F Simard. 2011. Chromosomal inversions, natural selection and adaptation in the malaria vector *Anopheles funestus*. *Mol Biol Evol* 28:745-758.
- Ayala, FJ, M Coluzzi. 2005. Chromosome speciation: humans, *Drosophila*, and mosquitoes. *Proc Natl Acad Sci U S A* 102 Suppl 1:6535-6542.

- Baack, EJ, LH Rieseberg. 2007. A genomic view of introgression and hybrid speciation. *Curr Opin Genet Dev* 17:513-518.
- Bailey, JA, R Baertsch, WJ Kent, D Haussler, EE Eichler. 2004. Hotspots of mammalian chromosomal evolution. *Genome Biol* 5:R23.
- Baricheva, EA, M Berrios, SS Bogachev, IV Borisevich, ER Lapik, IV Sharakhov, N Stuurman, PA Fisher. 1996. DNA from *Drosophila melanogaster* beta-heterochromatin binds specifically to nuclear lamins in vitro and the nuclear envelope in situ. *Gene* 171:171-176.
- Bartova, E, A Harnicarova, J Pachernik, S Kozubek. 2005. Nuclear topography and expression of the BCR/ABL fusion gene and its protein level influenced by cell differentiation and RNA interference. *Leuk Res* 29:901-913.
- Besansky, N. 2008. Genome Analysis Of Vectorial Capacity In Major Anopheles Vectors of Malaria Parasites. *White Paper VectorBase*:1-15.
- Besansky, NJ, J Krzywinski, T Lehmann, F Simard, M Kern, O Mukabayire, D Fontenille, Y Toure, N Sagnon. 2003. Semipermeable species boundaries between *Anopheles gambiae* and *Anopheles arabiensis*: evidence from multilocus DNA sequence variation. *Proc Natl Acad Sci U S A* 100:10818-10823.
- Bhutkar, A, WM Gelbart, TF Smith. 2007. Inferring genome-scale rearrangement phylogeny and ancestral gene order: a *Drosophila* case study. *Genome Biol* 8:R236.
- Bhutkar, A, SW Schaeffer, SM Russo, M Xu, TF Smith, WM Gelbart. 2008. Chromosomal rearrangement inferred from comparisons of 12 *Drosophila* genomes. *Genetics* 179:1657-1680.
- Blanco, E, M Pignatelli, S Beltran, A Punset, S Perez-Lluch, F Serras, R Guigo, M Corominas. 2008. Conserved chromosomal clustering of genes governed by chromatin regulators in *Drosophila*. *Genome Biol* 9:R134.
- Bonaccorsi, S, G Santini, M Gatti, S Pimpinelli, M Colluzzi. 1980. Intraspecific polymorphism of sex chromosome heterochromatin in two species of the *Anopheles gambiae* complex. *Chromosoma* 76:57-64.
- Bourque, G, PA Pevzner, G Tesler. 2004. Reconstructing the genomic architecture of ancestral mammals: lessons from human, mouse, and rat genomes. *Genome Res* 14:507-516.
- Brennecke, J, AA Aravin, A Stark, M Dus, M Kellis, R Sachidanandam, GJ Hannon. 2007. Discrete small RNA-generating loci as master regulators of transposon activity in *Drosophila*. *Cell* 128:1089-1103.
- Brideau, NJ, HA Flores, J Wang, S Maheshwari, X Wang, DA Barbash. 2006. Two Dobzhansky-Muller genes interact to cause hybrid lethality in *Drosophila*. *Science* 314:1292-1295.
- Brown, JD, RJ O'Neill. 2010. Chromosomes, conflict, and epigenetics: chromosomal speciation revisited. *Annu Rev Genomics Hum Genet* 11:291-316.
- Caceres, M, A Barbadilla, A Ruiz. 1997. Inversion length and breakpoint distribution in the *Drosophila buzzatii* species complex: is inversion length a selected trait? *Evolution* 51:1149-1155.
- Caceres, M, JM Ranz, A Barbadilla, M Long, A Ruiz. 1999. Generation of a widespread *Drosophila* inversion by a transposable element. *Science* 285:415-418.
- Cassone, BJ, K Mouline, MW Hahn, BJ White, M Pombi, F Simard, C Costantini, NJ Besansky. 2008. Differential gene expression in incipient species of *Anopheles gambiae*. *Mol Ecol* 17:2491-2504.

- Charlesworth, B, JA Coyne, NH Barton. 1987. The relative rates of evolution of sex chromosomes and autosomes. *The American Naturalist* 130:113-146.
- Cheng, C, BJ White, C Kamdem, K Mockaitis, C Costantini, MW Hahn, NJ Besansky. 2012. Ecological genomics of *Anopheles gambiae* along a latitudinal cline: a population-resequencing approach. *Genetics* 190:1417-1432.
- Clark, AGMB EisenDR Smith, et al. 2007. Evolution of genes and genomes on the *Drosophila* phylogeny. *Nature* 450:203-218.
- Coghlan, A, EE Eichler, SG Oliver, AH Paterson, L Stein. 2005. Chromosome evolution in eukaryotes: a multi-kingdom perspective. *Trends Genet* 21:673-682.
- Coluzzi, M. 1992. Malaria vector analysis and control. *Parasitol Today* 8:113-118.
- Coluzzi, M, A Sabatini. 1969. Cytogenetic observations on the salt water species, *Anopheles merus* and *Anopheles melas*, of the *gambiae* complex. *Parassitologia* 11:177-187.
- Coluzzi, M, A Sabatini, A della Torre, MA Di Deco, V Petrarca. 2002. A polytene chromosome analysis of the *Anopheles gambiae* species complex. *Science* 298:1415-1418.
- Coluzzi, M, A Sabatini, V Petrarca, MA Di Deco. 1977. Behavioural divergences between mosquitoes with different inversion karyotypes in polymorphic populations of the *Anopheles gambiae* complex. *Nature* 266:832-833.
- Coluzzi, M, A Sabatini, V Petrarca, MA Di Deco. 1979. Chromosomal differentiation and adaptation to human environments in the *Anopheles gambiae* complex. *Trans R Soc Trop Med Hyg* 73:483-497.
- Cornel, AJ, FH Collins. 2000. Maintenance of chromosome arm integrity between two *Anopheles* mosquito subgenera. *J Hered* 91:364-370.
- Corona, DF, G Siriaco, JA Armstrong, N Snarskaya, SA McClymont, MP Scott, JW Tamkun. 2007. ISWI regulates higher-order chromatin structure and histone H1 assembly in vivo. *PLoS Biol* 5:e232.
- Costantini, C, D Ayala, WM Guelbeogo, et al. 2009. Living at the edge: biogeographic patterns of habitat segregation conform to speciation by niche expansion in *Anopheles gambiae*. *BMC Ecol* 9:16.
- Costantini, C, N Sagnon, E Iboudo-Sanogo, M Coluzzi, D Boccolini. 1999. Chromosomal and bionomic heterogeneities suggest incipient speciation in *Anopheles funestus* from Burkina Faso. *Parassitologia* 41:595-611.
- Coulibaly, MB, NF Lobo, MC Fitzpatrick, M Kern, O Grushko, DV Thaner, SF Traore, FH Collins, NJ Besansky. 2007. Segmental duplication implicated in the genesis of inversion 2Rj of *Anopheles gambiae*. *PLoS ONE* 2:e849.
- Dechat, T, K Pfliegerhaa, K Sengupta, T Shimi, DK Shumaker, L Solimando, RD Goldman. 2008. Nuclear lamins: major factors in the structural organization and function of the nucleus and chromatin. *Genes Dev* 22:832-853.
- della Torre, A, C Costantini, NJ Besansky, A Caccone, V Petrarca, JR Powell, M Coluzzi. 2002. Speciation within *Anopheles gambiae* - the glass is half full. *Science* 298:115-117.
- della Torre, A, L Merzagora, JR Powell, M Coluzzi. 1997. Selective introgression of paracentric inversions between two sibling species of the *Anopheles gambiae* complex. *Genetics* 146:239-244.
- Desset, S, N Buchon, C Meignin, M Coiffet, C Vaury. 2008. In *Drosophila melanogaster* the COM locus directs the somatic silencing of two retrotransposons through both Piwi-dependent and -independent pathways. *PLoS ONE* 3:e1526.

- Desset, S, C Meignin, B Dastugue, C Vaury. 2003. COM, a heterochromatic locus governing the control of independent endogenous retroviruses from *Drosophila melanogaster*. *Genetics* 164:501-509.
- Dia, I, T Diop, I Rakotoarivony, P Kengne, D Fontenille. 2003. Bionomics of *Anopheles gambiae* Giles, *An. arabiensis* Patton, *An. funestus* Giles and *An. nili* (Theobald) (Diptera: Culicidae) and transmission of *Plasmodium falciparum* in a Sudano-Guinean zone (Ngari, Senegal). *J Med Entomol* 40:279-283.
- Diabate, A, RK Dabire, K Heidenberger, J Crawford, WO Lamp, LE Culler, T Lehmann. 2008. Evidence for divergent selection between the molecular forms of *Anopheles gambiae*: role of predation. *BMC Evol Biol* 8:5.
- Diabate, A, RK Dabire, P Kengne, C Brengues, T Baldet, A Ouari, F Simard, T Lehmann. 2006. Mixed swarms of the molecular M and S forms of *Anopheles gambiae* (Diptera: Culicidae) in sympatric area from Burkina Faso. *J Med Entomol* 43:480-483.
- Diabate, A, RK Dabire, N Millogo, T Lehmann. 2007. Evaluating the effect of postmating isolation between molecular forms of *Anopheles gambiae* (Diptera: Culicidae). *J Med Entomol* 44:60-64.
- Duan, Z, M Andronescu, K Schutz, S McIlwain, YJ Kim, C Lee, J Shendure, S Fields, CA Blau, WS Noble. 2010. A three-dimensional model of the yeast genome. *Nature* 465:363-367.
- Edillo, FE, F Tripet, YT Toure, GC Lanzaro, G Dolo, CE Taylor. 2006. Water quality and immatures of the M and S forms of *Anopheles gambiae* s.s. and *An. arabiensis* in a Malian village. *Malar J* 5:35.
- Evgen'ev, MB, H Zelentsova, H Poluectova, GT Lyozin, V Veleikodvorskaja, KI Pyatkov, LA Zhivotovsky, MG Kidwell. 2000. Mobile elements and chromosomal evolution in the virilis group of *Drosophila*. *Proc Natl Acad Sci U S A* 97:11337-11342.
- Feder, JL, JB Roethele, K Filchak, J Niedbalski, J Romero-Severson. 2003. Evidence for inversion polymorphism related to sympatric host race formation in the apple maggot fly, *Rhagoletis pomonella*. *Genetics* 163:939-953.
- Filion, GJ, JG van Bommel, U Braunschweig, et al. 2010. Systematic protein location mapping reveals five principal chromatin types in *Drosophila* cells. *Cell* 143:212-224.
- Fisher, AM, P Strike, C Scott, AV Moorman. 2005. Breakpoints of variant 9;22 translocations in chronic myeloid leukemia locate preferentially in the CG-richest regions of the genome. *Genes Chromosomes Cancer* 43:383-389.
- Folle, GA. 2008. Nuclear architecture, chromosome domains and genetic damage. *Mutat Res* 658:172-183.
- Fontanillas, P, DL Hartl, M Reuter. 2007. Genome organization and gene expression shape the transposable element distribution in the *Drosophila melanogaster* euchromatin. *PLoS Genet* 3:e210.
- Fontenille, D, F Simard. 2004. Unravelling complexities in human malaria transmission dynamics in Africa through a comprehensive knowledge of vector populations. *Comp Immunol Microbiol Infect Dis* 27:357-375.
- Foster, HA, JM Bridger. 2005. The genome and the nucleus: a marriage made by evolution. Genome organisation and nuclear architecture. *Chromosoma* 114:212-229.
- Gall, JG, EH Cohen, ML Polan. 1971. Reptitive DNA sequences in *drosophila*. *Chromosoma* 33:319-344.

- Gandhi, M, M Medvedovic, JR Stringer, YE Nikiforov. 2006. Interphase chromosome folding determines spatial proximity of genes participating in carcinogenic RET/PTC rearrangements. *Oncogene* 25:2360-2366.
- Gatti, M, G Santini, S Pimpinelli, M Coluzz. 1977. Fluorescence banding techniques in the identification of sibling species of the anopheles gambiae complex. *Heredity* 38:105-108.
- Gentile, G, M Slotman, V Ketmaier, JR Powell, A Caccone. 2001. Attempts to molecularly distinguish cryptic taxa in *Anopheles gambiae* s.s. *Insect Molecular Biology* 10:25-32.
- Gimonneau, G, J Bouyer, S Morand, NJ Besansky, A Diabate, F Simard. 2010. A behavioral mechanism underlying ecological divergence in the malaria mosquito *Anopheles gambiae*. *Behavioral Ecology* 21:1087-1092.
- Gimonneau, G, M Pombi, M Choisy, S Morand, RK Dabire, F Simard. 2012. Larval habitat segregation between the molecular forms of the mosquito *Anopheles gambiae* in a rice field area of Burkina Faso, West Africa. *Med Vet Entomol* 26:9-17.
- Goidts, V, JM Szamalek, H Hameister, H Kehrer-Sawatzki. 2004. Segmental duplication associated with the human-specific inversion of chromosome 18: a further example of the impact of segmental duplications on karyotype and genome evolution in primates. *Hum Genet* 115:116-122.
- Goltsev, Y, GL Rezende, K Vranizan, G Lanzaro, D Valle, M Levine. 2009. Developmental and evolutionary basis for drought tolerance of the *Anopheles gambiae* embryo. *Dev Biol* 330:462-470.
- Gonzalez, J, JM Ranz, A Ruiz. 2002. Chromosomal elements evolve at different rates in the *Drosophila* genome. *Genetics* 161:1137-1154.
- Gordeev, MI, MN Ejov. 2004. Global warming and variation in the chromosomal composition of Siberian malaria mosquito populations. *Dokl Biol Sci* 395:136-139.
- Gordon, L, S Yang, M Tran-Gyamfi, et al. 2007. Comparative analysis of chicken chromosome 28 provides new clues to the evolutionary fragility of gene-rich vertebrate regions. *Genome Res* 17:1603-1613.
- Gray, EM, KA Rocca, C Costantini, NJ Besansky. 2009. Inversion 2La is associated with enhanced desiccation resistance in *Anopheles gambiae*. *Malar J* 8:215.
- Green, C, R Hunt. 1980. Interpretation of variation in ovarian polytene chromosomes of *Anopheles funestus* Giles, *A. parensis* Gillies, and *A. aruni*? . *Genetica* 51:187-195.
- Green, CA. 1982. Cladistic analysis of mosquito chromosome data (*Anopheles* (Cellia) *Myzomyia*). *J Hered* 73:2-11.
- Grushko, OG, MV Sharakhova, VN Stegnii, IV Sharakhov. 2009. Molecular organization of heterochromatin in malaria mosquitoes of the *Anopheles maculipennis* subgroup. *Gene* 448:192-197.
- Guelen, L, L Pagie, E Brasset, et al. 2008. Domain organization of human chromosomes revealed by mapping of nuclear lamina interactions. *Nature* 453:948-951.
- Hanney, PW. 1960. The mosquitos of Zaria Province, Northern Nigeria. *Bull Entomol Res* 51:145-171.
- Harbach, RE. 2004. The classification of genus *Anopheles* (Diptera: Culicidae): a working hypothesis of phylogenetic relationships. *Bull Entomol Res* 94:537-553.
- Heitz, E. 1934. Über a- und b-Heterochromatin sowie Konstanz und Bau der Chromomeren bei *Drosophila*. *Biol. Zbl.* 54: 588-609.

- Hochstrasser, M, JW Sedat. 1987a. Three-dimensional organization of *Drosophila melanogaster* interphase nuclei. I. Tissue-specific aspects of polytene nuclear architecture. *J Cell Biol* 104:1455-1470.
- Hochstrasser, M, JW Sedat. 1987b. Three-dimensional organization of *Drosophila melanogaster* interphase nuclei. II. Chromosome spatial organization and gene regulation. *J Cell Biol* 104:1471-1483.
- Hoffmann, AA, PJ Daborn. 2007. Towards genetic markers in animal populations as biomonitors for human-induced environmental change. *Ecology Letters* 10:63-76.
- Hoffmann, AA, L Rieseberg. 2008. Revisiting the Impact of Inversions in Evolution: From Population Genetic Markers to Drivers of Adaptive Shifts and Speciation? *Annual Review of Ecology, Evolution, and Systematics* 39:21-42.
- Hoffmann, AA, CM Sgro, AR Weeks. 2004. Chromosomal inversion polymorphisms and adaptation. *Trends Ecol Evol* 19:482-488.
- Holt, RAGM SubramanianA Halpern, et al. 2002. The genome sequence of the malaria mosquito *Anopheles gambiae*. *Science* 298:129-149.
- Hurst, LD, C Pal, MJ Lercher. 2004. The evolutionary dynamics of eukaryotic gene order. *Nat Rev Genet* 5:299-310.
- Joffe, B, H Leonhardt, I Solovei. 2010. Differentiation and large scale spatial organization of the genome. *Curr Opin Genet Dev*.
- Kamali, M, A Xia, Z Tu, IV Sharakhov. 2012. A New Chromosomal Phylogeny Supports the Repeated Origin of Vectorial Capacity in Malaria Mosquitoes of the *Anopheles gambiae* Complex. *PLoS Pathog* 8(10): e1002960. doi:10.1371/journal.ppat.1002960.
- Kerah-Hinzoumbe, C, M Peka, C Antonio-Nkondjio, I Donan-Gouni, P Awono-Ambene, A Same-Ekobo, F Simard. 2009. Malaria vectors and transmission dynamics in Goulmoun, a rural city in south-western Chad. *BMC Infect Dis* 9:71.
- Kirkpatrick, M, N Barton. 2006. Chromosome inversions, local adaptation and speciation. *Genetics* 173:419-434.
- Knight, M, W Ittiprasert, EC Odoemelam, CM Adema, A Miller, N Raghavan, JM Bridger. 2011. Non-random organization of the *Biomphalaria glabrata* genome in interphase Bge cells and the spatial repositioning of activated genes in cells co-cultured with *Schistosoma mansoni*. *Int J Parasitol* 41:61-70.
- Kozubek, S, E Lukasova, A Mareckova, M Skalníkova, M Kozubek, E Bartova, V Kroha, E Krahulcova, J Slotova. 1999. The topological organization of chromosomes 9 and 22 in cell nuclei has a determinative role in the induction of t(9,22) translocations and in the pathogenesis of t(9,22) leukemias. *Chromosoma* 108:426-435.
- Krimbas, CB, JR Powell. 1992. *Drosophila Inversion Polymorphism*. Boca Raton: CRC Press.
- Krzywinski, J, NJ Besansky. 2003. Molecular systematics of *Anopheles*: from subgenera to subpopulations. *Annu Rev Entomol* 48:111-139.
- Krzywinski, J, OG Grushko, NJ Besansky. 2006. Analysis of the complete mitochondrial DNA from *Anopheles funestus*: an improved dipteran mitochondrial genome annotation and a temporal dimension of mosquito evolution. *Mol Phylogenet Evol* 39:417-423.
- Krzywinski, J, RC Wilkerson, NJ Besansky. 2001. Toward understanding Anophelinae (Diptera, Culicidae) phylogeny: insights from nuclear single-copy genes and the weight of evidence. *Syst Biol* 50:540-556.

- Larkin, DM, G Pape, R Donthu, L Auvil, M Welge, HA Lewin. 2009. Breakpoint regions and homologous syntenic blocks in chromosomes have different evolutionary histories. *Genome Res* 19:770-777.
- Lawniczak, MK, SJ Emrich, AK Holloway, et al. 2010. Widespread divergence between incipient *Anopheles gambiae* species revealed by whole genome sequences. *Science* 330:512-514.
- Lawson, D, P Arensburger, P Atkinson, et al. 2009. VectorBase: a data resource for invertebrate vector genomics. *Nucleic Acids Res* 37:D583-587.
- Lehmann, T, A Diabate. 2008. The molecular forms of *Anopheles gambiae*: a phenotypic perspective. *Infect Genet Evol* 8:737-746.
- Lemasson, JJ, D Fontenille, L Lochouart, I Dia, F Simard, K Ba, A Diop, M Diatta, JF Molez. 1997. Comparison of behavior and vector efficiency of *Anopheles gambiae* and *An. arabiensis* (Diptera: Culicidae) in Barkedji, a Sahelian area of Senegal. *J Med Entomol* 34:396-403.
- Lemeunier, F, M Ashburner. 1984. Relationships within the *Melanogaster* Species Subgroup of the Genus *Drosophila* (Sophophora) .4. The Chromosomes of 2 New Species. *Chromosoma* 89:343-351.
- Lewin, HA, DM Larkin, J Pontius, SJ O'Brien. 2009. Every genome sequence needs a good map. *Genome Res* 19:1925-1928.
- Lobachev, KS, A Rattray, V Narayanan. 2007. Hairpin- and cruciform-mediated chromosome breakage: causes and consequences in eukaryotic cells. *Front Biosci* 12:4208-4220.
- Lyttle, TW, DS Haymer. 1992. The role of the transposable element hobo in the origin of endemic inversions in wild populations of *Drosophila melanogaster*. *Genetica* 86:113-126.
- Machado, CA, TS Haselkorn, MA Noor. 2007. Evaluation of the genomic extent of effects of fixed inversion differences on intraspecific variation and interspecific gene flow in *Drosophila pseudoobscura* and *D. persimilis*. *Genetics* 175:1289-1306.
- Mahmood, F, RK Sakai. 1984. Inversion polymorphisms in natural populations of *Anopheles stephensi*. *Can J Genet Cytol* 26:538-546.
- Manoukis, NC, JR Powell, MB Toure, A Sacko, FE Edillo, MB Coulibaly, SF Traore, CE Taylor, NJ Besansky. 2008. A test of the chromosomal theory of ecotypic speciation in *Anopheles gambiae*. *Proc Natl Acad Sci U S A* 105:2940-2945.
- Marshall, WF. 2002. Order and disorder in the nucleus. *Curr Biol* 12:R185-192.
- Mathiopoulos, KD, A della Torre, V Predazzi, V Petrarca, M Coluzzi. 1998. Cloning of inversion breakpoints in the *Anopheles gambiae* complex traces a transposable element at the inversion junction. *Proc Natl Acad Sci U S A* 95:12444-12449.
- Mathiopoulos, KD, A della Torre, F Santolamazza, V Predazzi, V Petrarca, M Coluzzi. 1999. Are chromosomal inversions induced by transposable elements? A paradigm from the malaria mosquito *Anopheles gambiae*. *Parassitologia* 41:119-123.
- Matzkin, LM, TJ Merritt, CT Zhu, WF Eanes. 2005. The structure and population genetics of the breakpoints associated with the cosmopolitan chromosomal inversion In(3R)Payne in *Drosophila melanogaster*. *Genetics* 170:1143-1152.
- Mayer, W, R Fundele, T Haaf. 2000. Spatial separation of parental genomes during mouse interspecific (*Mus musculus* x *M. spretus*) spermiogenesis. *Chromosome Res* 8:555-558.
- Mayer, W, A Smith, R Fundele, T Haaf. 2000. Spatial separation of parental genomes in preimplantation mouse embryos. *J Cell Biol* 148:629-634.

- Meaburn, KJ, CN Parris, JM Bridger. 2005. The manipulation of chromosomes by mankind: the uses of microcell-mediated chromosome transfer. *Chromosoma* 114:263-274.
- Megy, K, SJ Emrich, D Lawson, et al. 2012. VectorBase: improvements to a bioinformatics resource for invertebrate vector genomics. *Nucleic Acids Res* 40:D729-734.
- Mevel-Ninio, M, A Pelisson, J Kinder, AR Campos, A Bucheton. 2007. The flamenco locus controls the gypsy and ZAM retroviruses and is required for *Drosophila* oogenesis. *Genetics* 175:1615-1624.
- Michalak, P. 2008. Coexpression, coregulation, and cofunctionality of neighboring genes in eukaryotic genomes. *Genomics* 91:243-248.
- Michalak, P. 2009. Epigenetic, transposon and small RNA determinants of hybrid dysfunctions. *Heredity* (Edinb) 102:45-50.
- Mikkelsen, TS, MJ Wakefield, B Aken, et al. 2007. Genome of the marsupial *Monodelphis domestica* reveals innovation in non-coding sequences. *Nature* 447:167-177.
- Mongin, E, K Dewar, M Blanchette. 2009. Long-range regulation is a major driving force in maintaining genome integrity. *BMC Evol Biol* 9:203.
- Moreno, M, O Marinotti, J Krzywinski, WP Tadei, AA James, NL Achee, JE Conn. 2010. Complete mtDNA genomes of *Anopheles darlingi* and an approach to anopheline divergence time. *Malar J* 9:127.
- Mouchet, J, P Carnevale, M Coosemans, J Julvez, S Manguin, D Richard-Lenoble, J Sircoulon. 2004. *Biodiversité du paludisme dans le monde Editions*. [Biodiversity of malaria in the world Editions]. John Libbey Eurotext. Montrouge, France.
- Murphy, WJ, DM Larkin, A Everts-van der Wind, et al. 2005. Dynamics of mammalian chromosome evolution inferred from multispecies comparative maps. *Science* 309:613-617.
- Nadeau, JH, BA Taylor. 1984. Lengths of chromosomal segments conserved since divergence of man and mouse. *Proc Natl Acad Sci U S A* 81:814-818.
- Neafsey, DE, MK Lawniczak, DJ Park, et al. 2010. SNP genotyping defines complex gene-flow boundaries among African malaria vector mosquitoes. *Science* 330:514-517.
- Neusser, M, V Schubel, A Koch, T Cremer, S Muller. 2007. Evolutionarily conserved, cell type and species-specific higher order chromatin arrangements in interphase nuclei of primates. *Chromosoma* 116:307-320.
- Neves, H, C Ramos, MG da Silva, A Parreira, L Parreira. 1999. The nuclear topography of ABL, BCR, PML, and RARalpha genes: evidence for gene proximity in specific phases of the cell cycle and stages of hematopoietic differentiation. *Blood* 93:1197-1207.
- Ng, YK, W Wu, L Zhang. 2009. Positive correlation between gene coexpression and positional clustering in the zebrafish genome. *BMC Genomics* 10:42.
- Noor, MA, DA Garfield, SW Schaeffer, CA Machado. 2007. Divergence between the *Drosophila pseudoobscura* and *D. persimilis* genome sequences in relation to chromosomal inversions. *Genetics* 177:1417-1428.
- Noor, MA, KL Grams, LA Bertucci, J Reiland. 2001. Chromosomal inversions and the reproductive isolation of species. *Proc Natl Acad Sci U S A* 98:12084-12088.
- O'Grady, PM, RH Baker, CM Durando, WJ Etges, R DeSalle. 2001. Polytene chromosomes as indicators of phylogeny in several species groups of *Drosophila*. *BMC Evol Biol* 1:6.
- Orr, HA. 2005. The probability of parallel evolution. *Evolution* 59:216-220.
- Pape, T. 1992. Cladistic analyses of mosquito chromosome data in *Anopheles* subgenus *Cellia* (Diptera: Culicidae). *Mosq. Syst.* 24:1– 11.

- Peric-Hupkes, D, W Meuleman, L Pagie, et al. 2010. Molecular maps of the reorganization of genome-nuclear lamina interactions during differentiation. *Mol Cell* 38:603-613.
- Petrarca, V, JC Beier. 1992. Intraspecific chromosomal polymorphism in the *Anopheles gambiae* complex as a factor affecting malaria transmission in the Kisumu area of Kenya. *Am J Trop Med Hyg* 46:229-237.
- Petrarca, V, AD Nugud, MA Ahmed, AM Haridi, MA Di Deco, M Coluzzi. 2000. Cytogenetics of the *Anopheles gambiae* complex in Sudan, with special reference to *An. arabiensis*: relationships with East and West African populations. *Med Vet Entomol* 14:149-164.
- Pickersgill, H, B Kalverda, E de Wit, W Talhout, M Fornerod, B van Steensel. 2006. Characterization of the *Drosophila melanogaster* genome at the nuclear lamina. *Nat Genet* 38:1005-1014.
- Pombi, M. 2005. *Anopheles gambiae* larval habitats in an arid savanna village of Burkina faso: Characterization of bionomical parameters and potential markers of ecological niche partitioning among three sympatric taxa of the complex. PhD Dissertation. La Sapienza.
- Pombi, M, B Caputo, F Simard, MA Di Deco, M Coluzzi, A della Torre, C Costantini, NJ Besansky, V Petrarca. 2008. Chromosomal plasticity and evolutionary potential in the malaria vector *Anopheles gambiae* sensu stricto: insights from three decades of rare paracentric inversions. *BMC Evol Biol* 8:309.
- Ponts, N, EY Harris, J Prudhomme, I Wick, C Eckhardt-Ludka, GR Hicks, G Hardiman, S Lonardi, KG Le Roch. 2010. Nucleosome landscape and control of transcription in the human malaria parasite. *Genome Res* 20:228-238.
- Postlethwait, JH, IG Woods, P Ngo-Hazelett, YL Yan, PD Kelly, F Chu, H Huang, A Hill-Force, WS Talbot. 2000. Zebrafish comparative genomics and the origins of vertebrate chromosomes. *Genome Res* 10:1890-1902.
- Powell, JR, V Petrarca, A della Torre, A Caccone, M Coluzzi. 1999. Population structure, speciation, and introgression in the *Anopheles gambiae* complex. *Parassitologia* 41:101-113.
- Ranz, JM, D Maurin, YS Chan, M von Grotthuss, LW Hillier, J Roote, M Ashburner, CM Bergman. 2007. Principles of Genome Evolution in the *Drosophila melanogaster* Species Group. *PLoS Biol* 5:e152.
- Richards, S, Y Liu, BR Bettencourt, et al. 2005. Comparative genome sequencing of *Drosophila pseudoobscura*: chromosomal, gene, and cis-element evolution. *Genome Res* 15:1-18.
- Riehle, MM, WM Guelbeogo, A Gneme, et al. 2011. A cryptic subgroup of *Anopheles gambiae* is highly susceptible to human malaria parasites. *Science* 331:596-598.
- Rieseberg, LH. 2001. Chromosomal rearrangements and speciation. *Trends Ecol Evol* 16:351-358.
- Rocca, KA, EM Gray, C Costantini, NJ Besansky. 2009. 2La chromosomal inversion enhances thermal tolerance of *Anopheles gambiae* larvae. *Malar J* 8:147.
- Rodriguez Delgado, CL, PD Waters, C Gilbert, TJ Robinson, JA Graves. 2009. Physical mapping of the elephant X chromosome: conservation of gene order over 105 million years. *Chromosome Res.*
- Saito, K, MC Siomi. 2010. Small RNA-mediated quiescence of transposable elements in animals. *Dev Cell* 19:687-697.
- Salcedo-Amaya, AM, MA van Driel, BT Alako, et al. 2009. Dynamic histone H3 epigenome marking during the intraerythrocytic cycle of *Plasmodium falciparum*. *Proc Natl Acad Sci U S A* 106:9655-9660.

- Schaeffer, SW, MP Goetting-Minesky, M Kovacevic, JR Peoples, JL Graybill, JM Miller, K Kim, JG Nelson, WW Anderson. 2003. Evolutionary genomics of inversions in *Drosophila pseudoobscura*: evidence for epistasis. *Proc Natl Acad Sci U S A* 100:8319-8324.
- Schaeffer, SW, DCW Group. 2008. Polytene Chromosomal Maps of 11 *Drosophila* species: The order of genomic scaffolds inferred from genetic and physical maps. *Genetics* 179:1-53.
- Scherthan, H. 2007. Telomere attachment and clustering during meiosis. *Cell Mol Life Sci* 64:117-124.
- Sharakhov, I, O Braginets, O Grushko, et al. 2004. A microsatellite map of the African human malaria vector *Anopheles funestus*. *J Hered* 95:29-34.
- Sharakhov, IV, AC Serazin, OG Grushko, et al. 2002. Inversions and gene order shuffling in *Anopheles gambiae* and *A. funestus*. *Science* 298:182-185.
- Sharakhov, IV, MV Sharakhova, CM Mbogo, LL Koekemoer, G Yan. 2001. Linear and spatial organization of polytene chromosomes of the African malaria mosquito *Anopheles funestus*. *Genetics* 159:211-218.
- Sharakhov, IV, BJ White, MV Sharakhova, J Kayondo, NF Lobo, F Santolamazza, A Della Torre, F Simard, FH Collins, NJ Besansky. 2006. Breakpoint structure reveals the unique origin of an interspecific chromosomal inversion (2La) in the *Anopheles gambiae* complex. *Proc Natl Acad Sci U S A* 103:6258-6262.
- Sharakhova, M, MP Hammond, NF Lobo, J Krzywinski, MF Unger, ME Hillenmeyer, RV Bruggner, E Birney, FH Collins. 2007. Update of the *Anopheles gambiae* PEST genome assembly. *Genome Biol* 8:R5.
- Sharakhova, MV, C Antonio-Nkondjio, A Xia, C Ndo, P Awono-Ambene, F Simard, IV Sharakhov. 2010a. Cytogenetic map for *Anopheles nili*: Application for population genetics and comparative physical mapping. *Infect Genet Evol* 10.1016/j.meegid.2010.06.015.
- Sharakhova, MV, P George, IV Brusentsova, SC Leman, JA Bailey, CD Smith, IV Sharakhov. 2010b. Genome mapping and characterization of the *Anopheles gambiae* heterochromatin. *BMC Genomics* 11:459.
- Sharakhova, MV, IV Sharakhov, LG Gosteva, IV Kotlikova, VN Stegnii. 2000. Pericentromeric and intercalary alpha-heterochromatin of polytene chromosomes of the malaria mosquito. *Genetika* 36:175-181.
- Sharakhova, MV, VN Stegnii, OP Braginets. 1997. Interspecies differences in the ovarian trophocyte precentromere heterochromatin structure and evolution of the malaria mosquito complex *Anopheles maculipennis*. *Genetika* 33:1640-1648.
- Sharakhova, MV, VN Stegnii, OV Timofeeva. 1997. Polymorphism of pericentromere heterochromatin of polytene chromosomes of ovarian trophocytes in natural populations of the malaria mosquito *Anopheles messeae* Fall. *Genetika* 33:281-283.
- Sharakhova, MV, A Xia, SC Leman, IV Sharakhov. 2011. Arm-specific dynamics of chromosome evolution in malaria mosquitoes. *BMC Evol Biol* 11:91.
- Singh, PB, SD Georgatos. 2002. HP1: facts, open questions, and speculation. *J Struct Biol* 140:10-16.
- Sinkins, SP, F Gould. 2006. Gene drive systems for insect disease vectors. *Nat Rev Genet* 7:427-435.
- Slotman, M, A Della Torre, JR Powell. 2004. The genetics of inviability and male sterility in hybrids between *Anopheles gambiae* and *An. arabiensis*. *Genetics* 167:275-287.

- Slotman, M, A Della Torre, JR Powell. 2005. Female sterility in hybrids between *Anopheles gambiae* and *A. arabiensis*, and the causes of Haldane's rule. *Evolution* 59:1016-1026.
- Stegnii, VN. 1987. Systemic reorganization of the architectonics of polytene chromosomes in the onto- and phylogenesis of malarial mosquitoes. II. Species specificity in the pattern of chromosome relations with the nuclear envelope of nutrient ovarian cells. *Genetika* 23:1194-1199.
- Stegnii, VN, IE Vasserlauf. 1994. Species architecture of generative tissue chromosomes and problems of phylogenetic relationships in the melanogaster subgroup of the *Drosophila* genus (*Sophophora*). *Genetika* 30:478-483.
- Takken, W, BGJ Knols. 2009. Malaria vector control: current and future strategies. *Trends Parasitol* 25:101-104.
- Talbert, PB, TD Bryson, S Henikoff. 2004. Adaptive evolution of centromere proteins in plants and animals. *J Biol* 3:18.
- Tanabe, H, S Muller, M Neusser, J von Hase, E Calcagno, M Cremer, I Solovei, C Cremer, T Cremer. 2002. Evolutionary conservation of chromosome territory arrangements in cell nuclei from higher primates. *Proc Natl Acad Sci U S A* 99:4424-4429.
- Taye, A, M Hadis, N Adugna, D Tilahun, RA Wirtz. 2006. Biting behavior and *Plasmodium* infection rates of *Anopheles arabiensis* from Sille, Ethiopia. *Acta Trop* 97:50-54.
- Tesler, G. 2002. GRIMM: genome rearrangements web server. *Bioinformatics* 18:492-493.
- Therman, E, B Susman, C Denniston. 1989. The nonrandom participation of human acrocentric chromosomes in Robertsonian translocations. *Ann Hum Genet* 53:49-65.
- Toure, YT, V Petrarca, SF Traore, A Coulibaly, HM Maiga, O Sankare, M Sow, MA Dideco, M Coluzzi. 1994. Ecological Genetic-Studies in the Chromosomal Form Mopti of *Anopheles-Gambiae* S-Str in Mali, West-Africa. *Genetica* 94:213-223.
- Toure, YT, V Petrarca, SF Traore, A Coulibaly, HM Maiga, O Sankare, M Sow, MA DiDeco, M Coluzzi. 1998. The distribution and inversion polymorphism of chromosomally recognized taxa of the *Anopheles gambiae* complex in Mali, West Africa. *Parassitologia* 40:477-511.
- Van Bortel, W, RE Harbach, HD Trung, P Roelants, T Backeljau, M Coosemans. 2001. Confirmation of *Anopheles varuna* in vietnam, previously misidentified and mistargeted as the malaria vector *Anopheles minimus*. *Am J Trop Med Hyg* 65:729-732.
- Vermaak, D, S Henikoff, HS Malik. 2005. Positive selection drives the evolution of rhino, a member of the heterochromatin protein 1 family in *Drosophila*. *PLoS Genet* 1:96-108.
- von Grotthuss, M, M Ashburner, JM Ranz. 2010. Fragile regions and not functional constraints predominate in shaping gene organization in the genus *Drosophila*. *Genome Res* 20:1084-1096.
- Wang, G, LA Christensen, KM Vasquez. 2006. Z-DNA-forming sequences generate large-scale deletions in mammalian cells. *Proc Natl Acad Sci U S A* 103:2677-2682.
- Wang-Sattler, R, S Blandin, Y Ning, et al. 2007. Mosaic Genome Architecture of the *Anopheles gambiae* Species Complex. *PLoS ONE* 2:e1249.
- Weetman, D, CS Wilding, K Steen, J Pinto, MJ Donnelly. 2012. Gene flow-dependent genomic divergence between *Anopheles gambiae* M and S forms. *Mol Biol Evol* 29:279-291.
- White, BJ, C Cheng, D Sangare, NF Lobo, FH Collins, NJ Besansky. 2009. The population genomics of trans-specific inversion polymorphisms in *Anopheles gambiae*. *Genetics* 183:275-288.

- White, BJ, C Cheng, F Simard, C Costantini, NJ Besansky. 2010. Genetic association of physically unlinked islands of genomic divergence in incipient species of *Anopheles gambiae*. *Mol Ecol*.
- Wondji, C, S Frederic, V Petrarca, J Etang, F Santolamazza, A Della Torre, D Fontenille. 2005. Species and populations of the *Anopheles gambiae* complex in Cameroon with special emphasis on chromosomal and molecular forms of *Anopheles gambiae* s.s. *J Med Entomol* 42:998-1005.
- Xia, A, M Sharakhova, S Leman, Z Tu, J Bailey, C Smith, IV Sharakhov. 2010. Genome landscape and evolutionary plasticity of chromosomes in malaria mosquitoes. *PLoS ONE* 5:e10592.
- Xia, A, MV Sharakhova, IV Sharakhov. 2008. Reconstructing ancestral autosomal arrangements in the *Anopheles gambiae* complex. *J Comput Biol* 15:965-980.
- Yang, F, PC O'Brien, BS Milne, AS Graphodatsky, N Solanky, V Trifonov, W Rens, D Sargan, MA Ferguson-Smith. 1999. A complete comparative chromosome map for the dog, red fox, and human and its integration with canine genetic maps. *Genomics* 62:189-202.
- Zhimulev, IF. 1998. Polytene chromosomes, heterochromatin, and position effect variegation. *Adv Genet* 37:1-566.

Chapter 5

TRANSCRIPTIONAL DIFFERENTIATION ACROSS THE FOUR SUBSPECIES OF *DROSOPHILA MOJAVENSIS*

Luciano M. Matzkin¹ and Therese A. Markow^{2,3}

¹University of Alabama in Huntsville, Department of Biological Sciences,
Huntsville, AL, US

²University of California San Diego, Section of Molecular and Cell Biology,
La Jolla, CA, US

³Centro de Investigaciones y Estudios Avanzados, Laboratorio Nacional de
Genómica de la Biodiversidad

ABSTRACT

Local adaptation can play a fundamental role in the isolation of populations. While less well-studied than differentiation in sequence variation, changes in transcriptional variation during speciation also are fundamental to the evolutionary process. *Drosophila mojavensis* offers an unprecedented opportunity to examine the role of transcriptional differentiation in local adaptation. *Drosophila mojavensis* is a cactophilic fly composed of four ecologically distinct subspecies that inhabit the deserts of western North America. Each of the four subspecies utilizes necrotic tissue of different cactus host species characterized by distinct chemical profiles. The subspecies in Baja California, Mexico uses *Stenocereus gummosus* (Agria), in mainland Sonora it uses *S. thurberi* (Organ Pipe), in the Mojave Desert the host is *Ferocactus cylindraceus* (Red Barrel) and in Santa Catalina Island, USA, *Opuntia littoralis* (Prickly Pear) is the host. In this chapter we examine how the adaptation to the different environmental conditions across the four subspecies have shaped their transcriptional profiles. Using complete *D. mojavensis* genome microarrays we examined the transcriptome of third instar larvae from all four subspecies reared in standard laboratory media free of necrotic cactus-derived compounds. This experimental strategy focused on differences between constitutively expressed genes and not genes induced by necrotic cactus-derived compounds. The subspecies exhibited significant differential expression of genes that likely underlie the adaptation to different cactus hosts, such as detoxification genes (Glutathione S-transferases, Cytochrome P450s and UDP-Glycosyltransferases) and chemosensory genes (Odorant Receptors, Gustatory Receptors and Odorant Binding Proteins).

INTRODUCTION

Increasing levels of genetic isolation between populations can lead to the formation of new species. The mechanisms involved could be broadly characterized as pre-zygotic or post-zygotic (Coyne and Orr 2004). Although natural selection can play an active role in maintaining the isolation (e.g., hybrid inviability), natural selection does not necessarily have to be implicated in the genetic divergence of the populations (Dobzhansky-Muller incompatibility model) (Dobzhansky 1937; Muller 1942). In certain cases local adaptation can amplify the divergence across populations and hence accelerate the speciation process (Funk, Nosil, and Etges 2006; Nosil 2007; Schluter and Conte 2009; Nosil 2012). Among many other characters, the pattern of variation of the transcriptome can be shaped by the adaptation to local ecological conditions. Knowledge of the regulatory differentiation present in ecologically distinct populations therefore informs our understanding the role of transcriptional changes in speciation.

Cactophilic *Drosophila* offer a powerful system to assess how a combination of geographic isolation, local adaptation and in some cases sexual selection can play a critical role in speciation. One such example is that of the North American endemic cactophilic species, *D. mojavensis*. Similar to all cactophiles, *D. mojavensis* feeds, oviposits and develops in the necrotic tissues of certain cactus species (Fellows and Heed 1972; Heed 1978; Heed 1982; Ruiz, Heed, and Wasserman 1990). In general all *Drosophila* are saprophytic, and just like others, *D. mojavensis* feeds on the several yeast species that are known to inhabit the cactus necroses (Starmer 1982b; Starmer 1982a; Fogleman and Starmer 1985; Starmer et al. 1990). In the process of consuming yeasts the flies also ingest and are exposed to the tissues and chemical composition of the cactus host (Kircher 1982). *Drosophila mojavensis* has relatively recently (<0.5 million years ago) diverged from its cactophilic sister species, *D. arizonae* (Heed 1978; Machado et al. 2007; Matzkin 2008; Smith et al. 2012). Currently, *D. mojavensis* is composed of four geographically (Figure 1) and ecologically distinct subspecies (Pfeiler, Castrezana, and Reed 2009). Each subspecies utilizes a distinct cactus host, each characterized by distinct microflora and chemical profiles. Two of the subspecies, Baja California and mainland Sonora (hereafter Sonora), utilize columnar cacti belonging to the same genus, *Stenocereus* (*S. gummosus* and *S. thurberi*, respectively). The subspecies in the Mojave Desert utilizes Red Barrel (*Ferocactus cylindraceus*), while the subspecies in Santa Catalina Island utilized Prickly Pear (*Opuntia littoralis*).

Given the toxic nature of some of the compounds found in the cactus necroses that are inhabited by *D. mojavensis*, much of the earlier transcriptional work has focused on the dietary induction of cactus-related compounds (Matzkin et al. 2006; Matzkin 2012). This approach has helped identify genes whose expression and pattern of sequence variation appear to have been shaped by local adaptation (Matzkin 2008). In this chapter our goal was to remove the possible confounding effects of the induction of genes in response to cactus-derived compounds, and thus focus on transcriptionally fixed differences across the subspecies. Using *D. mojavensis* specific microarrays, we investigated the transcriptional profile of third instar larvae (reared in cactus-free media) from 22 isofemale lines representative of all four subspecies. We observed that genes associated with metabolism are a major component of the transcriptional differences across the host subspecies. Surprisingly, several chemosensory genes were among those who

exhibited significant within subspecies variation, although many of these also exhibited between subspecies differences.

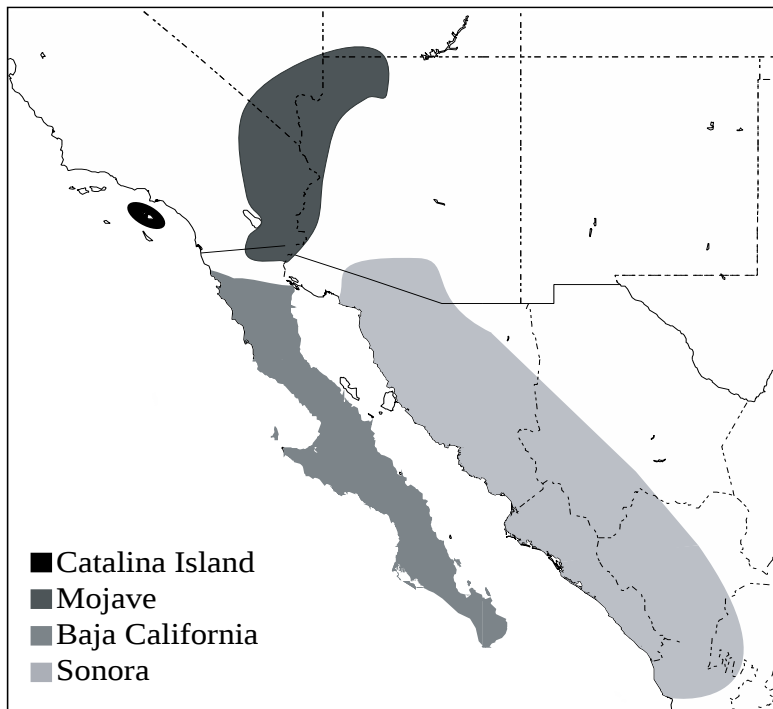


Figure 1. Distribution of the four *D. mojavensis* subspecies.

ANALYSIS OF TRANSCRIPTIONAL DIVERGENCE

A total of 22 isofemale lines were used in this chapter, six each for the Catalina, Sonora and Mojave subspecies and four from Baja California (Table 1). Isofemale lines were reared in banana-molasses media in vials. A generation prior to the experiment five females and five males were placed in a banana-molasses 8-dram vial with granules of live yeast for 24 hours. After this period the ten adults were removed and placed in new vials for next 24 hours. A total of five of these replicate vials were established per each of the 22 isofemale lines. Approximately eight days after oviposition third instar wandering larvae were collected. RNA was extracted for a total of two replicates per each of the 22 isofemale lines. All samples were hybridized onto a custom *D. mojavensis* microarray based on the previously sequenced *D. mojavensis* genome (*Drosophila* 12 Genomes Consortium 2007). As originally described in Bono *et al.* (2011) the array consists of 71,998 60 oligonucleotide probes representing 14,519 annotated *D. mojavensis* genes. The large majority (96%) of the genes in the microarray were represented by 6 probes each. Expression intensities were first normalized using the Robust Multichip Average (RMA) method (Bolstad *et al.* 2003; Irizarry *et al.* 2003). Statistical analysis of the $\log_2$ transformed data was performed using a two-step mixed model ANOVA (Wolfinger *et al.* 2001). The first step of this method is a global (data set wide) analysis that removes probe-

and hybridization-specific effects, while the second step is a gene-specific analysis with subspecies and lines-within-subspecies as factors (also including microarray as a random factor). Given the 14,519 tests performed, significance was determined using the False Discovery Rate (FDR) method (Storey and Tibshirani 2003).

Table 1. Collection sites and cactus host for lines used

Line	Collecting site	Subspecies	Cactus host
ANZA-0402-3	Anza-Borrego Desert (CA, USA)	Mojave	Red Barrel (<i>F. cylindraceus</i>)
ANZA-0402-4			
ANZA-0402-5			
ANZA-0402-8			
ANZA-0402-10			
ANZA-0402-17			
CI-1002-3	Santa Catalina Island Conservancy (CA, USA)	Catalina Island	Prickly Pear (<i>O. littoralis</i>)
CI-1002-6			
CI-1002-8			
CI-1002-9			
CI-1002-23			
CI-1002-27			
MJBC-35	La Paz (BC, Mexico)	Baja California	Agria (<i>S. gummosus</i>)
MJBC-103			
MJBC-155			
MJBC-216			
OPNM-407-2	Organ Pipe National Monument (AZ, USA)	Mainland Sonora	Organ Pipe (<i>S. thurberi</i>)
OPNM-407-3			
OPNM-407-5			
OPNM-407-6			
OPNM-407-8			
OPNM-407-10			

Two-way hierarchical clustering using the Ward method was performed for the differentially expressed genes. Clustering was first performed on mean expression intensity for each isofemale line and then by isofemale line. Analysis of the overrepresentation gene ontology (GO) terms was determined using the gene ontology enrichment analysis and visualization tool, GOrilla (Eden et al. 2007; Eden et al. 2009). Annotations of the *D. mojavensis* genome were based on the most current FlyBase annotation set (version FB2012-02) and those described in Matzkin (2012). Of the 14,519 *D. mojavensis* genes in the microarray we were able to obtain orthologous calls to *D. melanogaster* for just 10,685 genes. This smaller set of genes was used as the background gene set to test for overrepresentation of GO terms. For the purpose of examining overrepresented GO terms we set the *P*-value at less than 1×10^{-4} .

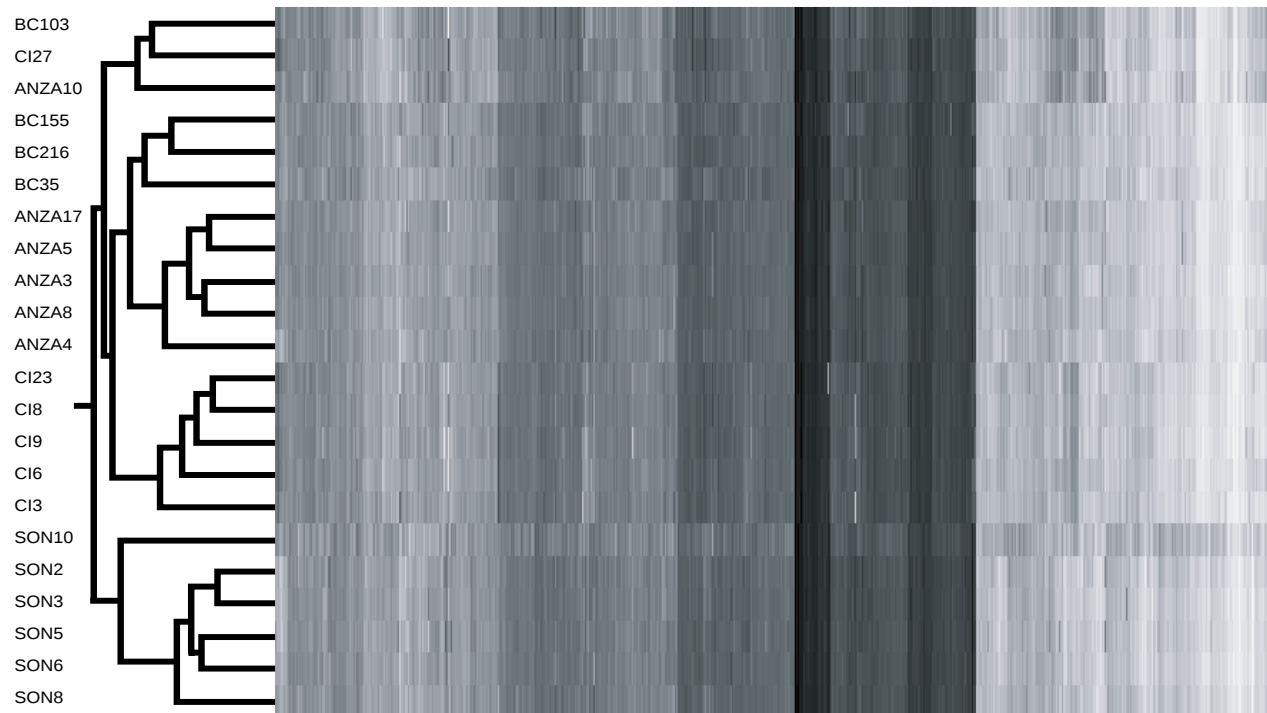


Figure 2. Two-level hierarchical clustering of expression intensities for the 14,519 genes examined. Expression intensities range from black (highest expression) to white (lowest expression). The tree indicates the transcriptional clustering across the 22 isofemale lines studied.

The entire transcriptional profile for each isofemale line can be observed in Figure 2. All expression data have been placed in the Gene Expression Omnibus under series entry #GSE41155. Although there are three exceptions (MJBC103, CI27 and ANZA10) all the isofemale lines tend to group together according to host subspecies (Figure 2). Genes with significant expression differences between and within subspecies are shown in Table 2. Regardless of the FDR cutoff used, the number of genes that exhibited a significant subspecies effect was roughly more than twice that showing significant within subspecies variation. We were interested in examining the most robust of differences between the subspecies and within subspecies and hence chose to use an FDR cutoff of 0.001 to assign significance. Even with this conservative FDR level, a total of 3,092 genes exhibited a significant subspecies effect, of which 655 also had a significant within subspecies effect (Figure 3). Among the genes with significant between subspecies differences include those involved in detoxification (Glutathione S-transferases, Cytochrome P450s, and UDP-Glycosyltransferases) and chemosensory (Odorant Receptors, Gustatory Receptors and Odorant Binding Proteins) (Table 3). For 660 genes we observed only a within subspecies effect (Figure 3). To examine subspecies pair differences, we performed a *post-hoc* test setting the FDR at 0.1% (Table 4). Comparisons using the mainland Sonora subspecies have the greatest number of observed transcriptional differences, while the comparison between Catalina Island and Baja California had the fewest (Table 4).

Table 2. Number of significant genes with significant subspecies and within subspecies variation [Line(Subspecies)] using different FDR cutoff values

	FDR		
	0.05	0.01	0.001
Subspecies	6343	4705	3092
Line(Subspecies)	2788	1971	1315

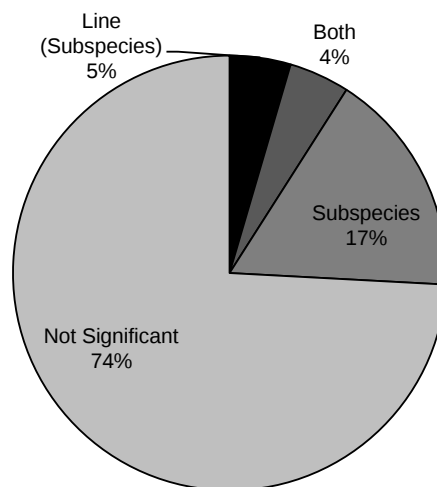


Figure 3. Proportion of genes with a significant within subspecies effect [Line(Subspecies)], subspecies effect, both within and between subspecies effects and not significant using and FDR cutoff value of 0.001.

Of the 3,092 genes with significant between subspecies differences, 2,033 had and orthologous calls to *D. melanogaster*. Results of the analysis of overrepresented Molecular Function and Biological Process GO terms are shown Figure 4. Most of the overrepresented GO terms appear to be associated with some aspect of metabolism. Also overrepresented were members of important detoxification gene families such as Glutathione S-transferases and Cytochrome P450s (within the heme-binding Biological Process GO terms). On the other hand, GO terms of genes that exhibited significant within subspecies variation (1,315 of which 744 had orthologous calls to *D. melanogaster*) were largely associated with chemosensory perception (Figure 5). The chromosomal location of the significant genes (Table 5) is significantly different ($\chi^2 = 14.2$, $df = 4$, $P = 0.0068$) from the expectation given the distribution of all genes in the *D. mojavensis* genome.

Table 3. Genes with significant between subspecies differences belonging to Gustatory Receptor (GR), Odorant Receptor (OR), Odorant Binding Protein (OBP), Glutathione S-Transferase (GST), Cytochrome P450 (P450) and UDP-Glycosyltransferase (UGT) gene families

<i>D. moj</i> Symbol	<i>D. moj</i> name	<i>D. mel</i> Ortholog	Gene Family
Dmoj\GI11424		Gr61a	GR
Dmoj\GI11499		Gr43b	GR
Dmoj\GI12801		Gr63a	GR
Dmoj\GI15700		Gr2a	GR
Dmoj\GI17782		Gr28a	GR
Dmoj\GI20824		Gr43a	GR
Dmoj\GI22003		Gr94a	GR
Dmoj\GI23037		Gr98b	GR
Dmoj\GI11973		CG6776	GST
Dmoj\GI11974		CG6781	GST
Dmoj\GI14608		CG1681	GST
Dmoj\GI16623		GstE2	GST
Dmoj\GI19388		GstE10	GST
Dmoj\GI19515		CG16936	GST
Dmoj\GI20072		CG4688	GST
Dmoj\GI20122		GstE6	GST
Dmoj\GI20123	Dmoj\GstE6b		GST
Dmoj\GI20124		GstE9	GST
Dmoj\GI22354	Dmoj\GstD1d	GstD1d	GST
Dmoj\GI22356		GstD10	GST
Dmoj\GI23193	Dmoj\GstD1b		GST
Dmoj\GI23194	Dmoj\GstD1e		GST
Dmoj\GI23195		GstD2	GST
Dmoj\GI23196		CG17639	GST
Dmoj\GI23596		CG9363	GST
Dmoj\GI17488	Dmoj\Obp28a	Pbprp5	OBP
Dmoj\GI15510		Obp8a	OBP
Dmoj\GI19754		Obp47a	OBP
Dmoj\GI19915		Obp49a	OBP
Dmoj\GI20270		Obp44a	OBP
Dmoj\GI21087		Obp56h	OBP
Dmoj\GI23726		Obp93a	OBP
Dmoj\GI14836		Or1a	OR
Dmoj\GI16944		Or13a	OR
Dmoj\GI17592	Dmoj\Or67a-1		OR
Dmoj\GI17593	Dmoj\Or67a-2		OR

Table 3. (Continued)

<i>D. moj</i> Symbol	<i>D. moj</i> name	<i>D. mel</i> Ortholog	Gene Family
Dmoj\GI19019		Or59b	OR
Dmoj\GI19311		Or49b	OR
Dmoj\GI19887		Or45b	OR
Dmoj\GI23263	Dmoj\Or85a-1		OR
Dmoj\GI23327	Dmoj\OrN2-2		OR
Dmoj\GI23643		Orco	OR
Dmoj\GI23646		Or83a	OR
Dmoj\GI23916		Or85d	OR
Dmoj\GI24760		Or33c	OR
Dmoj\GI10234		Cyp313b1	P450
Dmoj\GI11220		Cyp318a1	P450
Dmoj\GI12456		Cyp4d8	P450
Dmoj\GI12535		Cyp4d20	P450
Dmoj\GI13002		Cyp12d1-d	P450
Dmoj\GI15489		Cyp6v1	P450
Dmoj\GI16990		Cyp309a1	P450
Dmoj\GI17558		Cyp28a5	P450
Dmoj\GI18694		Cyp4e2	P450
Dmoj\GI18702		Cyp6a17	P450
Dmoj\GI18705		Cyp317a1	P450
Dmoj\GI20052			P450
Dmoj\GI20196		Cyp49a1	P450
Dmoj\GI20222		Cyp9h1	P450
Dmoj\GI20230		Cyp6a18	P450
Dmoj\GI20372		Cyp4p1	P450
Dmoj\GI20893	Dmoj\Cyp6a21b		P450
Dmoj\GI20894		Cyp6a21	P450
Dmoj\GI21924		Cyp4ac1	P450
Dmoj\GI22127		Cyp4c3	P450
Dmoj\GI23350		Cyp304a1	P450
Dmoj\GI24047		Cyp12e1	P450
Dmoj\GI24725		Cyp9f2	P450
Dmoj\GI10120		Ugt86Da	UGT
Dmoj\GI17057		Ugt37c1	UGT
Dmoj\GI17058		Ugt36Bb	UGT
Dmoj\GI17522		Ugt37a1	UGT
Dmoj\GI17523		Ugt37b1	UGT
Dmoj\GI19212		CG15661	UGT
Dmoj\GI19214		CG4302	UGT
Dmoj\GI22626		Ugt86Dj	UGT
Dmoj\GI22627		Ugt35a	UGT
Dmoj\GI22630		Ugt86Dd	UGT

Table 4. Number of significant transcriptional differences (FDR = 0.1%) between pairs of the *D. mojavensis* subspecies

	Catalina Island	Mojave	Baja California	Sonora
Catalina Island	-			
Mojave	861	-		
Baja California	396	580	-	
Sonora	2,493	1,263	837	-

Table 5. Chromosomal location of the significant differentially expressed between subspecies

Chromosome	Muller Element	Observed	Expected ¹
1	A	434	485.6
2	E	765	702.0
3	B	567	537.9
4	D	540	561.6
5	C	555	573.9

¹ Based upon the total number of genes in the respective chromosomes in the *D. mojavensis* genome.

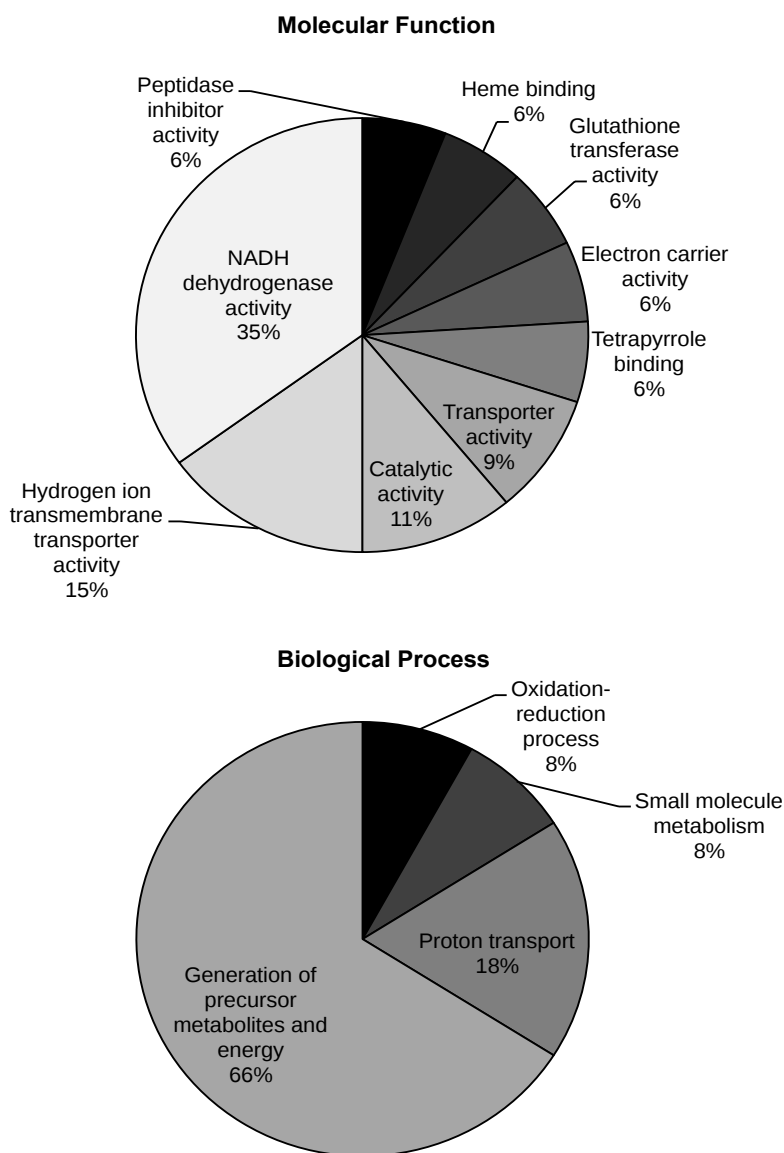


Figure 4. Composition of the overrepresented Molecular Function and Biological Process GO terms for those genes who exhibited a significant between subspecies effect.

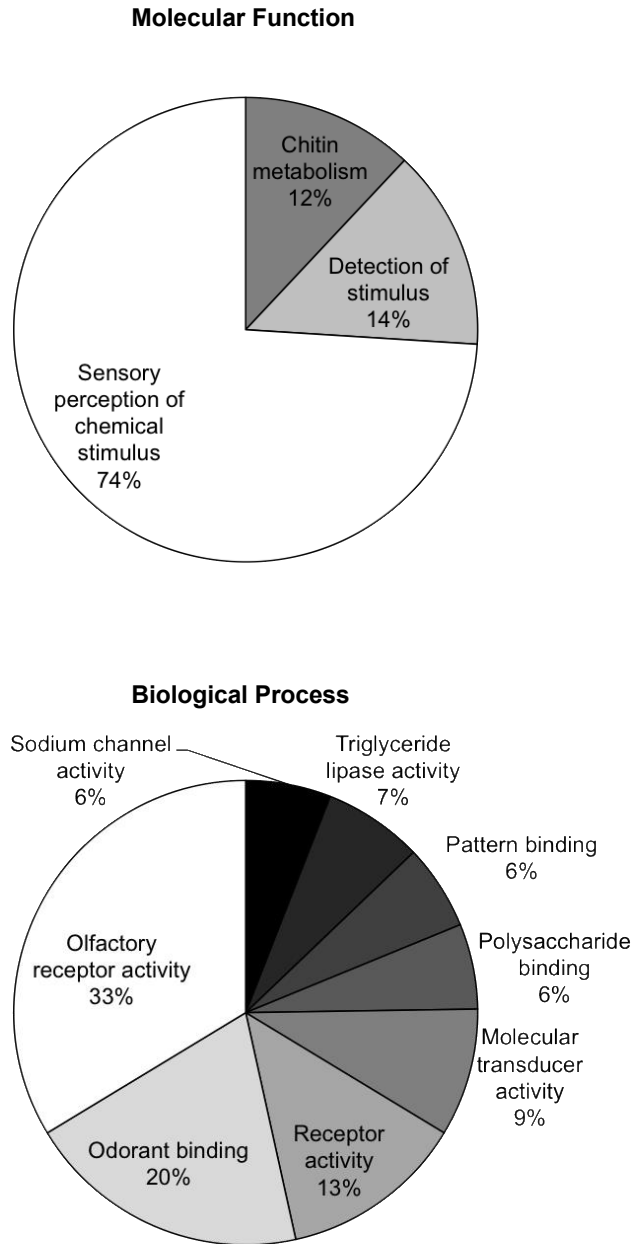


Figure 5. Composition of the overrepresented Molecular Function and Biological Process GO terms for those genes who exhibited a significant within subspecies effect.

ECOLOGICAL DIVERGENCE IN THE *D. MOJAVENSIS* TRANSCRIPTOME

The four host subspecies of *D. mojavensis* are presented by distinct ecological conditions, both biotic and abiotic in nature. These factors, among others (e.g., geographic isolation and sexual selection), have contributed to the morphological/structural (Etges et al. 2009; Pfeiler, Castrezana, and Reed 2009), life history (Etges and Heed 1987; Etges 1990), behavioral (Krebs

and Markow 1989; Markow 1991), molecular (Matzkin 2004; Machado et al. 2007; Matzkin 2008; Smith et al. 2012), biochemical (Matzkin 2005) and transcriptional (Matzkin et al. 2006; Matzkin 2012) variation in this species. Here we have identified the inherent modifications in the transcriptional profiles of the four *D. mojavensis* subspecies, independent of their diet.

Transcriptional regulatory differences accumulate as a function of isolation, having arisen by *cis* changes followed by *cis* and *trans* coevolution (Wittkopp, Haerum, and Clark 2004; Gordon and Ruvinsky 2012). As is clear from the clustering of the four subspecies by the entire transcriptome (Fig. 1), these subspecies have undergone a significant degree of transcriptional evolution. While we cannot specify the particular evolutionary force(s) that shaped the observed transcriptional divergence, the geographic isolation between the subspecies and historical bottlenecks (Smith et al. 2012) suggest that genetic drift could have influenced the transcriptional divergence (Stone and Wray 2001; Wray et al. 2003; Lynch, Scofield, and Hong 2005). At the same time, the known ecological differences between the subspecies ultimately can point to those genes whose expression difference may have been shaped by local adaptation. It is this adaptation to local conditions, partly via transcriptional changes, that contributes to the reduced success of migrant individuals, therefore aiding in the genetic isolation of the subspecies.

The necrotic cactus habitats utilized by each of the four host subspecies have marked chemical differences (Kircher 1982). In addition to the chemical composition of the different cactus species, the microfloral communities are critical in creating the chemical profile of what larvae and adult flies ingest (Starmer 1982b; Starmer 1982a; Fogleman and Starmer 1985; Starmer et al. 1990). Among these chemical differences are nutritional compounds, such as carbohydrates and lipids (Vacek 1979; Kircher 1982; Fogleman and Abril 1990). Of the 3,092 overall transcriptional differences across the subspecies, the significant majority had some role in metabolism (Fig. 3). Observed differences in metabolism-related genes observed thus may directly reflect adaptation to the different nutritional compositions of necrotic cactus hosts. As well as differences in nutritional compounds, several compounds present in cactus necroses can be toxic. Previous studies of transcriptional changes induced in response to necrotic cactus compounds revealed that many detoxification genes were affected (Matzkin et al. 2006; Matzkin 2012). In the present study we also observed expression differences in detoxification genes such as Glutathione S-transferases, Cytochrome P450 and UDP-Glycosyltransferases. Unlike in the prior studies, however, these differences are fixed and do not involve cactus compounds for their induction. Of course it is feasible that some of the significant gene expression differences across the subspecies could have been a result of a subspecies-specific response to the banana media. We reason that given the composition of the banana media (banana, yeast, molasses, corn syrup, antifungal and agar), many detoxification genes would not be influenced. Given that all the differences observed in the present study were seen in the absence of cactus compounds, many of the metabolism and detoxification genes underlying local adaptation appear to have been canalized with respect to their transcriptional profile.

TRANSCRIPTIONAL DIVERGENCE AND SPECIATION

The number of transcriptional differences differs drastically depending upon the particular subspecies analyzed (Table 4). Currently, several molecular studies (Machado et al. 2007;

Matzkin 2008; Smith et al. 2012) support the earlier idea by Ruiz *et al.* (1990) that the center of genetic diversity for *D. mojavensis* resides in Baja California. *Drosophila mojavensis* colonized mainland Sonora, Mojave and Catalina Island following its allopatric divergence from its sister species, *D. arizonae*, in Baja California. This historical model of the origin of the *D. mojavensis* subspecies predicts that, in the absence of selection, the fewest number of fixed transcriptional differences should involve comparisons involving Baja California and the other three subspecies, while the largest difference should occur between non-Baja California comparisons. Our observed pair-wise transcriptional differences (Table 4) support this prediction. Interestingly we observed that comparisons using the mainland Sonora subspecies exhibited the largest number of transcriptional differences. One major difference between the Sonora subspecies and the other three is the fact that it is the only one that is sympatric with its sister species *D. arizonae* (Fellows and Heed 1972; Ruiz, Heed, and Wasserman 1990). Reinforcement has markedly shaped the behavior of the sympatric Sonoran subspecies (Markow 1981) and this in turn could have potentially shaped its transcriptome. Given the expected multifaceted nature of reinforcement, it would be expected that several genomic regions would be affected, which would include transcriptional differences.

Chromosomal inversions have been proposed to play a role in speciation (Noor et al. 2001; Rieseberg 2001; Machado, Haselkorn, and Noor 2007). Inversions facilitate the persistence of linkage blocks that contain sterility factors and as a result higher levels of divergence should be expected around the inverted regions. We observed a greater number of expression differences in chromosomes 2 and 3 and lower number in chromosomes 1, 4 and 5 than expected by chance (Table 5). In *D. mojavensis* only chromosomes 2 and 3 are polymorphic and several of these inversions are fixed between the subspecies (Wasserman 1962; Mettler 1963; Johnson 1980; Ruiz, Heed, and Wasserman 1990). According to the model, elevated levels of fixed sequence differences should be observed around the inversion breakpoints, it is probable that these sequence differences could also affect transcriptional levels as it was observed here. Although we have not yet mapped the breakpoint sequences of the inversions in *D. mojavensis*, the increased level of transcriptional divergence in chromosomes 2 and 3 as a whole, provides us with some initial tantalizing evidence on the role of inversions in local adaptation and isolation in these subspecies. Current genome sequencing of the three remaining *D. mojavensis* subspecies by Matzkin will provide the information needed to begin to answer this question.

Although the largest number of transcriptional differences observed were between subspecies, approximately 9% the transcriptome exhibited significant within subspecies variation of which roughly half also differed between subspecies (Fig. 2). A disproportionate large number were involved with chemosensory pathways mostly dealing with odorant behavior (Fig. 4). Thirty Odorant Receptors had significant within subspecies variation, 11 of which also had significant between subspecies differences. Significant within subspecies variation could possibly be a result of relaxation of selection or balancing selection. This would suggest that at least directional selection has not played a role in shaping the pattern transcriptional variation in these chemosensory genes. Alternatively, the significant within subspecies variation could be result of the lack of stimuli (i.e., cactus derived compound) present in the rearing medium (banana-molasses food). Several chemosensory genes have been previously shown to respond to cactus rearing (Matzkin et al. 2006; Matzkin 2012). Given variation in host chemical composition, it is possible that for certain genes it would be advantageous to have a more plastic transcriptional profile. Additional studies comparing

cactus vs. non-cactus rearing are necessary to determine roles, if any, of chemosensory genes in the local adaptation of the different host subspecies.

CONCLUSION

Adaptation to local ecological conditions can be a potent driver of divergence between isolated populations (Funk, Nosil, and Etges 2006; Nosil 2007; Schluter and Conte 2009; Nosil 2012). *Drosophila mojavensis* offers a powerful system to investigate the role of local adaptation in the process of speciation. These subspecies already have accumulated a series of behavioral, morphological, life history and molecular differences (Etges and Heed 1987; Markow 1991; Matzkin 2005; Matzkin et al. 2006; Machado et al. 2007; Etges et al. 2009; Smith et al. 2012). Here we presented evidence that local adaptation might have been responsible in shaping the transcriptome of these ecologically different cactus host subspecies. Studies aiming to link genome level differences with the observed transcriptional differences as well as to functional, physiological and behavior differences are ongoing in our laboratories.

ACKNOWLEDGMENTS

We thank Georgina J. Matzkin (a NSF Research Experience for Teachers intern) and Jasmin Uribe for their assistance in this project. We also would like to thank the UCSD *Drosophila* Stock Center. This project was supported by a grant from the Science Foundation Arizona to LMM and TAM and a NSF award (DEB-1219387) to LMM.

REFERENCES

- Bolstad, B. M., R. A. Irizarry, M. Astrand, and T. P. Speed. 2003. A comparison of normalization methods for high density oligonucleotide array data based on variance and bias. *Bioinformatics* 19:185-193.
- Bono, J. M., L. M. Matzkin, E. S. Kelleher, and T. A. Markow. 2011. Postmating transcriptional changes in reproductive tracts of con- and heterospecifically mated *Drosophila mojavensis* females. *Proceedings of the National Academy of Sciences* 108:7878-7883.
- Coyne, J. A., and H. A. Orr. 2004. *Speciation*. Sinauer Associates, Inc., Sunderland.
- Dobzhansky, T. 1937. *Genetics and the Origin of Species*. Columbia University, New York.
- Drosophila* 12 Genomes Consortium. 2007. Evolution of genes and genomes on the *Drosophila* phylogeny. *Nature* 450:203-218.
- Eden, E., D. Lipson, S. Yogev, and Z. Yakhini. 2007. Discovering Motifs in Ranked Lists of DNA Sequences. *PLoS Computational Biology* 3:e39.
- Eden, E., R. Navon, I. Steinfeld, D. Lipson, and Z. Yakhini. 2009. GOrilla: a tool for discovery and visualization of enriched GO terms in ranked gene lists. *BMC Bioinformatics* 10:48.
- Etges, W. J. 1990. Direction of life history evolution in *Drosophila mojavensis*. Pp. 37-56 in J. S. F. Barker, W. T. Starmer, and R. J. MacIntyre, eds. *Ecological and evolutionary genetics of Drosophila*. Plenum Press, New York.

- Etges, W. J., C. C. De Oliveira, M. G. Ritchie, and M. A. F. Noor. 2009. Genetics of Incipient Speciation in *Drosophila mojavensis*: II. Host Plants and Mating Status Influence Cuticular Hydrocarbon QTL Expression and GxE Interactions. *Evolution* 63:1712-1730.
- Etges, W. J., and W. B. Heed. 1987. Sensitivity to larval density in populations of *Drosophila mojavensis*: Influences of host plant variation on components fitness. *Oecologia* 71:375-381.
- Fellows, D. F., and W. B. Heed. 1972. Factors affecting host plant selection in desert-adapted cactiphilic *Drosophila*. *Ecology* 53:850-858.
- Fogleman, J. C., and J. R. Abril. 1990. Ecological and evolutionary importance of host plant chemistry. Pp. 121-143 in J. S. F. Barker, W. T. Starmer, and R. J. MacIntyre, eds. *Ecological and evolutionary genetics of Drosophila*. Plenum Press, New York.
- Fogleman, J. C., and W. T. Starmer. 1985. Analysis of the community structure of yeasts associated with the decaying stems of cactus. III. *Stenocereus thurberi*. *Microbial Ecology* 11:165-173.
- Funk, D., P. Nosil, and W. Etges. 2006. Ecological divergence exhibits consistently positive associations with reproductive isolation across disparate taxa. *Proceedings of the National Academy of Science* 103:3209-3213.
- Gordon, K. L., and I. Ruvinsky. 2012. Tempo and Mode in Evolution of Transcriptional Regulation. *Plos Genetics* 8:e1002432.
- Heed, W. B. 1982. The origin of *Drosophila* in the sonoran desert. Pp. 65-80 in J. S. F. Barker, and W. T. Starmer, eds. *Ecological genetics and evolution: The cactus-yeast-Drosophila model system*. Academic Press, New York.
- Heed, W. B. 1978. Ecology and genetics of Sonoran desert *Drosophila*. Pp. 109-126 in P. F. Brussard, ed. *Ecological genetics: The interface*. Springer-Verlag.
- Irizarry, R. A., B. Hobbs, F. Collin, Y. D. Beazer-Barclay, K. J. Antonellis, U. Scherf, and T. P. Speed. 2003. Exploration, normalization, and summaries of high density oligonucleotide array probe level data. *Biostat* 4:249-264.
- Johnson, W. R. 1980. *Chromosomal polymorphism in natural populations of the desert adapted species, Drosophila mojavensis*. Pp. 75. Department of Ecology and Evolutionary Biology. University of Arizona, Tucson, AZ.
- Kircher, H. W. 1982. Chemical composition of cacti and its relationship to Sonoran Desert *Drosophila*. Pp. 143-158 in J. S. F. Barker, and W. T. Starmer, eds. *Ecological genetics and evolution: The cactus-yeast-Drosophila model system*. Academic Press, New York.
- Krebs, R. A., and T. A. Markow. 1989. Courtship behavior and control of reproductive isolation in *Drosophila mojavensis*. *Evolution* 43:908-913.
- Lynch, M., D. G. Scofield, and X. Hong. 2005. The Evolution of Transcription-Initiation Sites. *Molecular Biology and Evolution* 22:1137-1146.
- Machado, C. A., T. S. Haselkorn, and M. A. F. Noor. 2007. Evaluation of the genomic extent of effects of fixed inversion differences on intraspecific variation and interspecific gene flow in *Drosophila pseudoobscura* and *D. persimilis*. *Genetics* 175:1289-1306.
- Machado, C. A., L. M. Matzkin, L. K. Reed, and T. A. Markow. 2007. Multilocus nuclear sequences reveal intra- and interspecific relationships among chromosomally polymorphic species of cactophilic *Drosophila*. *Molecular Ecology* 16:3009-3024.
- Markow, T. A. 1981. Courtship behavior and control of reproductive isolation between *Drosophila mojavensis* and *Drosophila arizonensis*. *Evolution* 35:1022-1026.

- Markow, T. A. 1991. Sexual isolation among populations of *Drosophila mojavensis*. *Evolution* 45:1525-1529.
- Matzkin, L. M. 2008. The Molecular Basis of Host Adaptation in Cactophilic *Drosophila*: Molecular Evolution of a Glutathione S-Transferase Gene (GstD1) in *Drosophila mojavensis*. *Genetics* 178:1073-1083.
- Matzkin, L. M. 2004. Population genetics and geographic variation of alcohol dehydrogenase (Adh) paralogs and glucose-6-phosphate dehydrogenase (G6pd) in *Drosophila mojavensis*. *Molecular Biology and Evolution* 21:276-285.
- Matzkin, L. M. 2005. Activity variation in alcohol dehydrogenase paralogs is associated with adaptation to cactus host use in cactophilic *Drosophila*. *Molecular Ecology* 14:2223-2231.
- Matzkin, L. M. 2012. Population transcriptomics of cactus host shifts in *Drosophila mojavensis*. *Molecular Ecology* 21:2428-2439.
- Matzkin, L. M., T. D. Watts, B. G. Bitler, C. A. Machado, and T. A. Markow. 2006. Functional genomics of cactus host shifts in *Drosophila mojavensis*. *Molecular Ecology* 15:4635-4643.
- Mettler, L. E. 1963. *Drosophila mojavensis* baja, a new form in the mulleri complex. *Drosophila Information Service* 38:57-58.
- Muller, H. J. 1942. Isolating mechanisms, evolution, and temperature. *Biological Symposia* 6:71-125.
- Noor, M. A. F., K. L. Grams, L. A. Bertucci, and J. Reiland. 2001. Chromosomal inversions and the reproductive isolation of species. *Proceedings of the National Academy of Sciences* 98:12084-12088.
- Nosil, P. 2012. *Ecological Speciation*. Oxford University Press, Oxford.
- Nosil, P. 2007. Divergent Host Plant Adaptation and Reproductive Isolation between Ecotypes of *Timema cristinae* Walking Sticks. *The American Naturalist* 169:151-162.
- Pfeiler, E., S. Castrezana, and L. K. Reed. 2009. Genetic, ecological and morphological differences among populations of the cactophilic *Drosophila mojavensis* from southwestern USA and northwestern Mexico, with descriptions of two new subspecies. *Journal of Natural History* 43:923-938.
- Rieseberg, L. H. 2001. Chromosomal rearrangements and speciation. *Trends in Ecology and Evolution* 16:351-358.
- Ruiz, A., W. B. Heed, and M. Wasserman. 1990. Evolution of the Mojavensis cluster of cactophilic *Drosophila* with descriptions of two new species. *Journal of Heredity* 81:30-42.
- Schluter, D., and G. L. Conte. 2009. Genetics and ecological speciation. *Proceedings of the National Academy of Sciences* 106:9955-9962.
- Smith, G., K. Lohse, W. J. Etges, and M. G. Ritchie. 2012. Model-based comparisons of phylogeographic scenarios resolve the intraspecific divergence of cactophilic *Drosophila mojavensis*. *Molecular Ecology* 21:3293-3307.
- Starmer, W. T. 1982a. Analysis of the Community Structure of Yeasts Associated with the Decaying Stems of Cactus. I. *Stenocereus gummosus*. *Microbial Ecology* 8:71-81.
- Starmer, W. T. 1982b. Associations and Interactions Among Yeasts, *Drosophila* and Their Habitats. Pp. 159-174 in J. S. F. Barker, and W. T. Starmer, eds. *Ecological genetics and evolution: The cactus-yeast-Drosophila model system*. Academic Press, New York.
- Starmer, W. T., M. A. Lachance, H. J. Phaff, and W. B. Heed. 1990. The biogeography of yeasts Associated with decaying cactus tissue in North America, the Caribbean, and

- Northern Venezuela. Pp. 253-296 in M. K. Hecht, B. Wallace, and R. J. Macintyre, eds. *Evolutionary Biology*. Plenum Publishing Corporation.
- Stone, J., and G. Wray. 2001. Rapid evolution of *cis*-regulatory sequences via local point mutations. *Molecular Biology and Evolution* 18:1764-1770.
- Storey, J., and R. Tibshirani. 2003. Statistical significance for genomewide studies. *Proceedings of the National Academy of Sciences, USA* 100:9440-9445.
- Vacek, D. C. 1979. The microbial ecology of the host plants of *Drosophila mojavensis*. Pp. 112. *Department of Ecology and Evolutionary Biology*. University of Arizona, Tucson, AZ.
- Wasserman, M. 1962. Cytological studies of the repleta group of the genus *Drosophila*. V. The mulleri subgroup. *University of Texas Publication* 6205:85-118.
- Wittkopp, P. J., B. K. Haerum, and A. G. Clark. 2004. Evolutionary changes in *cis* and *trans* gene regulation. *Nature* 430:85-88.
- Wolfinger, R. D., G. Gibson, E. D. Wolfinger, L. Bennett, H. Hamadeh, P. Bushel, C. Afshari, and R. S. Paules. 2001. Assessing gene significance from cDNA microarray expression data via mixed models. *Journal of Computational Biology* 8:625-637.
- Wray, G. A., M. W. Hahn, E. Abouheif, J. P. Balhoff, M. Pizer, M. V. Rockman, and L. A. Romano. 2003. The Evolution of Transcriptional Regulation in Eukaryotes. *Molecular Biology and Evolution* 20:1377-1419.

Chapter 6

**EVOLUTION OF MALE COURTSHIP SONGS
IN THE *DROSOPHILA BUZZATII*
SPECIES CLUSTER**

**Cássia C. Oliveira^{1,2,*}, Maura H. Manfrin³,
Fábio de M. Sene⁴ and William J. Etges¹**

¹Program in Ecology and Evolutionary Biology,
Department of Biological Sciences, University of Arkansas,
Fayetteville, AR, US

²Math and Science Division, Lyon College, Batesville, AR, US

³Departamento de Biologia, Faculdade de Filosofia, Ciências e Letras
de Ribeirão Preto, Universidade de São Paulo, Ribeirão Preto-SP, Brazil

⁴Departamento de Genética, Faculdade de Medicina de Ribeirão Preto,
Universidade de São Paulo, Ribeirão Preto-SP, Brazil

ABSTRACT

Acoustic signals produced to attract mates before, during, and after courtship are frequently involved with sexual selection, sexual isolation, and reproductive isolation in *Drosophila* spp. and other animals, yet few studies have revealed how courtship songs evolve in a larger phylogenetic context. Therefore, we mapped different acoustic components of courtship songs in the monophyletic *Drosophila buzzatii* species cluster onto an independently derived *period (per)* gene + chromosome inversion phylogeny to assess the concordance of courtship song evolution with species divergence. These cactophilic flies are distributed throughout several biomes in southern South America and include the sibling species *D. buzzatii*, *D. koepferae*, *D. serido*, *D. borborema*, *D. seriema*, *D. antonietae*, and *D. gouveai*. All seven species produced two song types; primary and secondary pulse songs, except for *D. borborema* and *D. gouveai* that produced no secondary songs. Courtship songs were characterized by analyzing six commonly studied acoustic components including burst duration (BD), carrier frequency (CF), pulse length (PL), pulse number (PN), inter-burst interval (IBI), and inter-pulse interval (IPI).

* Corresponding Author's Email: cassia.oliveira@lyon.edu.

Significant intra- and inter-specific song variation was observed for BD, PN, and IBI, while CF, PL, and IPI varied in a more species-specific manner, albeit with some overlap. Thus, some song components may be better species recognition signals than others. Multivariate clustering analyses resolved all species into distinct, non-overlapping groups. Mapping individual song traits (BD, IBI, and IPI) as well composites of these song variables onto our (*per*) gene + chromosome inversion phylogeny revealed no phylogenetic signal when different comparative mapping methods were used. Hence, the evolution of courtship songs in *D. buzzatii* cluster species was uncorrelated with the degree of species divergence. These findings reinforce previous observations that courtship songs evolve rapidly enough to erase any signature of evolutionary affinity between closely related animal species.

INTRODUCTION

In a number of *Drosophila* species, courtship “love” songs have been implicated in promoting sexual isolation between species (Ewing 1989; Ritchie et al. 1999). Despite the potential importance of courtship songs in the speciation process and that songs have been characterized in over 100 *Drosophila* species (Hoikkala 2005), only a few studies have investigated the correlation between song traits and species phylogenetic history (Ewing and Miyan 1986; Gleason and Ritchie 1998; Etges 2002). Comparative studies involving mate signaling cues in closely related species are crucial to unraveling not only which traits are repeatedly involved in the early stages of species formation, but also determining their divergence rates across taxa (Etges 2002; Coyne and Orr 2004). In short, we are interested in identifying key behavioral traits that are responsible for large-scale diversification of species. Thus, we analyzed the evolution of quantitative differences in courtship song traits in the *D. buzzatii* cluster, a group of recently diverged species, in order to assess the concordance of love song evolution in relation to patterns of species divergence in a phylogenetic context.

During courtship, males of most *Drosophila* species produce acoustic signals, courtship love songs, by vibrating their wings in attempts to gain female acceptance and successful copulation (Ewing 1983). Courtship songs are typically species-specific in the majority of *Drosophila* species (Cowling and Burnet 1981; Cobb et al. 1988; Ritchie and Gleason 1995), and so acoustic signaling is thought to allow courting adults to ascertain the appropriateness of attempting to mate with a member of the opposite sex (Ewing 1989; Saarikettu et al. 2005; Mendelson and Shaw 2012), even though these species differences may evolve by sexual selection (Ritchie and Gleason 1995; Ritchie et al. 1998).

Drosophila love songs are typically characterized by low frequency pulses that can be produced individually or in structured bursts. Some species, such as those in the *D. melanogaster* group, have two types of song, pulse song and sine song (Cowling and Burnet 1981). In the *D. repleta* group, two kinds (A and B) of pulse songs have been described where A songs have short inter-pulse intervals (S-IPs), and B songs have longer inter-pulse intervals (L-IPs) (Figure 1). However, not all species within the group exhibit both song types, and variation between species is considerable (Ewing and Miyan 1986).

The rate of pulse production measured by the inter-pulse interval or IPI has been shown to be a common mate recognition signal recognized by female *Drosophila*. However, other courtship song traits have been found to be species-specific in *Drosophila*, including burst duration, inter-burst interval, pulse number per burst, sine song, cycle number per pulse, and intra-pulse frequency (Bennet-Clark and Ewing 1969; Kyriacou and Hall 1980; Ritchie and Gleason 1995; Byrne 1999; Yamada et al. 2002; Etges et al. 2006). Females can use one or more love song components when selecting among potential mates (Kyriacou and Hall 1982; Tomaru et al. 1995; Ritchie, et al. 1998). For example, *Drosophila montana* females will not mate with wingless males, implying that courtship song is an obligatory component of courtship in this species (Liimatainen et al. 1992). In *D. melanogaster*, male courtship song is not necessary for mate recognition, since wingless males can copulate, even though the time it takes to achieve copulation is longer than for control males (von Schilcher 1976). Using recorded love songs in playback experiments has shown a role for courtship song in sexual isolation between different populations or species by exposing females to wingless males and then playing different types of songs. The large role of love songs has been confirmed in these studies where male mating success was restored or increased by play-back of songs of the same population or species. For example, this is the case for two other members of the *D. repleta* group, *D. mojavensis* and *D. arizonae* (Byrne 1999).

PHYLOGENY OF *D. BUZZATII* CLUSTER SPECIES

The species of the *D. buzzatii* cluster belong to the large *D. repleta* group (Ruiz and Wasserman 1993; Durando et al. 2000; Oliveira et al. 2012) including *D. buzzatii*, *D. koepferae*, *D. serido*, *D. borborema*, *D. seriema*, *D. antonietae*, and *D. gouveai*. All are closely related “sibling” species that form a monophyletic group (Manfrin and Sene 2006). These species are endemic to South America, except for *D. buzzatii* that has a cosmopolitan distribution due to its association with species of *Opuntia* cactus that have been propagated around the world for fruit production (Wasserman 1992).

The monophyly of the *D. buzzatii* cluster was first defined on the basis of a complex arrangement of chromosomal inversions (Ruiz and Wasserman 1993), yet only four fixed inversions can be used for species identification (Ruiz et al. 2000; Manfrin et al. 2001) (Figure 2). Other traits including mtDNA COI gene sequences (Manfrin, et al. 2001) and Xanthine dehydrogenase (*Xdh*) nucleotide sequences (Rodriguez-Trelles et al. 2000), as well as wing morphology (Moraes et al. 2004) have also been used to infer phylogenetic relationships among these species. Phylogenetic analysis using mtDNA COI indicated that the *D. buzzatii* cluster was a well-supported monophyletic group (Manfrin, et al. 2001; de Brito et al. 2002), but these mtDNA sequences did not help to resolve the pattern of species relationships within this group. At present, sequence variation in the *period* (*per*) gene (Franco et al. 2010) has produced a phylogeny that best resolves these branching patterns. The *per* phylogeny also reinforced the monophyletic nature of this cluster, but more importantly it resolved the relationships of *D. gouveai*, *D. borborema* and *D. seriema*, which had been difficult to understand when chromosomal inversions and mtDNA sequences were used.

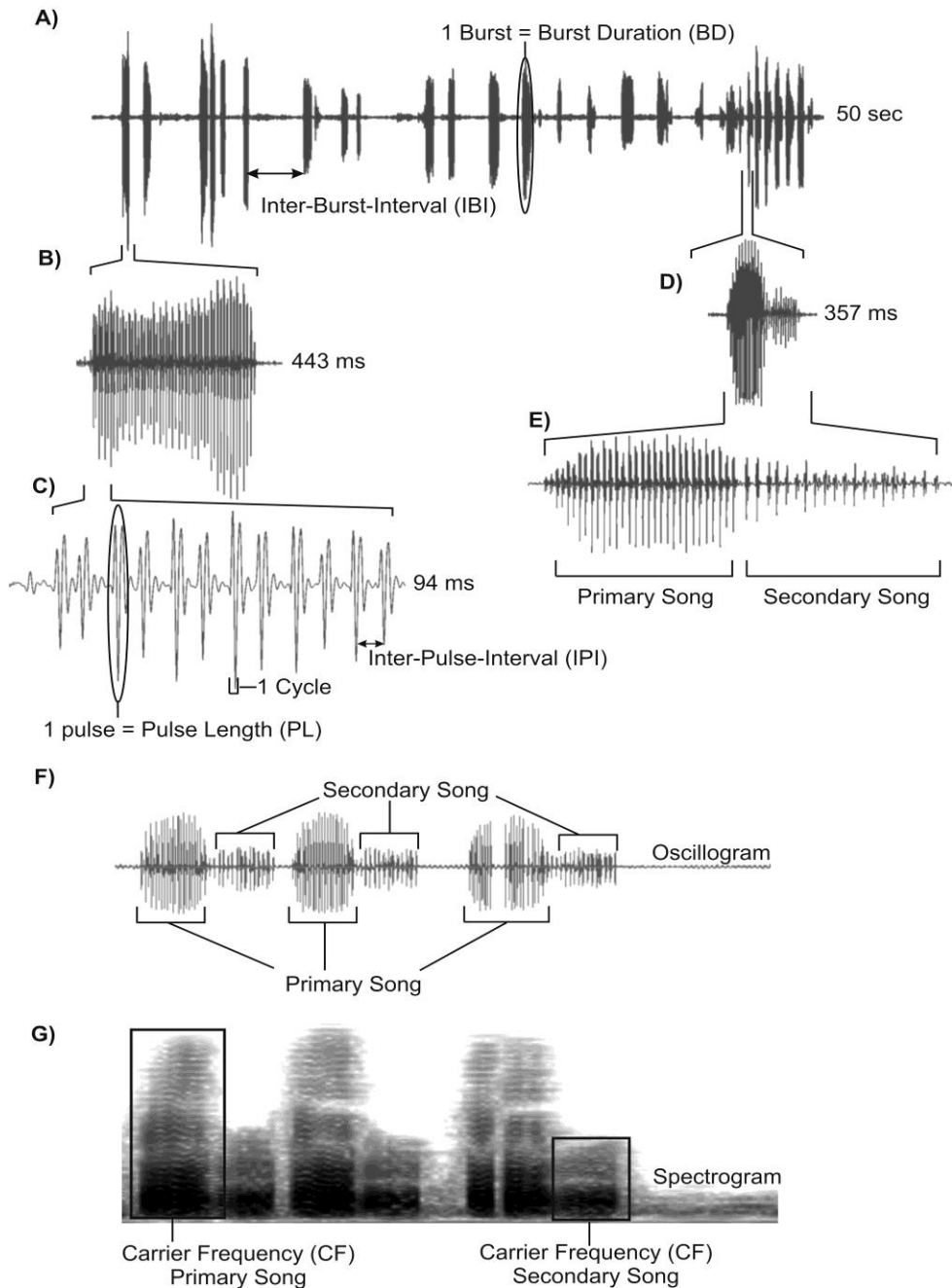


Figure 1. A-G. Typical courtship song of *D. buzzatii* species composed of pulses arranged into bursts. Oscillograms are used to illustrate the *Drosophila* courtship song terminology. (A) Fifty seconds of pulse song showing both primary and secondary song. IBI = inter-burst interval. (B) Single burst of primary song showing composed of 52 pulses. (C) Expanded view of B showing the first 12 polycyclic pulses. IPI = inter-pulse interval. (D) Two bursts showing primary and secondary song, respectively. (E) Enlarged view of D. (F) Oscillogram of six bursts: three bursts of primary song intercalated by three bursts of secondary song. (G) Spectrogram of the oscillogram showed in F. Bursts of primary song present higher frequency than the bursts of secondary song.

HOST CACTUS, BIOGEOGRAPHY AND ECOLOGY OF *D. BUZZATII* CLUSTER SPECIES

The *D. buzzatii* cluster species are distributed over a vast geographical area in South America, ranging from northeastern Brazil to Paraguay, Bolivia and Argentina (Figure 3). The vegetation in these areas includes the morpho- climatic biomes of caatinga (thorny scrub), cerrado (savannah), Atlantic forest, and Chaco. Like other cactophilic species belonging to the *mulleri* subgroup of the *D. repleta* group, *D. buzzatii* cluster species use fermenting cactus tissues as feeding and breeding substrates, and the level of host specificity varies among the different species (Sene et al. 1982; Ruiz, et al. 2000; Kokudai et al. 2011).

The ecology and biogeography of the *D. buzzatii* cluster species, as well as varying levels of genetic divergence within this clade make the *D. buzzatii* cluster an ideal system to address questions regarding how mate recognition systems evolve in the early stages of species divergence. First, we recorded and described the courtship songs of these species and used a comparative approach to assess whether courtship song evolution was correlated with species divergence. Our results revealed strong species-specific differentiation in multiple acoustic characteristics of male courtship songs signifying rapid evolution in this central component of acoustic courtship signaling.

DESCRIPTION OF COURTSHIP SONGS AND COMPARATIVE METHODS

We recorded the courtship songs of all seven species of the *D. buzzatii* cluster and quantified variation in courtship song components (Figure 3). Fly stocks and handling procedures are described in Oliveira et al. (2011). All flies were aged at least 8 days before use to ensure sexual maturity (Bizzo 1983; Moraes 1992). Courtship songs of ten males of each species were recorded with an ultra-sensitive microphone (Bennet-Clark 1984) in an acrylic chamber (3 x 3 x 1 cm³) as described by Sene and Manfrin (1998). Each male was housed with two virgin females of the same species and the wings were removed from the females prior to recording. Temperature inside the recording chamber was monitored continuously with a digital thermometer because courtship songs can be temperature dependent (Byrne 1999; Ritchie et al. 2001). The time of day of recording was not controlled for. Approximately three minutes of song were recorded for each male. We digitized the song recordings at 8 KHz using Sonic Sound Forge software (2006, Creative Software Inc., Madison, Wisconsin, USA).

We analyzed courtship song components with Raven Pro 1.3 software (2003, Cornell Laboratory of Ornithology, Ithaca, New York, USA). All song measurements were made directly from the waveform tracings from Raven. For each male, five bursts of each type of song (i.e., primary and secondary) were analyzed but not all males produced a secondary song. A total of six song components was analyzed for each type of song including burst duration (BD), carrier frequency (CF), pulse number (PN), and inter-burst interval (IBI) that were measured from five randomly selected bursts. For pulse length (PL) and inter-pulse interval (IPI) five randomly selected pulses or inter-pulses, respectively, were measured per burst (Table 1, Figure 1).

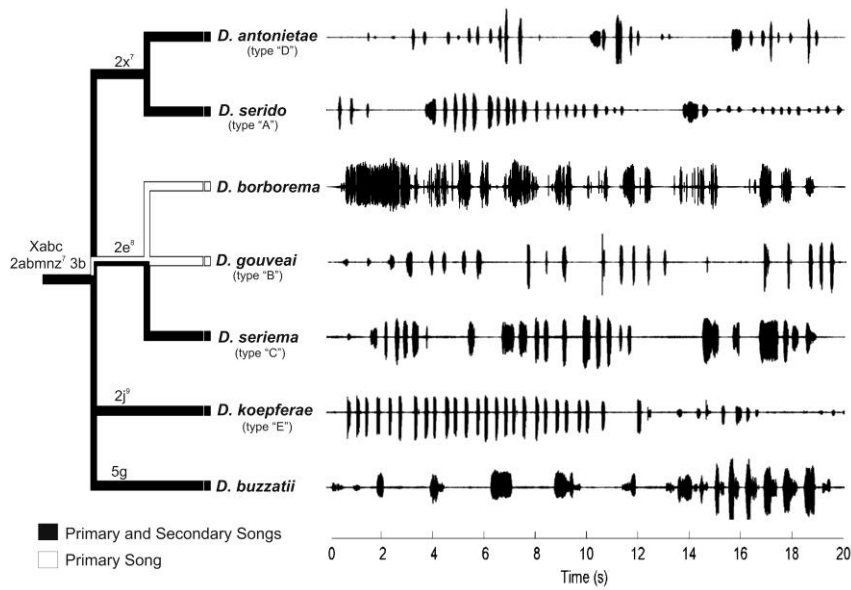


Figure 2. Left side: Consensus phylogeny based on chromosomal inversions for *D. buzzatii* species cluster. Male genitalia (type A – E) are according to Silva and Sene (1991). Chromosomal inversions, shown above the tree branches, were based on the work of Ruiz et al. (1997; 2000). Black branches characterize species that have both primary and secondary song, while white branches represent species that possess only primary song, i.e., have lost secondary song. Right side: Typical wave pattern of the male courtship songs of the species of the *D. buzzatii* cluster.

Song differences among species were assessed for all song variables using PROC GLM (SAS Institute Inc. 2004) and temperature effects during recording were evaluated with analysis of covariance (ANCOVA). Species were considered a fixed effect and temperature was log transformed to improve normality. Principal components analysis (PCA) was used to reduce the dimensionality of the data in PROC PRINCOMP, and canonical discriminant function analysis (CDF) was used to help visualize species differences with PROC CANDISC. We used both data corrected for temperature variation and the residuals in PCA and CDF analysis. Principal Components (PCs) and canonical variates (CVs) were later used for character mapping analysis (see below).

Table 1. Definition of song parameters analyzed in the species of the *D. buzzatii* cluster

Song Parameter	Abbreviation	Unit	Definition
Burst Duration	BD	Milliseconds (ms)	The time between the first and last pulse in a burst.
Carrier frequency	CF	Hertz (Hz)	Highest peak frequency from a fast Fourier transformation.
Pulse Length	PL	Milliseconds (ms)	The length of a pulse.
Pulse Number	PN	----	Number of pulses per burst.
Inter-Burst Interval	IBI	Milliseconds (ms)	The time between the end of a burst and the beginning to the next one.
Inter-Pulse Interval	IPI	Milliseconds (ms)	The time between pulses, measured from peak-to-peak.

For each male five bursts were analyzed for each type of song (primary and secondary song).

Phylogenetic Reconstruction

Phylogenetic relationships for the seven *D. buzzatii* cluster species were reconstructed using chromosomal inversion differences (Ruiz et al. 1997; Ruiz, et al. 2000) and nucleotide variation in a 443 bp fragment of the X linked *period* (*per*) gene (Franco, et al. 2010). Two outgroup species, *D. mojavensis* and *D. hydei* were also included. Because no song data were available for *D. hydei* this species was removed before the tree was used for phylogenetic analysis of song evolution. Three song components were available for *D. mojavensis*, i.e., BD, IBI, and IPI, from Etges et al. (2007). Because we were also interested in the song evolution for *D. mojavensis* as well as its effects as an outgroup, we kept this species in the character reconstruction analysis when individual song components were mapped onto the phylogeny, but removed it when PCs or CVs were used (see below).

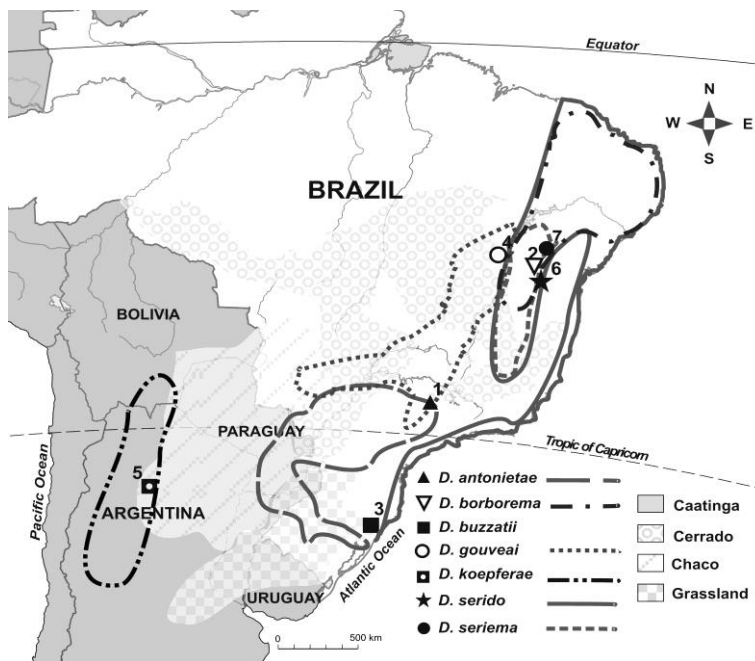


Figure 3. Partial view of South American map showing the geographic distribution of the species of the *D. buzzatii* cluster and the four major vegetation types with which these species are associated. The distribution of *D. buzzatii* is not marked because this species is found in all areas where the other species occur. Numbers represent the description for each of the species used in the courtship song analysis, i.e., species name, stock number, locality and year of collection. (1) *D. antonietae* (J41P1M), Serrana, São Paulo, 1999; (2) *D. borborema* (N70), Junco do Seridó, Paraíba, 2008; (3) *D. buzzatii* (J26A45), Osório, Rio Grande do Sul, 1998; (4) *D. gouveai* (J78M1), Ibotirama, Bahia, 2001; (5) *D. koepferae* (B20D2), Tapia, Tucumán; (6) *D. serido* (J92A91M), Milagres, Bahia, 2002; (7) *D. seriema* (D73C5BM), Morro do Chapéu, Bahia, 1990. Except for *D. koepferae*, from Argentina, all other species were collected in Brazil.

Phylogenetic analysis using the combined data was performed using PAUP* 4.0 (Swofford 2000) as in Oliveira et al. (2011). Maximum parsimony was used to search for optimal tree(s) and heuristic searches were carried out with 100 random addition analyses and tree bisection reconnection (TBR) branch swapping. Nodal support was obtained using bootstrap analysis (1,000 replicates).

Mapping Song Traits onto the Phylogeny

Patterns of courtship song evolution were inferred by mapping individual song traits, i.e., BD, IBI, and IPI, Principal Components (PCs), and canonical variates (CVs) onto the reconstructed phylogeny using Mesquite 2.74 (Maddison and Maddison 2010). Because some song components were temperature dependent (e.g., Ritchie, et al. 2001; Etges, et al. 2007), all six song components were regressed against temperature using PROC REG to generate predicted (PRD) and residual (RES) values used in character mapping. These values were mapped onto the first of two most parsimonious trees instead of the strict consensus tree because one of the models used, Squared Change Parsimony Gradual (see below), requires branch length information. Character reconstruction analysis was used to infer phylogenetic signal and was performed using three parsimony methods, Linear Parsimony (LP), Squared Change Parsimony Gradual (SCPG), and Squared Change Parsimony Punctuated (SCPP). In addition to these three parsimony methods, we also used the test for serial independence (TFSI) to detect phylogenetic signal as described in Abouheif (1999) using Phylogenetic Independence 2.0 (Reeve and Abouheif 2003). Phylogenetic signal was used as a measure of congruence between the phylogeny and variation in the song variables.

We tested the null hypothesis that courtship songs have evolved independently of species evolution due to non-phylogenetic influences such as developmental noise, ecological effects (e.g., rearing conditions), or species-specific sexual selection. Our alternative hypothesis was that positive phylogenetic signal should be observed due to the phylogenetic affinities of these species and song traits. The presence of phylogenetic signal was tested with all three parsimony methods by randomly modifying the most parsimonious tree, named here as a reference tree (Oliveira, et al. 2011). The terminal taxa on the reference tree were reshuffled 10,000 times to generate a population of random trees for each of the variables tested, i.e., PCs, CVs, and individual variables (BD, IBI, and IPI). These random trees with reshuffled taxa were then compared with the reference tree to test whether the mapped variables were more conserved than expected by chance alone. Presence of phylogenetic signal was inferred if the number of parsimony character steps in the reference tree was less than in 95% of the trees with reshuffled taxa and fell on the extreme left of the distribution. For all three parsimony methods and TFSI, *P* values were corrected for multiple comparisons via false discovery rate (FDR) analysis (Benjamini and Hochberg 1995; Laurin et al. 2009).

Differences in Courtship Song Components

Male courtship songs consisted of low-frequency, polycyclic pulses arranged into pulse trains or bursts (Figures 1 and 2). Courtship songs were produced by vibration of both wings during courtship and until copulation, but no song was produced during or after copulation. Primary song was produced during most of the courtship sequence and secondary songs were usually produced later in courtship, immediately before copulation. Secondary song was absent in males of *D. borborema* and *D. gouveai*. Ambient recording temperature ($\bar{X} \pm \text{SD} = 25.33 \pm 1.01^\circ\text{C}$, $N = 70$, range $23 - 27^\circ\text{C}$) had little significant effect on variation in any of the song components except for primary song IPI (Table 2). ANCOVA revealed heterogeneity of slopes for a few song components caused by different species. Along with some missing data and only

10 males per species recorded, we observed differences in significance between Type I and Type III sums of squares for species differences (results not shown) and statistical significance for the overall model sums of squares. We report Type III sums of squares and their significance in Table 2 to be conservative, but Type I sums of squares for BD, CF, PN, IBI, and IPI were all statistically significant. Further, significant pair-wise species differences were observed when least square means were analyzed (See Figure 4). As the ANCOVAs used contained one fixed effect (species), temperature as a covariate, and a species X temperature interaction term, we concluded that Type I sums of squares were appropriate for revealing species differences in these song components. Differences among species as well as differences in type of song for each song component are described below. Since *D. borborema* and *D. gouveai* lacked secondary songs, no comparison for type of song was available for these species.

Pair-wise comparisons using least square means revealed that burst duration was variable for several of the species pairs (Figure 4A). Further- more, BD did not vary consistently for individuals of the same species or even in the same individual. In fact, song bursts in the same individual had different shapes, amplitudes and durations (Figure 2). Mean BD for primary song ranged from 129.34 ms to 660.44 ms, and for secondary song, ranged from 78.67 ms to 440.17 ms (Table 3). Except for *D. borborema*, CF was relatively similar among the other species (Figure 4B). For all species, CF was characterized by low frequency peaks with mean CF for primary song ranging from 213.13 Hz to 467.51 Hz, and for secondary song ranging from 274.67 Hz to 379.69 Hz (Table 3). Pulse length or pulse duration was more conserved for primary song than for secondary song (Figure 4C). All seven species produced songs with polycyclic pulses consisting of two to four cycles per pulse. Mean PL for primary song ranged from 5.25 ms to 6.14 ms, and for secondary song, from 6.22 ms to 7.75 ms (Table 3).

Pulse number influenced burst duration. For instance, *D. borborema* produced long bursts (Figure 4A), which had more pulses (Figure 4D). Mean PN for primary song ranged from 9.6 pulses to 42.3 pulses (Table 3). For secondary song, mean PN ranged from 5.8 pulses to 24.3 pulses (Table 4). Because of its correlation with burst duration, pulse number was also highly variable. Inter-burst interval, measured as the distance between bursts, was difficult to calculate because some males stopped and started singing multiple times. Least square mean comparisons revealed that *D. borborema* had the highest IBI values. Even though IBI for secondary song was not statistically different among species ($P = 0.2482$), there was variation among individuals of the same species, especially for *D. antonietae*, as indicated by a large standard error (Figure 4E). Mean IBI for primary song ranged from 347.56 ms to 1348.22 ms, and for secondary song, from 480.73 ms to 1536.84 ms (Table 3).

Based on least square mean comparisons for primary song, *D. borborema* had the highest mean IPI, which was significantly different from all other species (Figure 4F). For secondary song, *D. buzzatii*, *D. koepferae*, and, *D. seriema* had the highest mean IPI followed by *D. serido* and lastly by *D. antonietae*. Mean IPI for primary song ranged from 7.92 ms to 14.20 ms, and for secondary song, from 8.85 ms to 12.30 ms (Table 3). Differences in IPI between primary and secondary songs were significant among the five species that possessed both types of songs ($P = 0.0113$). Significant pair-wise species differences were observed when least square means were analyzed, but only *D. buzzatii* showed significant differences for both song types where IPIs were shorter for primary song and longer for secondary song (Figure 5). Therefore, except for *D. buzzatii*, the other four species had unimodal IPIs, i.e., just one type of IPI. Bimodal IPIs, i.e., short and long IPIs, are considered a characteristic of the *D. repleta* group (Ewing and Miyan 1986).

Table 2. Results of ANCOVAs for six song parameters analyzed for the species of the *D. buzzatii* cluster

Song Parameter	Source of Variation	df	Type III SS	F	P
Primary Song					
Burst Duration (BD)	Model	12	2227914.86	9.73	<0.0001
	Species	5	34867.21	0.37	0.8701
	Lgtemp	1	8456.86	0.44	0.5083
	Lgtemp x Species	5	36531.98	0.38	0.8585
	Error	57	1087763.35		
Carrier Frequency (CF)	Model	12	470873.61	10.70	<0.0001
	Species	5	10812.81	0.59	0.7077
	Lgtemp	1	5316.89	1.45	0.2335
	Lgtemp x Species	5	10980.11	0.60	0.7008
	Error	57	208968.10		
Pulse Number (PN)	Model	12	10326.37	10.11	<0.0001
	Species	5	292.94	0.69	0.6341
	Lgtemp	1	187.01	2.20	0.1437
	Lgtemp x Species	5	306.34	0.72	0.6110
	Error	57	4849.79		
Pulse Length (PL)	Model	12	8.92	1.28	0.2576
	Species	5	2.37	0.81	0.5452
	Lgtemp	1	0.05	0.08	0.7801
	Lgtemp x Species	5	2.36	0.81	0.5466
	Error	57	33.19		
Inter-Burst-Interval (IBI)	Model	12	12948801.41	2.43	0.0126
	Species	5	1582033.81	0.71	0.6164
	Lgtemp	1	43479.23	0.10	0.7555
	Lgtemp x Species	5	1585975.11	0.71	0.6151
	Error	57	25303658.70		
Inter-Pulse-Interval (IPI)	Model	12	295.59	41.76	<0.0001
	Species	5	4.00	1.36	0.2549
	Lgtemp	1	5.27	8.93	0.0041
	Lgtemp x Species	5	3.57	1.21	0.3157
	Error	57	33.62		
Burst Duration (BD)	Model	9	542983.86	1.72	0.1285
	Species	4	1083.37	0.01	0.9999
	Lgtemp	1	72.66	0.00	0.9640
	Lgtemp x Species	4	798.02	0.01	0.9999
	Error	30	1053962.14		
Carrier Frequency (CF)	Model	9	70692.77	1.43	0.2216
	Species	4	7335.99	0.33	0.8537
	Lgtemp	1	6278.32	1.14	0.2943
	Lgtemp x Species	4	7367.70	0.33	0.8527
	Error	30	165310.43		

Song Parameter	Source of Variation	df	Type III SS	F	P
Secondary Song					
Pulse Number (PN)	Model	9	1406.84	1.46	0.2095
	Species	4	18.75	0.04	0.9962
	Lgtemp	1	3.49	0.03	0.8581
	Lgtemp x Species	4	17.90	0.04	0.9965
	Error	30	3220.48		
Pulse Length (PL)	Model	9	28.33	0.73	0.6794
	Species	4	16.08	0.93	0.4593
	Lgtemp	1	4.74	1.10	0.3030
	Lgtemp x Species	4	15.95	0.92	0.4635
	Error	30	129.58		
Inter-Burst-Interval (IBI)	Model	9	24012175.07	1.70	0.1335
	Species	4	8993207.14	1.43	0.2482
	Lgtemp	1	151204.19	0.10	0.7586
	Lgtemp x Species	4	9094390.70	1.45	0.2432
	Error	30	47160005.47		
Inter-Pulse-Interval (IPI)	Model	9	66.66	2.08	0.0637
	Species	4	9.01	0.63	0.6425
	Lgtemp	1	2.25	0.63	0.4323
	Lgtemp x Species	4	8.84	0.62	0.6505
	Error	30	106.61		

See Figure 4 for least square mean differences between species. Lgtemp = $\log_{10}$ temperature.

Principal components analysis (PCA) revealed that the five principal components (PCs) accounted for 96% of the variation in the data for the seven species. The first principal component (PC1) accounted for 51% of the variance and was mainly driven by the differences between primary and secondary songs. PC1 scores were all negative for primary song traits (except for carrier frequency), and positive for secondary song traits (Table 4). Such differences were accentuated because secondary song was absent in some males and completely absent in *D. borborema* and *D. gouveai*. Accordingly, PC1 separated these two species from the others (Figure 6). The second principal component (PC2) accounted for 20% of the variation and separated species largely based on differences in primary songs traits, i.e., BD, CF, PN, and IBI. The third and fourth PCs, which represented 16% and 6% of the variation, respectively, were also mostly influenced by differences in primary songs (Table 4).

Canonical discriminant function (CDF) analysis using the residuals of the song characters yielded significant multivariate differences among species (Wilks $\lambda = 0.0000$, $F = 6.13 \times 10^{11}$, $P < 0.0001$). The first three canonical variables accounted for 98% of the total variation in courtship songs. As observed with PC analysis, the first canonical variate (CV1) was largely influenced by differences in type of song, primary and secondary, and the second and third canonical variates (CV2 and CV3) expressed differences among species as a result of variation in primary song (results not shown). Altogether, the results from PCA and CDF analysis confirmed that courtship songs were species-specific in the *D. buzzatii* cluster and primary song was mainly responsible for species differences.

CHARACTER MAPPING ANALYSIS OF COURTSHIP SONG

We used the first of two most parsimonious trees to perform the character reconstruction analysis (Figure 7). No phylogenetic signal was observed for any of the song traits mapped onto the phylogeny using either temperature corrected data or the residuals, i.e., individual song traits (BD, IBI, and IPI), CVs or PCs, using four different reconstruction methods (Table 5). Even though *D. mojavensis* is not closely related to the *D. buzzatii* cluster, this species had long bursts similar to *D. buzzatii*, *D. serido*, and *D. seriema*. However, *D. koepferae*, closely related to *D. buzzatii*, had short bursts (Figure 7A). Furthermore, *D. borborema* and *D. seriema* are closely related, but the former had the longest bursts of all species (Table 3, Figure 7). Similar differences were observed for IBI and IPI (Figure 7B, C) and the other variables (PCs and CVs). When PCs and CVs were mapped onto a phylogeny with or without *D. mojavensis* as an outgroup, phylogenetic signal was not detected (Table 5) indicating that this species did not influence the results. Overall, our results demonstrated no congruence between species differences in these song traits and phylogenetic structure in this clade of *Drosophila* species.

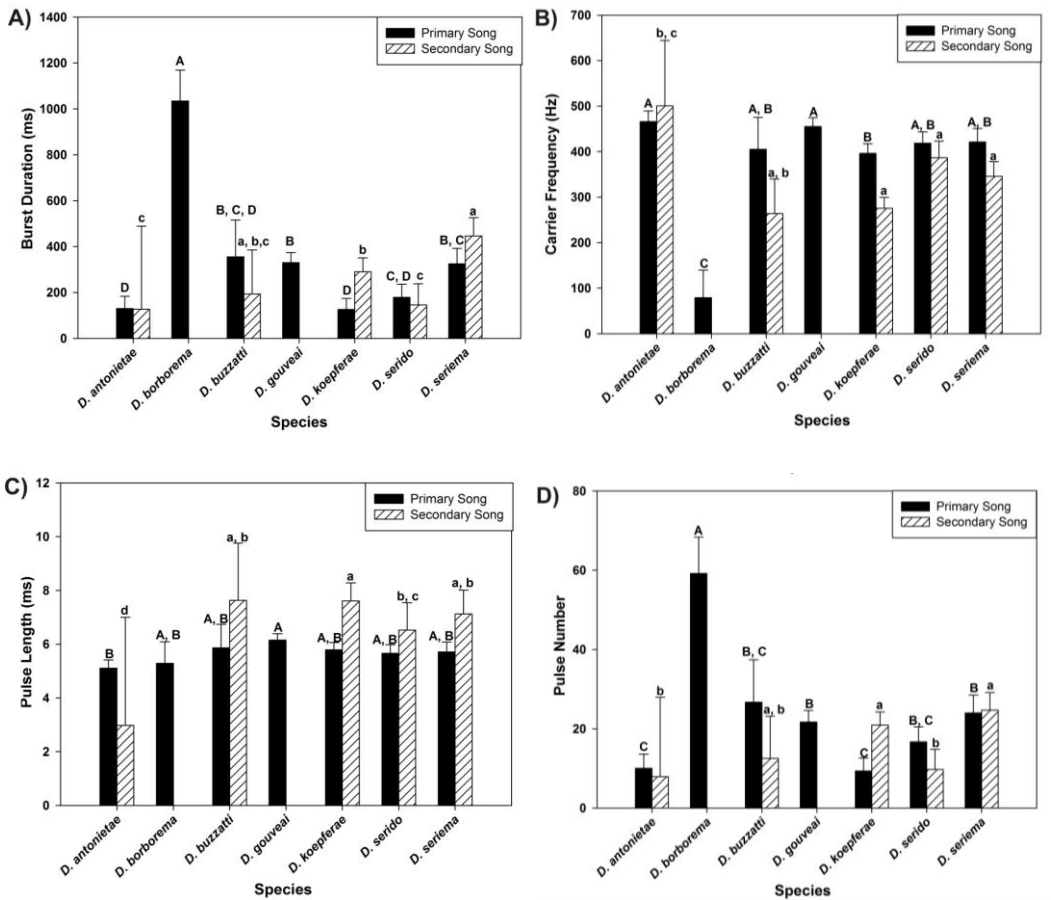


Figure 4. A-F. (Continued).

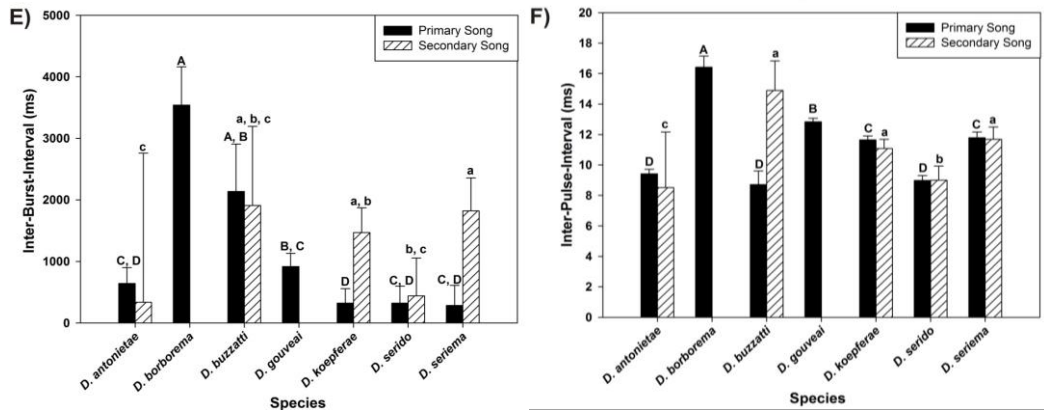


Figure 4. A-F. Least square means $\pm$ SE of six song components. A) Burst duration, B) Carrier frequency, C) Pulse number, D) Pulse length, E) Inter-burst-interval, and F) Inter-pulse-interval. Nonsignificant means share the same letter. Interspecific comparisons were performed for primary and secondary song separately.

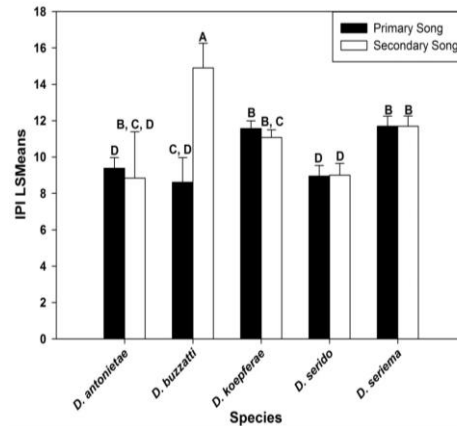


Figure 5. Differences in IPI, inter-pulse interval, between primary and secondary songs among five species using least square means $\pm$ SE. No comparison was performed for *D. borborema* and *D. gouveai* because these species lacked secondary song. Nonsignificant means share the same letter.

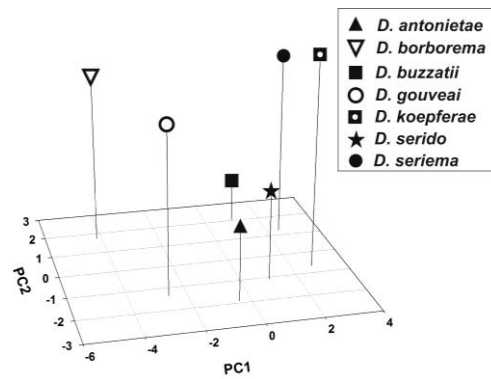


Figure 6. Three dimensional plot of the *D. buzzatii* species cluster based on the first three principal components (PCs) obtained from six song components (see Table 1). Altogether, the first three PCs explained 87% of the variance in the data (PC1 = 51%, PC2 = 20%, and PC3 = 16%).

Table 3. Mean $\pm$ SD of courtship song parameters of the *D. buzzatii* species cluster

Species		Burst Duration (ms)			Carrier Frequency (Hz)			Pulse Length (ms)			Pulse Number			Inter-burst-interval (ms)			Inter-pulse-interval (ms)		
		Mean	SD	n	Mean	SD	n	Mean	SD	n	Mean	SD	n	Mean	SD	n	Mean	SD	n
<i>D. antonietae</i>	PS	142.72	45.45	50	447.51	51.14	50	5.25	0.60	250	10.82	3.91	50	588.26	272.69	50	9.52	0.40	250
	SS	78.67	7.36	50	359.34	60.43	50	6.48	1.50	250	5.77	1.30	50	501.35	149.60	50	9.90	0.90	250
<i>D. borborema</i>	PS	660.44	249.12	50	213.13	111.32	50	5.54	1.20	250	38.58	13.63	50	1312.00	1295.20	50	14.20	1.30	250
	SS	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
<i>D. buzzatii</i>	PS	438.74	121.28	50	467.51	34.71	50	5.29	0.70	250	42.30	11.13	50	1348.80	666.46	50	7.92	1.00	250
	SS	201.22	20.47	50	324.30	105.78	50	7.75	0.70	250	13.67	1.76	50	791.85	695.21	50	12.30	1.50	250
<i>D. gouveai</i>	PS	330.2	117.28	50	456.26	52.28	50	6.14	0.60	250	21.80	7.62	50	938.48	765.76	50	12.80	1.10	250
	SS	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
<i>D. koepferae</i>	PS	129.34	33.91	50	396.87	43.24	50	5.92	0.90	250	9.62	2.73	50	350.60	107.25	50	11.50	0.70	250
	SS	290.18	68.58	50	274.67	30.09	50	7.42	3.20	250	20.63	8.68	50	1536.80	2351.90	50	11.10	2.20	250
<i>D. serido</i>	PS	189.26	85.49	50	425.00	27.01	50	5.41	0.50	250	17.08	7.46	50	347.56	130.95	50	9.19	0.50	250
	SS	145.48	94.84	50	379.69	45.15	50	6.22	1.70	250	11.28	6.06	50	480.73	180.28	50	8.95	1.10	250
<i>D. seriema</i>	PS	385.42	154.76	50	417.51	56.12	50	5.64	0.60	250	28.10	11.48	50	534.84	466.07	50	11.40	0.50	250
	SS	440.17	343.09	50	347.04	83.39	50	7.27	1.70	250	24.28	16.91	50	1250.70	666.56	50	11.6	2.30	250

ms = milliseconds; Hz = Hertz; N = number of flies recorded, i.e., 10 males per species; n = number of sample size of each parameter; SD = standard deviation; PS = primary song; SS = secondary song.

Table 4. Eigenvectors from a principal component analysis for six courtship song traits

Variable	PC1 (51%)	PC2 (20%)	PC3 (16%)	PC4 (6%)	PC5 (3%)
BD – PS	-0.233	0.516	0.063	-0.003	0.091
CF – PS	0.203	-0.268	-0.329	0.620	0.394
PN – PS	-0.135	0.569	-0.154	0.180	0.061
PL – PS	-0.017	-0.127	0.565	0.592	-0.429
IBI – PS	-0.240	0.394	-0.184	0.386	-0.130
IPI – PS	-0.221	0.040	0.571	-0.159	0.146
BD – SS	0.351	0.202	0.204	-0.011	0.424
CF – SS	0.375	0.112	-0.120	-0.138	-0.331
PN – SS	0.367	0.170	0.170	-0.070	0.250
PL – SS	0.377	0.159	-0.056	-0.056	-0.353
IBI – SS	0.309	0.151	0.310	0.176	0.225
IPI – SS	0.379	0.187	-0.033	0.035	-0.289

Variables were generated using values corrected for temperature variation. Only the first five principal components (PC1 – PC5) are shown that accounted for 96% of the variation in the data. Values in parentheses are the percentages of the variance explained by each PC. Numbers in bold represent the song traits that most contributed to the PC scores. See Table 1 for description of song traits. BD = burst duration; CF = carrier frequency; PN = pulse number; PL = pulse length; IBI = inter-burst interval; IPI = inter-pulse interval; PS = primary song; SS = secondary song.

Table 5. Analysis of congruence between the chromosomal inversion plus per gene phylogeny and courtship song data

	Parsimony methods									Test for Serial Independency (TFSI)	
	Linear parsimony (LP)			Squared change parsimony gradual (SCPG)			Squared change parsimony punctuated (SCPP)				
<i>Characters</i>	<i>Reference Tree</i>	<i>Random Trees</i>	<i>P</i>	<i>Reference Tree</i>	<i>Random Trees</i>	<i>P</i>	<i>Reference Tree</i>	<i>Random Trees</i>	<i>P</i>	Observed Mean C- Statistics	<i>P</i>
BD-PRD	926.13	879.83	0.5409	30856.68	31145.87	0.5709	140094.71	133425.24	0.6271	-0.0538	0.3940
BD-RES	181.61	177.43	0.5901	2668.23	1954.91	0.7543	9250.87	8415.69	0.5844	-0.1063	0.4740
IBI-PRD	4.64	4.29	0.5467	0.79	0.56	0.8254	2.92	2.43	0.7834	-0.1547	0.2670
IBI-RES	0.23	0.23	0.4710	0.01	0.006	0.7441	0.03	0.03	0.8253	-0.1316	0.1160
IPI-PRD	26.26	25.30	0.4487	16.86	19.92	0.4412	79.62	85.67	0.3648	0.0663	0.2870
IPI-RES	1.31	1.27	0.7381	0.17	0.09	0.9409	0.47	0.40	0.7327	-0.1535	0.2330
CV1-PRD	314.69	389.05	0.1079	3372.06	6074.17	0.0794	15141.00	24097.36	0.0896	0.2304	0.2010
CV1-RES	2.37	3.62	0.0438	0.31	0.57	0.1065	1.62	2.25	0.1107	0.18	0.2760
CV2-PRD	164.95	177.71	0.2599	2366.06	1406.55	0.8578	5365.56	5527.05	0.4406	0.0411	0.4800
CV2-RES	1.94	1.78	0.5990	0.13	0.14	0.5272	0.78	0.54	0.9838	-0.33	0.0530
CV3-PRD	53.59	53.61	0.1685	163.13	190.85	0.5152	654.49	761.93	0.1769	0.0466	0.2650
CV3-RES	1.22	1.50	0.1358	0.08	0.09	0.5113	0.26	0.34	0.1771	0.18	0.1340
PC1-PRD	11.99	12.40	0.2345	12.25	6.76	0.9075	27.98	26.91	0.5137	-0.0506	0.4750
PC1-RES	10.00	10.94	0.1032	3.01	5.38	0.1255	22.00	21.33	0.5395	-0.05	0.5430
PC2-PRD	7.07	8.18	0.2973	1.51	2.68	0.1380	9.32	10.57	0.3282	0.0654	0.4460
PC2-RES	4.00	4.00	0.5000	0.53	1.55	0.0464	8.42	6.13	0.9061	-0.24	0.1020
PC3-PRD	6.02	6.88	0.2060	0.90	1.99	0.0472	8.68	7.87	0.5831	-0.0230	0.4070

	Parsimony methods									Test for Serial Indepdency (TFSI)	
	Linear parsimony (LP)			Squared change parsimony gradual (SCPG)			Squared change parsimony punctuated (SCPP)				
Characters	Reference Tree	Random Trees	P	Reference Tree	Random Trees	P	Reference Tree	Random Trees	P	Observed Mean C-Statistics	P
PC3-RES	8.00	7.42	0.4673	3.53	2.55	0.8553	10.64	10.06	0.6180	-0.12	0.1860
PC4-PRD	4.35	4.02	0.9084	0.74	0.73	0.6158	3.20	2.89	0.6300	-0.1597	0.2760
PC4-RES	15.00	13.94	0.4607	4.74	9.43	0.0514	44.00	37.26	0.8321	-0.14	0.2390
PC5-PRD	1.48	1.66	0.1448	0.24	0.15	0.8154	0.54	0.57	0.4503	0.0180	0.4250
PC5-RES	11.00	11.83	0.1625	3.83	5.73	0.9551	19.76	22.64	0.2994	0.05	0.5580

Individual song components, i.e., burst duration (BD), inter-burst interval (IBI), and inter-pulse interval (IPI), were mapped onto a phylogeny that included the seven species of the *D. buzzatii* cluster plus *D. movajensis* as an outgroup. Furthermore, CDF analysis and PC analysis were used to obtain canonical variates (CVs) and principal components (PCs). These variables were also mapped onto a phylogeny, but this time it did not include an outgroup. All variables mapped onto the phylogeny were based on predicted (PRD) and residual (RES) values. Three different parsimony methods were used in Mesquite: linear parsimony (LP), squared-change parsimony assuming a gradual model of evolution (SCPG), and squared-change parsimony with a punctuated model of evolution (SCPP). In all three models, presence of phylogenetic signal for each character was assessed by comparing the mean parsimony character steps from the reference tree (as shown on Figure 7) with those of a population of random trees. Terminal taxa were reshuffled 10,000 times to generate the random trees. Phylogenetic signal was positive when the mean parsimony character steps for the reference tree were significantly smaller than the mean parsimony character steps for the random trees. The detection of phylogenetic signal was also examined with the test for serial independence (TFSI) run with 1,000 replicates using the program Phylogenetic Independence 2.0. P-values were adjusted using false discovery rate (FDR) analysis. No p values were significant after the adjustment.

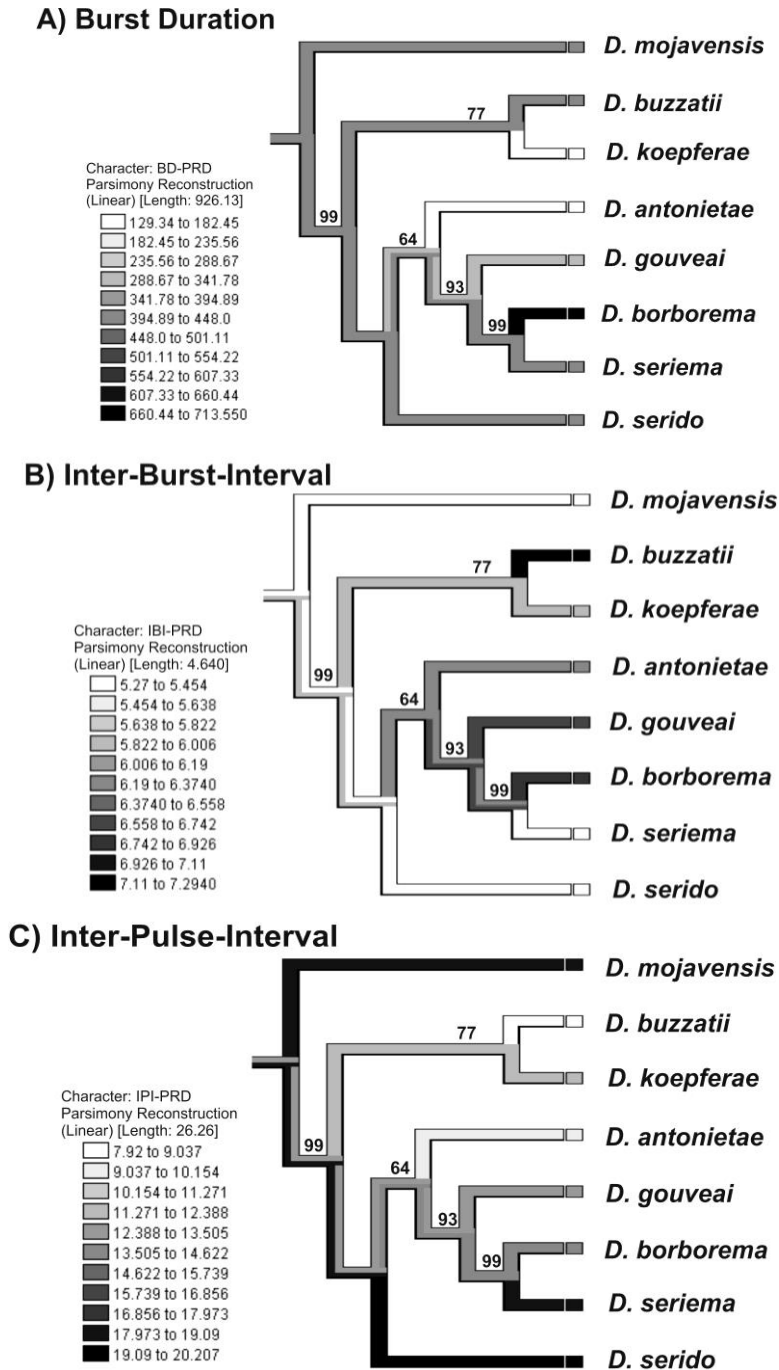


Figure 7. A-C. Phylogenetic character mapping using the linear parsimony model. This phylogeny represents a most parsimonious tree (one of two trees) of species of the *D. buzzatii* cluster inferred from chromosomal inversions (Ruiz et al. 1997; 2000) and *period* gene (Franco et al. 2011). *D. mojavensis* was used as an outgroup species. A) Burst Duration; B) Inter-Burst-Interval; and C) Inter-Pulse-Interval. Bootstrap values (shown above the nodes) were based on 1,000 replicates and 100 random additions. Only bootstrap values above 50% are shown.

CONCLUSION

Our comparative analysis of quantitative variation in male courtship songs revealed that song evolution was uncorrelated with the phylogenetic relationships among species. Mapping primary and secondary pulse songs types onto the phylogeny revealed that the presence of two songs is ancestral in the *D. buzzatii* cluster (Figure 2). These findings are in agreement with Ewing and Miyan (1986), who suggested that two song types is a primitive character state in the *D. repleta* group. They also proposed that differences in IPI (short and long IPIs) were responsible for the differences between A and B songs in species of the *D. repleta* group. We did not find clear correspondence between A and B songs and our designations of primary and secondary songs. Only *D. buzzatii* males produced significantly different primary and secondary songs, i.e., short IPI for primary song and long IPI for secondary song (Figure 5). In the other four species that possessed both types of songs, IPIs were unimodal. Ewing and Miyan (1986) also reported that *D. buzzatii* produced A song, but lacked B song. Males of the strain of *D. buzzatii* that we analyzed clearly presented two types of songs, so it is likely that there is intraspecific variation in song types for *D. buzzatii*. Intriguingly, they also described *D. mojavensis* as having only A song, but other studies have found both types of songs (Byrne 1999; Etges, et al. 2006).

The high levels of variation in courtship song among the different species of the *D. buzzatii* cluster are in agreement with other studies involving closely related *Drosophila* species (Cowling and Burnet 1981; Ewing and Miyan 1986; Hoy et al. 1988; Hoikkala et al. 1994; Tomaru and Oguma 1994; Ritchie and Gleason 1995; but see Noor et al. 2000). These findings have suggested that aspects of courtship behavior and mate recognition can be more distinct than morphology or other traits in closely related species (Butlin and Ritchie 1994; Mendelson and Shaw 2005). In fact, it is common for cryptic species to show low levels of genetic divergence in contrast to major differences in courtship behavior phenotypes (Henry et al. 2002). These patterns of differentiation are consistent with a significant role of sexual selection promoting sexual isolation and speciation (Panhuis et al. 2001). However, Noor et al. (2000) observed a lack of divergence in courtship songs between recently diverged subspecies *Drosophila pseudoobscura pseudoobscura* and *D. p. bogotana*, implying that rates of evolution in courtship songs may vary among different species groups. Also, Costa et al. (2000) observed low levels of intraspecific courtship song variation in *D. meridionalis*, also a member of the *D. repleta* group, despite the fact that karyotypic differentiation has been reported in different populations. Future studies will have to include fine-scale intraspecific, population level studies to gauge accurate rates of courtship signal and male-female signaling system evolution.

Similar to many other *Drosophila* species, courtship songs in the *D. buzzatii* cluster were characterized by low-frequency songs (Figure 1) limiting the use of male song to close-range courtship (Ewing 1983). Hawaiian *Drosophila* species are a remarkable exception, since these species produce high-frequency songs (Hoy, et al. 1988). High levels of quantitative variation were observed within *D. buzzatii* cluster species for some of the song components, e.g., burst duration and pulse number (Figure 4A, D), implying that these two traits may not be reliable species-specific signals as they would not serve as consistent species recognition signals. Furthermore, for song traits that were more species specific, e.g., pulse length and IPI (Figure 4C, F), there was significant overlap among species suggesting that females may use more than one song component during mate and/or species recognition. Despite large variation at the

individual level, courtship songs, particularly primary songs, were species-specific in the *D. buzzatii* cluster (Table 2, Figure 6). Certainly, the large differences in male courtship songs are likely to play a role in interspecific sexual isolation in the *D. buzzatii* cluster (Oliveira et al., unpubl. data), but song playback experiments with wingless males (Byrne 1999) have yet to be performed.

In the *fasciola* subgroup, a basal clade of species in the *D. repleta* group, Costa and Sene (2002) reported that IPI was species-specific with little intraspecific variation, suggesting a potential role in species recognition. In *D. montana*, females prefer songs with short pulses and high carrier frequency (Ritchie et al. 2005). In *D. ananassae* and *D. pallidosa* females recognize the frequency spectra of bursts (determined by inter-pulse-interval, intra-pulse-frequency and cycles per pulse) as a species-specific signal rather than individual song components (Yamada, et al. 2002).

EVOLUTION OF COURTSHIP SONGS

No phylogenetic signal was observed when we mapped quantitative song traits onto an independently derived phylogeny of the *D. buzzatii* cluster (Table 5, Figure 7) consistent with rapid evolution of male courtship songs. Although rates of species divergence in this group given their genetic affinities have been well described (e.g., Manfrin and Sene 2006), our analyses need to be broadened to a larger sampling of *D. repleta* group species in order to assess the validity and generality of our conclusions. Few comparative studies have investigated the evolution of behavioral traits involved in mate recognition and reproductive isolation (e.g., Ewing and Miyan 1986; Kusmierski et al. 1997; Gleason and Ritchie 1998; Henry et al. 1999; Etges and Noor 2002; Symonds and Elgar 2004; Grace and Shaw 2012). Even scarcer are studies that have mapped quantitative variation, rather than categorical differences, in behavioral traits onto a phylogeny. This is in part because it has been difficult to identify sufficient numbers of species clusters whose members are in different stages of reproductive isolation and for which there is comparative data for phenotypes involved with pre- and/or postmating isolation. Furthermore, the long-standing view that the evolution of behavioral traits is weakly or uncorrelated with phylogeny (e.g., Atz 1970; Baroni Urbani 1989; Blomberg et al. 2003) has certainly contributed to a *a priori* view that all behavioral traits are labile.

In their comparative study of the courtship songs in 22 species of the *D. repleta* group, Ewing and Miyan (1986) mapped the evolution of song types, A and B songs, onto a phylogeny based on chromosomal inversions. They observed that presence of two song types was ancestral in the group, but over evolutionary time some species have lost one of the song types while in others B song has become more complex. Etges (2002) mapped song types onto the phylogeny of the group using an available species phylogeny (Durando, et al. 2000), and showed that song type evolution was not concordant with the observed phylogenetic relationships among species and has been characterized by diversification, character loss, and reversal. Contrasting results have been found in other animal species regarding the evolution of courtship songs. In the *Drosophila willistoni* group, Gleason and Ritchie (1998) observed that song divergence was variable and not correlated with genetic divergence. In green lacewings songs were homoplastic (Henry, et al. 1999; 2012), while in oropendola birds songs were more conserved (Price and Lanyon 2002).

The inclusion of more comparative data is needed to calibrate rates of song evolution. Clearly courtship songs have evolved more rapidly than species diversification in several *Drosophila* species groups, but other factors must be involved in shaping larger phylogenetic trends in song evolution. The *D. repleta* group is a potentially useful group for this analysis since it is one of the largest monophyletic groups of *Drosophila*, with over 100 species (Throckmorton 1982; Vilela 1983; Durando, et al. 2000; Oliveira, et al. 2012), and is composed of species that have been intensively studied for decades, as is the case of the species of the *D. mojavensis* and *D. buzzatii* clusters (Byrne 1999; Etges 2002; Etges, et al. 2006; Manfrin and Sene 2006; Etges, et al. 2007). Using the comparative approach in a hypothesis testing framework, assessing rates of character evolution and the influence of phylogenetic affinities for mate recognition signals should help to resolve broad-scale evolutionary trends in mating signal evolution and provide some clarity into the origins of the spectacular diversity of mate communication systems we seek to understand.

ACKNOWLEDGMENTS

We thank P.R. Epifânio for technical assistance and C. H. A. S. Motta for recording the songs of *D. borborema*. F. F. Franco and A. L. H. Esguicero assisted with field collection of cactus and flies. Funding was provided by a supplement to National Science Foundation DEB-0211125 to WJE, a Dissertation Research Award from the J. William Fulbright College of Arts & Sciences, University of Arkansas to CCO, and grants from FAPESP, CNPq and USP to MHM and FMS.

REFERENCES

- Abouheif E. 1999. A method for testing the assumption of phylogenetic independence in comparative data. *Evol Ecol Res* 1: 895-909.
- Atz J. W. 1970. *The application of the idea of homology to behavior*. In: Aronson L. R., Tobach E., Lehrman D. S., Rosenblatt J. S., editors. *Development and Evolution of Behavior: Essays in Memory of T. C. Schneierla*. San Francisco: W. H. Freeman. p. 53-74.
- Baroni Urbani C 1989. Phylogeny and behavioural evolution in ants, with a discussion of the role of behaviour in evolutionary processes. *Ethol Ecol Evol* 1: 137-168.
- Benjamini Y., Hochberg Y. 1995. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J R Stat Soc B* 57: 289-300.
- Bennet-Clark H. C. 1984. A particle velocity microphone for the song of small insects and other acoustic measurements. *J Exp Biol* 108: 459-463.
- Bennet-Clark H. C., Ewing A. W. 1969. Pulse interval as a critical parameter in the courtship song of *Drosophila melanogaster*. *Anim Behav* 17: 755-759.
- Bizzo N. M. V. 1983. *Estudos sobre a biologia e isolamento reprodutivo em Drosophila serido*. [M.S.]. [Studies on the biology and reproductive isolation in *Drosophila serido*]. [São Paulo]: Universidade de São Paulo.
- Blomberg S. P., Garland T., Jr., Ives A. R. 2003. Testing for phylogenetic signal in comparative data: Behavioral traits are more labile. *Evolution* 57: 717-745.

- Butlin R. K., Ritchie M. G. 1994. *Mating behaviour and speciation*. In: Slater P. J. B., Halliday T. R., editors. *Behaviour and Evolution*: Cambridge University Press. p. 43-79.
- Byrne B. C. 1999. *Behaviour-genetic analysis of lovesongs in desert species of Drosophila*. [Ph.D.]. [Leicester]: University of Leicester.
- Cobb M., Burnet B., Connolly K. 1988. Sexual isolation and courtship behaviour in *Drosophila simulans*, *D. mauritiana* and their interspecific hybrids. *Behav Genet* 18: 211-255.
- Costa C. T. A., Kuhn G. C. S., Sene F. M. 2000. Low courtship song variation in South and Southeastern Brazilian populations of *Drosophila meridionalis* (Diptera, Drosophilidae). *Rev Bras Biol* 60: 53-61.
- Costa C. T. A., Sene F. M. 2002. Characterization of courtship sounds of species of the subgroup *fasciola* (Diptera, Drosophilidae, *Drosophila repleta* group): Interspecific and interpopulational analyses. *Brazil J Genet* 62: 573-583.
- Cowling D. E., Burnet B. 1981. Courtship songs and genetic control of their acoustic characteristics in sibling species of the *Drosophila melanogaster* subgroup. *Anim Behav* 29: 924-935.
- Coyne J. A., Orr H. A. 2004. *Speciation*. Sunderland, MA: Sinauer.
- de Brito R. A., Manfrin M. H., Sene F. M. 2002. Mitochondrial DNA phylogeography of Brazilian populations of *Drosophila buzzatii*. *Genet Mol Biol* 25: 161-171.
- Durando C. M., Baker R. H., Etges W. J., Heed W. B., Wasserman M., de Salle R 2000. Phylogenetic analysis of the *repleta* species group of the genus *Drosophila* using multiple sources of characters. *Mol Phylogenet Evol* 16: 296-307.
- Etges W. J. 2002. Divergence in mate choice systems: does evolution play by rules? *Genetica* 116: 151-166.
- Etges W. J., Noor M. A. F. 2002. *Genetics of mate choice: From sexual selection to sexual isolation*. New York: Kluwer. p. 260.
- Etges W. J., Oliveira C. C., Gragg E., Ortiz-Barrientos D., Noor M. A. F., Ritchie M. G. 2007. Genetics of incipient speciation in *Drosophila mojavensis*. I. Male courtship song, mating success, and genotype x environment interactions. *Evolution* 61: 1106-1119.
- Etges W. J., Over K. F., Oliveira C. C., Ritchie M. G. 2006. Inheritance of courtship song variation among geographically isolated populations of *Drosophila mojavensis*. *Anim Behav* 71: 1205-1214.
- Ewing A. W. 1983. Functional aspects of *Drosophila* courtship. *Biol Rev* 58: 275-292.
- Ewing A. W. 1989. *Arthropod Bioacoustics: Neurobiology and Behaviour*. Edinburgh, Scotland: Edinburgh University Press.
- Ewing A. W., Miyan J. A. 1986. Sexual selection, sexual isolation and the evolution of song in the *Drosophila repleta* group of species. *Anim Behav* 34: 421-429.
- Franco F. F., Silva-Bernardi E. C. C., Sene F. M., Hasson E. R., Manfrin M. H. 2010. Intra- and interspecific divergence in the nuclear sequences of the clock gene *period* in species of the *Drosophila buzzatii* cluster. *J Zool Syst Evol Res* 48: 322-331.
- Gleason J. M., Ritchie M. G. 1998. Evolution of courtship song and reproductive isolation in the *Drosophila willistoni* species complex: Do sexual signals diverge the most quickly? *Evolution* 52: 1493-1500.
- Grace J. L., Shaw K. L. 2012. Incipient sexual isolation in *Laupala cerasina*: Females discriminate population-level divergence in acoustic characters. *Curr Zool* 58: 416-425.

- Henry C. S., Brooks S. J., Duelli P., Johnson J. B., Wells M. M., Mochizuki A. 2012. Parallel evolution in courtship songs of North American and European green lacewings (Neuroptera: Chrysopidae). *Biol J Linn Soc* 105: 776-796.
- Henry C. S., Wells M. L. M., Holsinger K. E. 2002. The inheritance of mating songs in two cryptic, sibling lacewing species (Neuroptera: Chrysopidae: *Chrysoperla*). *Genetica* 116: 269-289.
- Henry C. S., Wells M. L. M., Simon C. M. 1999. Convergent evolution of courtship songs among cryptic species of the *carnea* group of green lacewings (Neuroptera: Chrysopidae: *Chrysoperla*). *Evolution* 53: 1165-1179.
- Hoikkala A. 2005. Inheritance of male sound characteristics in *Drosophila* species. In: Drosopoulos S, Claridge MF, editors. *Insect sounds and communication: Physiology, behaviour, ecology and evolution*. Boca Raton: CRC Taylor & Francis. p. 167-177.
- Hoikkala A., Kaneshiro K. Y., Hoy R. R. 1994. Courtship songs of the picture-winged *Drosophila planitibia* subgroup species. *Anim Behav* 47: 1363-1374.
- Hoy R. R., Hoikkala A., Kaneshiro K. 1988. Hawaiian courtship songs: Evolutionary innovation in communication signals of *Drosophila*. *Science* 240: 217-219.
- Kokudai C. B. S., Sene F. M., Manfrin M. H. 2011. Sympatry and asymmetric introgression between the cactophilic species *Drosophila serido* and *Drosophila antonietae* (Diptera: Drosophilidae). *Ann Entomol Soc Am* 104: 434-442.
- Kusmierski R., Borgia G., Uy A., Crozier R. H. 1997. Labile evolution of display traits in bowerbirds indicates reduced effects of phylogenetic constraint. *Proc R Soc Biol Sci Ser B* 264: 307-313.
- Kyriacou C. P., Hall J. C. 1980. Circadian rhythm mutations in *Drosophila melanogaster* affect short-term fluctuations in the male's courtship song. *Proc Natl Acad Sci USA* 77: 6729-6733.
- Kyriacou C. P., Hall J. C. 1982. The function of courtship song rhythms in *Drosophila*. *Anim Behav* 30: 794-801.
- Laurin M., Canoville A., Quilhac A. 2009. Use of paleontological and molecular data in supertrees for comparative studies: the example of lissamphibian femoral microanatomy. *J Anat* 215: 110-123.
- Liimatainen J., Hoikkala A., Aspi J., Welbergen P. 1992. Courtship in *Drosophila montana*: the effects of male auditory signals on the behaviour of flies. *Anim Behav* 43: 35-48.
- Maddison W. P., Maddison D. R. 2010. *Mesquite: A modular system for evolutionary analysis*. Version 2.74.
- Manfrin M. H., de Brito R. O. A., Sene F. M. 2001. Systematics and evolution of the *Drosophila buzzatii* (Diptera: Drosophilidae) cluster using mtDNA. *Ann Entomol Soc Am* 94: 333-346.
- Manfrin M. H., Sene F. M. 2006. Cactophilic *Drosophila* in South America: A model for evolutionary studies. *Genetica* 126: 57-75.
- Mendelson T. C., Shaw K. L. 2005. Sexual behaviour: Rapid speciation in an arthropod. *Nature* 433: 375-376.
- Mendelson T. C., Shaw K. L. 2012. The (mis)concept of species recognition. *Trends Ecol Evol* 427: 421-427.
- Moraes E. M., Spressola V. L., Prado P. R. R., Costa L. F., Sene F. M. 2004. Divergence in wing morphology among sibling species of the *Drosophila buzzatii* cluster. *J Zool Syst Evol Res* 42: 154-158.

- Moraes M. A. M. 1992. Isolamento sexual entre populações da superespécie *Drosophila serido* (Diptera, Drosophilidae). [M. S.]. [Sexual isolation between populations of the superspecies *Drosophila serido*]. [São Paulo]: Universidade de São Paulo.
- Noor M. A. F., Williams M. A., Alvarez D., Ruiz-Garcia M. 2000. Lack of evolutionary divergence in courtship songs of *Drosophila pseudoobscura* subspecies. *J Insect Behav* 13: 255-262.
- Oliveira C. C., Manfrin M. H., Sene F. M., Jackson L. L., Etges W. J. 2011. Variations on a theme: diversification of cuticular hydrocarbons in a clade of cactophilic *Drosophila*. *BMC Evol Biol* 11:179.
- Oliveira D. C. S. G., Almeida F. C., O'Grady P., Armella M. A., de Salle R., Etges W. J. 2012. Monophyly, divergence times, and evolution of host plant use inferred from a revised phylogeny of the *Drosophila repleta* species group. *Mol Phylogenet Evol* 427: 421-427.
- Panhuis T. M., Butlin R., Zuk M., Tregenza T. 2001. Sexual selection and speciation. *Trends Ecol Evol* 16: 364-371.
- Price J. J., Lanyon S. M. 2002. Reconstructing the evolution of complex bird song in the oropendolas. *Evolution* 56: 1514-1529.
- Raven. 2003. ver. Pro 1.3. Version Pro 1.3. Ithaca, New York, USA. www.birds.cornell.edu/raven: Cornell Laboratory of Ornithology.
- Reeve J., Abouheif E. 2003. *Phylogenetic Independence*. Version 2.0.
- Ritchie M. G., Gleason J. M. 1995. Rapid evolution of courtship song pattern in *Drosophila willistoni* sibling species. *J Evol Biol* 8: 463-479.
- Ritchie M. G., Halsey E. J., Gleason J. M. 1999. *Drosophila* song as a species-specific mating signal and the behavioural importance of Kyriacou & Hall cycles in *D. melanogaster* song. *Anim Behav* 58: 649-657.
- Ritchie M. G., Saarikettu M., Hoikkala A. 2005. Variation, but no covariance, in female preference functions and male song in a natural population of *Drosophila montana*. *Anim Behav* 70: 849-854.
- Ritchie M. G., Saarikettu M., Livingstone S., Hoikkala A. 2001. Characterization of female preference functions for *Drosophila montana* courtship song and a test of the temperature coupling hypothesis. *Evolution* 55: 721-727.
- Ritchie M. G., Townhill R. M., Hoikkala A. 1998. Female preference for fly song: playback experiments confirm the targets of sexual selection. *Anim Behav* 56: 713-717.
- Rodriguez-Trelles F., Alarcón L., Fontdevila A. 2000. Molecular Evolution and phylogeny of the buzzatii complex (*Drosophila repleta* group): A maximum-likelihood approach. *Mol Biol Evol* 17: 1112-1122.
- Ruiz A., Cansian A. M., Kuhn G. C. S., Alves M. A. R., Sene F. M. 2000. The *Drosophila serido* speciation puzzle: putting new pieces together. *Genetica* 108: 217-227.
- Ruiz A., Ranz J. M., Cáceres M., Segarra C., Navarro A., Barbadilla A. 1997. Chromosomal evolution and comparative gene mapping in the *Drosophila repleta* species group. *Brazil J Genet* 20: 553-565.
- Ruiz A., Wasserman M. 1993. Evolutionary cytogenetics of the *Drosophila buzzatii* species complex. *Heredity* 70: 582-596.
- Saarikettu M., Liimatainen J. O., Hoikkala A. 2005. The role of male courtship song in species recognition in *Drosophila montana*. *Behav Genet* 35: 257-263.
- SAS Institute Inc. 2004. SAS/STAT 9.1.2. Cary, NC: SAS Institute, Inc.

- Sene F. M., Manfrin M. H. 1998. A chamber to record the courtship sound in *Drosophila*. *Dros Inf Serv* 81: 203-204.
- Sene F. M., Pereira M. A. Q. R., Vilela C. R. 1982. *Evolutionary aspects of cactus breeding Drosophila in South America*. In: Barker J. S. F., Starmer W. T., editors. *Ecological Genetics and Evolution. The Cactus-Yeast-Drosophila Model System*. Sydney: Academic Press. p. 97-106.
- Silva A. F. G., Sene F. M. 1991. Morphological geographic variability in *Drosophila serido* (Diptera, Drosophilidae). *Rev Bras Entomol* 35: 455-468.
- Sony. 2006. *Sound Forge 7.0*. Sony Creative Software Inc., Madison, Wisconsin, USA <http://www.sonycreativesoftware.com/>.
- Swofford D. L. 2000. *PAUP**. Version 4.0. Sunderland, MA: Sinauer.
- Symonds M. R. E., Elgar M. A. 2004. The mode of pheromone evolution: evidence from bark beetles. *Proc R Soc Lond B Biol Sci* 271: 839-846.
- Throckmorton L. H. 1982. Pathways of evolution in the genus *Drosophila* and the founding of the *repleta* group. In: Barker J. S. F., Starmer W. T., editors. *Ecological genetics and evolution*. Sydney, Australia: Academic Press. p. 32-47.
- Tomaru M., Matsubayashi H., Oguma Y. 1995. Heterospecific inter-pulse intervals of courtship songs elicit female rejection in *Drosophila bauraria*. *Anim Behav* 50: 905-914.
- Tomaru M., Oguma Y. 1994. Genetic basis and evolution of species-specific courtship song in the *Drosophila auraria* complex. *Genetic Research*, Cambridge 63: 11-17.
- Vilela C. R. 1983. A revision of the *Drosophila repleta* species group (Diptera: Drosophilidae). *Rev Bras Entomol* 27: 1-114.
- von Schilcher F. 1976. The role of auditory stimuli in the courtship of *Drosophila melanogaster*. *Anim Behav* 24: 18-26.
- Wasserman M. 1992. Cytological evolution of the *Drosophila repleta* species group. In: Krimbas C. B., Powell J. R., editors. *Drosophila Inversion Polymorphism*. Boca Raton: CRC Press. p. 455-552.
- Yamada H., Sakai T., Tomaru M., Doi M., Matsuda M., Oguma Y. 2002. Search for species-specific mating signal in courtship songs of sympatric sibling species, *Drosophila ananassae* and *D. pallidosa*. *Genes Genet Syst* 77: 97-106.

Chapter 7

PREZYGOTIC ISOLATION IN THE PARASITOID WASP GENUS *NASONIA*

***Maartje C. W. G. Giesbers¹, Sylvia Gerritsma¹,
Jan Buellesbach², Wenwen Diao¹, Bart A. Pannebakker³,
Louis van de Zande¹, Thomas Schmitt⁴ and Leo W. Beukeboom¹***

¹Evolutionary Genetics, Center for Ecological and Evolutionary Studies,
University of Groningen, The Netherlands

²Department of Entomology, North Carolina State University, Raleigh, NC, US

³Laboratory of Genetics, University of Wageningen, The Netherlands

⁴Ecological Networks, Faculty of Biology,
Technical University of Darmstadt, Darmstadt, Germany

ABSTRACT

Nasonia (Hymenoptera: Pteromalidae) are small haplodiploid parasitoids of flesh- and blowfly pupae that have become model organisms for speciation research. The genus consists of four closely related species that harbor species-specific *Wolbachia* bacteria that cause postmating reproductive isolation. Antibiotic curing allows for interspecific crosses and genetic exchange between species which, together with haploidy of males, facilitates genetic analysis of fitness traits. In this chapter we synthesize the current knowledge on the different prezygotic isolation factors that act in the *Nasonia* genus, and on the genetic basis of these traits. A major prezygotic isolation factor is courtship behaviour. Species differ in male courtship behaviour, and there is large variation in interspecific mate discrimination depending on species pair. We summarize data on the strength of prezygotic isolation barriers between all possible species pairs and present new data on mate discrimination in choice and no-choice experiments. In tests of reinforcement, we found no stronger female mate discrimination of *N. vitripennis* strains occurring in microsympatry with *N. giraulti* compared to that of allopatric *N. vitripennis* strains. Additionally, we present data on the significance of cuticular hydrocarbon profiles for assortative mating in males and discuss other factors that may be involved in prezygotic isolation, including pheromone communication, within-host-mating and sneaking behaviour.

INTRODUCTION

The Biological Species concept states that species are ‘groups of interbreeding natural populations that are reproductively isolated from other such groups’ (Mayr 1942). When a reproductive barrier between two populations arises, local adaptation and drift can lead to divergence of these populations, and eventually to speciation (Mayr 1963). These reproductive isolation barriers can be divided into those that act before fertilization, i.e., prezygotic isolation, and those that act after fertilization, i.e., postzygotic isolation (Dobzhansky 1935, Mallet 1998). Prezygotic isolation barriers typically manifest prior to mating, for example through temporal or geographic isolation between individuals, morphological differences between mating partners, or behavioural isolation (Kondrashov and Shpak 1998, Coyne and Orr 1998). However, post-mating prezygotic isolation can also occur, for example through incompatibilities between egg and sperm (e.g., Shaw et al. 1994).

Assortative mating, i.e., preferred mating among individuals of the same species over matings between individuals of different species, is arguably the most prevalent, or at least best studied, form of prezygotic isolation. Despite the fact that prezygotic reproductive barriers do not always lead to completely isolated species for prolonged periods of time, behavioural isolation can apparently evolve rapidly (Coyne and Orr 1997, 2004). The evolution of behavioural barriers can be studied especially well in young species complexes, where hybridization can still occur and a full barrier to gene flow has not yet been established (e.g., Coyne et al. 2002, Gow et al. 2006). A well-known, and frequently studied, behavioural barrier is female interspecific mate discrimination, in which females of a particular species show strong assortative mating. This behaviour can be strengthened by reinforcement, a process in which prezygotic isolation barriers are increased by natural selection against unfit hybrids (Dobzhansky 1940, Servedio and Noor 2003).

Despite the attention for behavioural reproductive isolation in the process of speciation, the genetic basis of traits that play a role in maintaining and/or creating prezygotic isolation is not well known (e.g., Arbuthnott 2009). We do not have good answers yet to many questions, such as whether behavioural isolation is based on few genes with large effects or many genes with small effects, whether most behavioural changes cause reproductive isolation (hence behavioural genes being speciation genes, *sensu* Orr et al. 2004) or come into being after the speciation event (hence contributing to species differences). Such knowledge is required to understand how selection and drift may cause changes in reproductive isolation between populations in the course of evolution.

In this chapter we present the results of many years of research on speciation in the parasitoid wasp genus *Nasonia* (Hymenoptera: Pteromalidae). *Nasonia* have been used extensively for research on speciation and adaptation for a couple of decades now (e.g., Werren 1983, Pultz and Leaf 2003, Leonard and Boake 2006), and their haplodiploid sex determination system makes them very well suited for genetic analyses (Beukeboom and Desplan 2003, Werren and Loehlin 2009, Werren et al. 2010). The complete genome sequences are available for all four species, making this young species complex ideal for genomic studies on complex traits (Werren et al. 2010). In this chapter, we present an overview of the knowledge on prezygotic isolation between the *Nasonia* species. We summarize previously published data, and provide new information that has been gathered by our research groups in the course of time.

***Nasonia* Wasps**

Nasonia (Hymenoptera: Pteromalidae) is a genus of 2-3 mm sized parasitoid wasps (Figure 1) that lay their eggs in pupae of various fly species, such as Protocalliphora, that occur in bird nests and on carcasses (Whiting 1967, Darling and Werren 1990). A clutch typically consists of 20-30 eggs depending on host size (Charnoy and Skinner 1984). Like all Hymenoptera, *Nasonia* has a haplodiploid mode of reproduction (Whiting 1967). Males are uniparentally produced and haploid; they develop from unfertilized eggs. Females have a mother and a father, are diploid and develop from fertilized eggs. Females store sperm after mating and can control the process of egg fertilization (Holmes 1972, Werren 1984, King and Skinner 1991). In the field, clutches typically contain mostly fertilized (female) eggs (Whiting 1967, Werren 1983), of which the developmental time is approximately 14-16 days at 25°C. Male offspring emerges first and stay at the natal patch where they mate with emerging females, that can both be their sisters and offspring of other founding parental females, if present (Whiting 1967, van den Assem et al. 1980, Leonard and Boake 2006). Females typically mate only once and disperse in search of new host patches, but they may become receptive a second time 24 hours after mating (van den Assem and Visser 1976, Grillenberger et al. 2009a). The combination of these life history traits makes *Nasonia* a prime example of local mate competition (Hamilton 1967, Werren 1980, 1983).

The *Nasonia* genus consists of four species: *N. vitripennis*, *N. giraulti*, *N. longicornis* (Darling and Werren 1990), and the recently discovered species *N. oneida* (Raychoudhury et al. 2010a). *N. vitripennis* diverged from its sister species approximately 1.0 million years ago (Campbell et al. 1993), while *N. longicornis* and *N. giraulti* diverged approximately 0.5 million years ago (Raychoudhury et al. 2010b), and *N. oneida* diverged from *N. giraulti* 0.4 million years ago (Raychoudhury et al. 2010a). *N. vitripennis* is a cosmopolitan species and occurs in sympatry with *N. giraulti* and *N. oneida* in eastern and with *N. longicornis* in western North America, respectively (Darling and Werren 1990; Raychoudhury et al. 2010a, Figure 2). *N. longicornis* is spatially isolated from *N. giraulti* and *N. oneida*, but *N. vitripennis* is the only species that occurs in true allopatry outside North America. In areas of sympatry, the opportunity for hybrid matings is potentially high because the species are morphologically very similar (Darling and Werren 1990), they can occupy the same host patch or even hatch from the same host pupa (Darling and Werren 1990, Grillenberger et al. 2009b) and they can hybridize in the laboratory. Exact frequencies of hybrid matings under natural conditions have not been investigated so far. However, there is strong postzygotic reproductive isolation between most of the species in nature due to infection with species-specific strains of *Wolbachia* bacteria that cause cytoplasmic incompatibility and hybrid breakdown (Breeuwer and Werren 1990, 1995). Such incompatibility, between egg and sperm that are infected with different *Wolbachia* strains, results in improper condensation, eventual damage and subsequent loss of the paternal chromosomes in fertilized eggs (Ryan and Saul 1968, Breeuwer and Werren 1990, Tram and Sullivan 2002). The resulting haploid embryos may either die or develop into haploid males (Breeuwer and Werren 1990, Bordenstein and Werren 1998, Bordenstein et al. 2003, Tram et al. 2006). Interspecific matings therefore usually result in all-male progenies, the haploid males being non-hybrid because of haplodiploidy, and sometimes in reduced offspring numbers. The strength of cytoplasmic incompatibility varies between species pairs, and is generally proportional to their divergence times. The *Wolbachia* strains of *N. vitripennis* induce complete incompatibilities with those of all three other species, *N. giraulti* and *N.*

longicornis strains are partially incompatible (Breeuwer and Werren 1990, Bordenstein et al. 2001), but the *Wolbachia* strains from *N. oneida* are fully compatible with those of *N. giraulti* and partially compatible with those of *N. longicornis* (see Raychoudhury et al. 2010a and Figure 4 for details). Antibiotic curing from the *Wolbachia* bacteria allows for heterospecific crosses in the laboratory that yield viable and (partially) fertile F1 females and F2 males, despite other pre- and postzygotic barriers that are present between the different species pairs at varying degrees.

Apart from cytoplasmic incompatibilities reducing the number of hybrid offspring, F2 hybrid males, between *N. vitripennis* and both *N. giraulti* and *N. longicornis*, suffer from strong postzygotic hybrid breakdown (Breeuwer and Werren 1995, Gadau et al. 1999, Niehuis et al. 2008, Clark et al. 2010, Koevoets et al. 2012) Together, through these postzygotic isolation mechanisms a substantial fitness disadvantage can be expected from females who permit courtship from and copulations with heterospecific males. Similarly, from the male perspective, investment in interspecific copulations means a complete absence of mating success in terms of paternal genetic contribution to the offspring (Bordenstein and Werren 2007). Thus, selection pressure towards mate discrimination and assortative mating is a likely consequence, favoring the ability to discriminate conspecific from heterospecific mating partners (Butlin 1987). In this chapter we present the current knowledge on prezygotic isolation mechanisms in the *Nasonia* genus and consider whether the observed species differences could have evolved in response to maladaptive hybridization.



Figure 1. *Nasonia vitripennis* female, parasitizing a host, and male (top left corner). Photos by Peter Koomen.

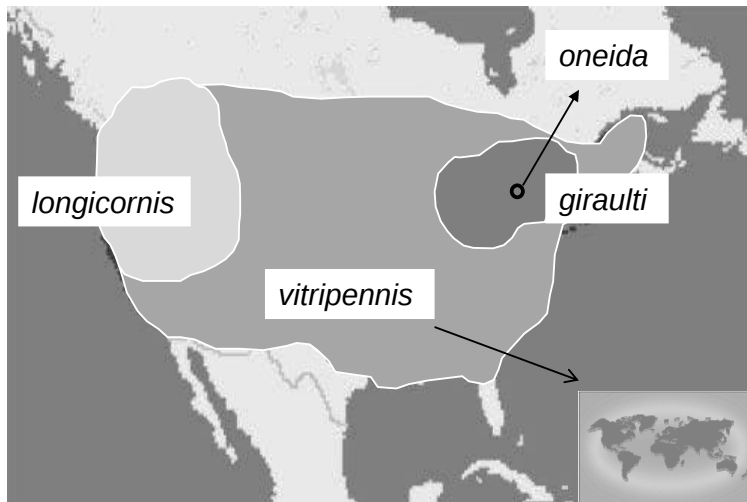


Figure 2. Geographical distribution of the four *Nasonia* species. *N. oneida* was recently discovered, and has as yet only been found in the region of Ithaca (NY). *N. vitripennis* is a cosmopolitan species.

COURTSHIP AND MATING BEHAVIOUR

The *Nasonia* species differ in male courtship behaviour and can be discerned by their species-specific alterations of a common courtship pattern (van den Assem and Werren 1994, Drapeau and Werren 1999). This difference alone is not sufficient to prevent interspecific matings. In fact, it is still not well known to what extent these differences in male courtship behaviour are perceived by females to affect their receptivity. Moreover, it is unclear whether male courtship behaviour has mostly been shaped by intraspecific sexual selection as suggested by Barrass (1976) and Jachmann and van den Assem (1996), or has also been modulated by interspecific interactions. In *Nasonia*, female reproductive behaviour appears to induce a stronger isolation than male behaviour does. In addition to information obtained from male courtship behaviour, females might discriminate against males of other species using species-specific differences in sex pheromones, aphrodisiacs and sound (Ruther et al. 2009, Robertson et al. 2010, Niehuis et al. 2011).

Nasonia has been extensively used for behavioural studies for several decades (Barrass 1960, van den Assem and Visser 1976; van den Assem 1986, Jachmann and van den Assem 1993; van den Assem and Werren 1994), including some studies of the underlying genetics of behaviour (Beukeboom and van den Assem 2001, 2002, Velthuis et al. 2005). In this chapter, female interspecific mate discriminating behaviour is considered to be the main prezygotic isolation mechanism that separates the *Nasonia* species. This behaviour, in which females of a species are reluctant to mate with heterospecific males, can be influenced by a number of different cues that are embedded in the mating sequence. However, research of the previous years focused almost solely on male courtship behaviour and female response. In laboratory observations, the start of a mating sequence is marked by the latency time. This is the time it takes the male and female to approach each other, and defined as the time period between the moment of introducing the two partners to the mating arena and the moment that the male mounts the female and takes up the courtship position on top of the female. Following some

evidence from van den Assem et al. (1980), Ruther et al. (2007) showed that *Nasonia vitripennis* males release a sex pheromone from their abdomen to attract females. This pheromone can be considered as a cue for the female, and might determine the duration of the latency time.

After the male has mounted the female, the male courtship behaviour of *Nasonia* males is composed of stereotypic movements, which differ quantitatively and qualitatively between the species (see for details van den Assem and Werren 1994, Drapeau and Werren 1999, Beukeboom and van den Assem 2001 and Figure 3). Typically, courting males take up an invariable courtship position on top of the female, with the front legs placed on the female's head (Barrass 1960). A conspicuous feature of the courtship performance is head nodding: repeated bouts of nods that are separated by short pauses (Barrass 1960). The first head nod in a bout coincides with mouthpart extrusions, which is probably linked to the release of a pheromone (van den Assem et al. 1980). This pheromone is believed to be involved in the process of induction of female receptivity (van den Assem et al. 1980), but its chemical composition has to date not been elucidated. If indeed the first head nod in a series coincides with pheromone release head nod frequency has at least a modulating role in the mating behaviour sequence. Therefore, head nodding behaviour and pheromone release are considered additional cues for the female to assess the male species.

A third factor that may play a role in *Nasonia* mate choice is male song. Male song, created by wing vibrations, is an important part of male courtship behaviour in *Drosophila*, where it has been studied in many species (e.g., Kyriacou and Hall 1982, Ritchie and Gleason 1995, Li et al. 2012). Male *Nasonia* wasps also vibrate their wings while performing head-nods, and these songs are also seen as part of their courtship behaviour. These vibrations can be recorded (van den Assem 1986, Diao et al. unpublished) and potentially differ between the four species, for example by reflecting the differences in male wing lengths between the species. However, the functional significance of song patterns in the *Nasonia* mating sequence remains to be analysed and requires additional manipulative experimentation.

After assessing the above mentioned cues, females typically show their willingness to copulate, and thus acceptance, during the first head nod of a series by sweeping their antennae downwards and opening their genital orifice by raising their abdomen. Males immediately back up and copulation follows (Barrass 1960). Copulations last for 12-14 seconds and do not seem to differ between the species. Immediately after copulation the male performs a sequence of postcopulatory courtship before dismounting. If females do not become receptive, as manifested by their unwillingness to adopt the copulatory position, males will eventually terminate their courtship performance and dismount. When presented with an already mated female, males of the four *Nasonia* species differ in the number of head nod cycles they perform until dismounting. For example, *N. vitripennis* typically terminates courtship faster (after 7-8 cycles) than *N. longicornis* (after 12-13 cycles) (Beukeboom and van den Assem 2001). In addition, the number of cycles is affected by the number of courtship bouts that the male has performed in the recent past (Jachmann and van den Assem 1996).

The degree of prezygotic isolation between the *Nasonia* species is likely determined by an interaction between male courtship behaviour, chemical communication, male song on the one hand, and the females' response to these traits, in terms of interspecific mate discrimination behaviour on the other. In contrast to males, which mate multiple times, females typically mate only once in nature, therefore selection for conspecific mating is expected to be stronger in females (van den Assem 1986, Burton-Chellew et al. 2007, Grillenberger et al. 2008).

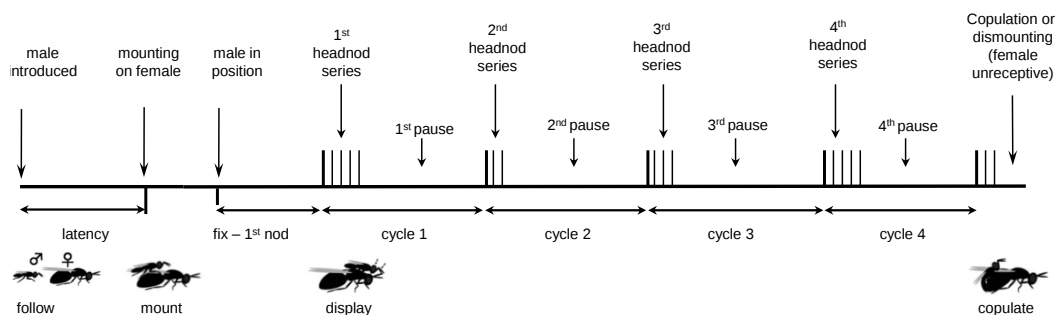


Figure 3. Schematic representation of the courtship pattern of *Nasonia*. Pictures by Michael Clark.

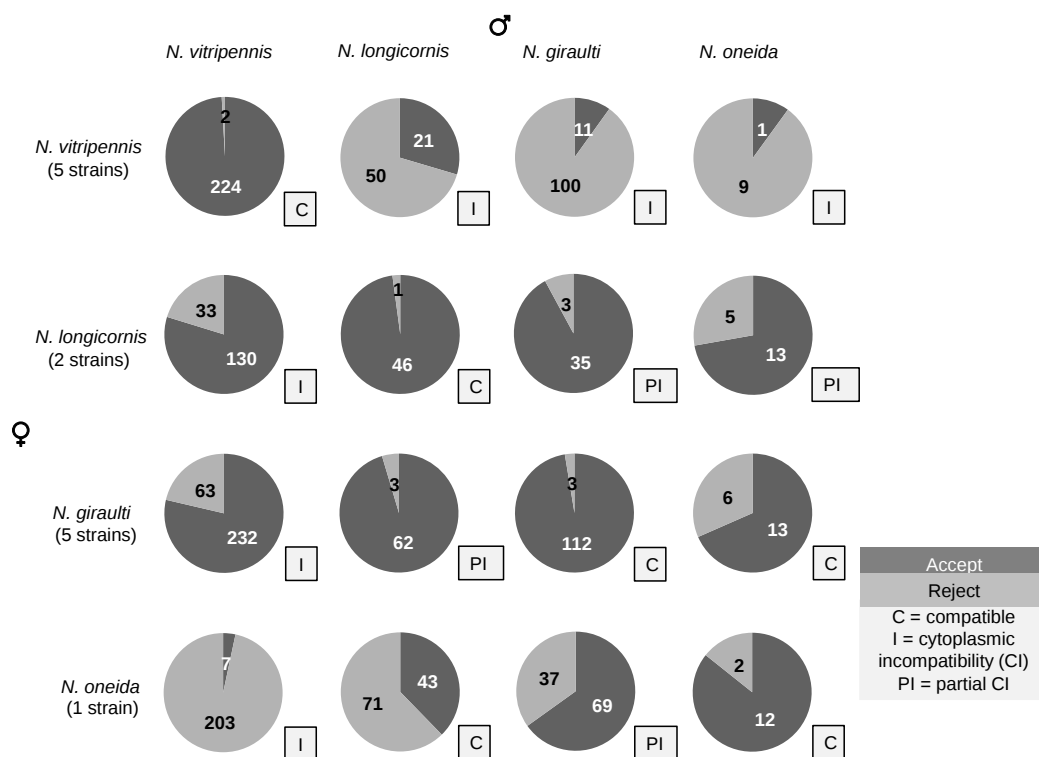


Figure 4. General patterns of intraspecific and interspecific female mate discrimination in the four *Nasonia* species. For each species pair, the number of females is shown that rejected or accepted the male partner. Observations are performed using a variety of strains; the number of strains used per species is indicated below the female species name. Occurrence and strength of cytoplasmic incompatibilities, induced by species-specific *Wolbachia* infections, are shown for all species pairs.

PATTERNS OF MATE DISCRIMINATION: NO CHOICE EXPERIMENTS

Tests for mate discrimination usually consist of behavioural observation trials (or mating trials) in which successful copulations are scored as present or absent. Female mate discrimination is then defined as the percentage of females that rejected their mate. Mating trials in *Nasonia* are typically performed as no-choice experiments by placing one virgin male

and one virgin female of standardized age (1 or 2 days old) in a small arena and observing them for 5-30 minutes. In intraspecific setups, the vast majority (>96%, Figure 4) of mating pairs result in copulations within the observation time frame. True choice experiments with two or more males or females are hard to perform as it remains difficult to interpret the interaction between interspecific male-male competition and female choice. The main reason for this is that males of the different species differ strongly in their aggressiveness, with *N. vitripennis* males being the most aggressive (Leonard and Boake 2006, this chapter). By combining a number of no-choice experiment datasets that have been collected in our laboratories over the years, a general overview can be made of female mate discrimination patterns between the four *Nasonia* species (Figure 4). In these no-choice experiments, mate discrimination is scored when a male successfully mounts and courts a female, but the female does not signal receptivity and no copulation occurs. Although significant quantitative differences are found between the studies, this does not lead to qualitative differences in the observed patterns, and therefore the results of the different studies can be combined. Furthermore, results found in our study are quite similar to data shown in Bordenstein et al. (2000). In the next section we will shortly discuss these similarities.

Strong asymmetric sexual isolation exists between the different *Nasonia* species, with *N. vitripennis* showing the highest female mate interspecific discrimination. *N. vitripennis* females have high mate discrimination against *N. giraulti* (90%) and *N. oneida* males (90%), but accept *N. longicornis* males more frequently (71% discrimination). *N. giraulti* and *N. longicornis* females show comparable heterospecific mate discrimination patterns against males of all three species: low mate discrimination against *N. vitripennis* males (20%), and slightly higher against *N. oneida* males (approximately 25%). As *N. longicornis* and *N. giraulti* do not occur in sympatry (Darling and Werren 1990), mate discrimination between these two species is expected to be low. Our results are consistent with this, *N. longicornis* females hardly discriminate against *N. giraulti* males (8%), and neither do *N. giraulti* females against *N. longicornis* males (5%). Similarly, *N. oneida* females discriminate strongly (97%) against males of the sympatric species *N. vitripennis*, but discrimination against *N. longicornis*, which is allopatric to *N. oneida*, is lower (63%). *N. giraulti* males are more often accepted by *N. oneida*, with only 25% of the females rejecting the *N. giraulti* males. This last result contradicts the prediction that mate discrimination of species that occur in sympatry is higher than allopatric species pairs. To further quantify the degree of prezygotic isolation between the four *Nasonia* species we calculated a prezygotic isolation index, using the following formula:

$$i = \frac{\text{frequency of heterospecific matings of both reciprocal crosses}}{\text{frequency of homospecific matings of both species}} = \frac{[A \times B] + [B \times A]}{[A \times A] + [B \times B]} \quad (1)$$

Indexes are shown in Table 1. We can infer that prezygotic isolation is strongest between *N. vitripennis* and *N. oneida*, and weakest between *N. giraulti* and *N. longicornis*, as is expected by their allopatric status. The prezygotic isolation index between *N. giraulti* and *N. oneida* is fairly low, contradicting what would be expected based on the theory of reproductive character displacement, which states that differences in behaviour (in this case displayed by mate discrimination) are stronger in sympatrically occurring species as opposed to allopatrically occurring species (Brown and Wilson 1956, Grant 1972).

Table 1. Patterns of prezygotic isolation between the four *Nasonia* species, as indicated by prezygotic isolation index (i). As this index is designed for heterospecific matings, no values can be calculated for intraspecific crosses

species	<i>N. vitripennis</i>	<i>N. longicornis</i>	<i>N. giraulti</i>	<i>N. oneida</i>
<i>N. vitripennis</i>				
<i>N. longicornis</i>	0.445			
<i>N. giraulti</i>	0.553	0.040		
<i>N. oneida</i>	0.932	0.434	0.204	

0 = no isolation.

1 = complete isolation.

Interestingly, asymmetric patterns exist in mate discrimination. This means that for two species A and B, species A could strongly discriminate against species B, but species B might not necessarily show such strong discrimination against species A. To indicate asymmetric patterns of mate discrimination, the prezygotic isolation index was slightly adjusted:

$$i' = \frac{\text{frequency of heterospecific matings with female of species A}}{\text{frequency of homospecific matings of species A}} \quad (2)$$

Indexes for asymmetrical isolation are shown in Table 2. From these indexes it can be concluded that strong asymmetrical patterns of prezygotic isolations exist for crosses with *N. vitripennis* (except for *N. vitripennis* and *N. oneida*), reflecting the strong discriminatory behaviour of *N. vitripennis* females. Asymmetry was also found in the *N. longicornis* and *N. oneida* species pair, as a result of strong mate discrimination of *N. oneida* females.

These results show that prezygotic isolation in the *Nasonia* complex is incomplete and confirm previous results (van den Assem and Werren 1994, Bordenstein et al. 2000). Additionally, the data are partially reflective of the geographical distribution and sympatry of the four *Nasonia* species, but are not entirely explained by these factors.

Table 2. Patterns of asymmetrical prezygotic isolation between the four *Nasonia* species, as indicated by the asymmetrical prezygotic isolation index (i'). The upper right section and lower left section can be compared to show asymmetries in isolation

		Male species			
		<i>N. vitripennis</i>	<i>N. longicornis</i>	<i>N. giraulti</i>	<i>N. oneida</i>
Female species	<i>N. vitripennis</i>		0.701	0.907	0.899
	<i>N. longicornis</i>	0.185		0.059	0.262
	<i>N. giraulti</i>	0.193	0.021		0.800
	<i>N. oneida</i>	0.965	0.609	0.792	

Table 3. Variation between strains of the same species for interspecific mate discrimination (IMD) against *N. vitripennis* and *N. giraulti* males

Female species	Male species	Total IMD	1	2	3	4	5	6	7	8	9	10
<i>N. vitripennis</i>	<i>N. giraulti</i>	91±1% (n=664)	97±2% (n=46)	89±3% (n=115)	100% (n=90)	93±4% (n=44)	100% (n=19)	98±2% (n=43)	82±4% (n=95)	92±4% (n=39)	73±5% (n=70)	96±2% (n=103)
<i>N. giraulti</i>	<i>N. vitripennis</i>	21±2% (n=295)	47±8% (n=36)	15±4% (n=68)	31±7% (n=48)	26±5% (n=74)	3±2% (n=69)					

Given is the sample size, percentage and standard error of females that discriminate against males of the other species. *N. vitripennis* strains used are 1-Sal12, 2-Sal8, 3-ITH4c, 4-ITH3F from Ithaca, New York; 5-LabII standard lab strain from Leiden, The Netherlands; 6-HV2004, 7-HV2003, 8-HVRx from The Netherlands; 9-ITA2 from Italy and 10-Russia Bait from Russia. *N. giraulti* strains used are 1-RV2 standard lab strain from New York; 2-GMix mixed population from 5 *N. giraulti* strains originating from Virginia, Pennsylvania and New York; 3-NGVA1 and 4-NGVA2 from Virginia; 5-NGPA from Pennsylvania.

In addition to these differences between the four *Nasonia* species, intraspecific differences in mating behaviour are common, and occasionally a *N. vitripennis* strain has been collected in Europe (e.g., ITA2 from Italy) that shows low interspecific mate discrimination. Also *N. giraulti* strains show large variations for interspecific mate discrimination against *N. vitripennis* (Table 3). Intraspecific crossings, however, are almost always successful regardless of the tested strains. For example, crosses between different *N. vitripennis* strains result in copulation in almost 100% of the cases. This has been tested by observing a total of 600 matings pairs of 20 strain combinations collected along a latitudinal gradient of 20 degrees (latitude 42-62) within Europe. Mate acceptance was high (97%) between all strain combinations (data not shown), indicating that there is no isolation between the strains. Only significant differences in latency time were found (GLM, $df=19$, $\chi^2=54.38$, $p < 0.0001$). These results point at absence of intraspecific isolation within *N. vitripennis* strains. This is consistent with the observation that European and North-American *N. vitripennis* strains cross readily in the laboratory (see also the section on reinforcement below).

CHOICE EXPERIMENTS: IMPLICATIONS OF MALE-MALE COMPETITION

In an experimental set-up in which we aimed to investigate female mate discrimination in *N. vitripennis*, when given a choice between a conspecific male and a *N. giraulti* male, we noticed a number of issues relating to male-male competition. A total of 200 *N. vitripennis* females were given a choice between a heterospecific and conspecific male. The conspecific *N. vitripennis* male was significantly more often the first male to court the *N. vitripennis* female ($\chi^2=39.8$, $df=3$, $p < 0.0001$). This shows that, when paired with a *N. vitripennis* female, *N. vitripennis* males almost always win the male-male competition with *N. giraulti* in terms of first mounting the female (Table 4). This is consistent with *N. vitripennis* males being more aggressive (Leonard and Boake 2006). Additionally, sneaking behaviour (van den Assem 1986) was observed in *N. vitripennis* males for 35% of the observed crosses, i.e., while the *N. giraulti* male performed courtship behaviour to induce female receptivity, the *N. vitripennis* male sits on the female's abdomen and copulates with her upon becoming receptive. In a choice experiment with *N. giraulti* females paired with both a *N. vitripennis* and a conspecific male, *N. vitripennis* males showed sneaking behaviour in 54% of the observed crosses. In contrast, *N. giraulti* males never performed sneaking behaviour, in choice experiments with either a *N. vitripennis* or *N. giraulti* female (see Table 4). Sneaking behaviour appears to be a strategy that has evolved in response to strong male-male competition within *N. vitripennis*, which is supported by observations of sneaking attempts in intraspecific crosses (data not shown). As male-male competition and differences in aggression (for example sneaking attempts), between the *Nasonia* males of the different species, likely interact with female choice, it is ambiguous to make conclusions on female interspecific mate discrimination using this experimental set-up. Testing female mate discrimination in a true choice experiment will require an experimental set-up that does not allow for male-male interactions.

Table 4. Mating behaviour differences in choice mating trials between *N. vitripennis* and *N. giraulti*

Species	First mounting of <i>N. vitripennis</i> ♂ in heterospecific contests. Cross with <i>N. vitripennis</i> ♀	Male sneaking behaviour in heterospecific contests. Cross with <i>N. vitripennis</i> ♀	Male sneaking behaviour in heterospecific contests. Cross with <i>N. giraulti</i> ♀
<i>N. vitripennis</i>	72% (N=200)	35% (N=192)	54% (N=94)
<i>N. giraulti</i>	58% (N=100)	0% (N=192)	0% (N=94)

GENETICS OF MATE DISCRIMINATION

The genetics of male reproductive behaviour in *Nasonia* has been investigated using interspecific crosses (Beukeboom and van den Assen 2001; Beukeboom et al., in prep). These studies reveal a complex regulation with multiple QTL for different courtship components and abundant epistatic interactions between different loci. The genetic architecture of female mating behaviour has not been studied to the same extent (but see Velthuis et al. 2005). Recently, Giesbers et al. (in prep) analysed the genetic basis of female interspecific mate discrimination (IMD) in *N. vitripennis* against *N. giraulti* males using artificial selection. They found a fast and significant response to selection resulting in a 35% decrease in mate discrimination within four generations of selection. The cumulative heritability of IMD in the artificial selection experiment was $h^2=0.29$, which shows the existence of substantial genetic variation for interspecific mate discrimination in *N. vitripennis*. Although more detailed genetic studies are required, these studies indicate that the *Nasonia* system is very well suited for investigating the genetic basis of female mate discrimination. Such studies can particularly be useful for testing the theoretically predicted genetic linkage between male signal and female choice traits, as well as for investigating the genetic basis of intraspecific versus interspecific mate choice.

OTHER TRAITS INVOLVED IN PREZYGOTIC ISOLATION

Another interesting prezygotic isolation inducing behaviour that has only been found in a few parasitic Hymenoptera species (da Costa Lima 1928, Suzuki and Hiehata 1985, Drapeau and Werren 1999) is within-host-mating, in which males and females mate inside the host before emergence. The *Nasonia* species differ in their preference for within-host-mating (Drapeau and Werren, 1999, Giesbers et al. in prep, this chapter). For our experiments, within-host-mating was scored by collecting all females from inside the host, at the moment when the first female has emerged from the host. Females that have mated inside the host will produce daughters and sons, while unmated females will only produce sons. Our results confirm earlier work by Drapeau and Werren (1999): *N. giraulti* has a high preference for within-host-mating, while *N. vitripennis* almost always mates outside the host (Figure 5). *N. longicornis* and *N. oneida* show low levels of within-host-mating. Drapeau and Werren (1999) already suggested that within-host-mating might have evolved in *N. giraulti* to avoid interspecific hybridization. As a consequence of within-host-mating, *N. giraulti* females will only encounter *N. vitripennis*

males after having mated conspecifically within the host. As re-mating rates in natural populations of at least *N. vitripennis* are very low (van den Assem et al. 1980, van den Assem and Jachmann, 1999, Grillenberger et al. 2009b), this would produce a strong barrier against interspecific mating. Additionally, it might explain our observation that *N. giraulti* females discriminate less against *N. vitripennis* males than *N. vitripennis* females discriminate against *N. giraulti* males. Our results are generally consistent with the prediction that the rate of within-host-mating is negatively correlated with the degree of interspecific mate discrimination. It remains to be seen if within-host-mating is a male or female mediated trait, or the result of an interaction between both sexes.

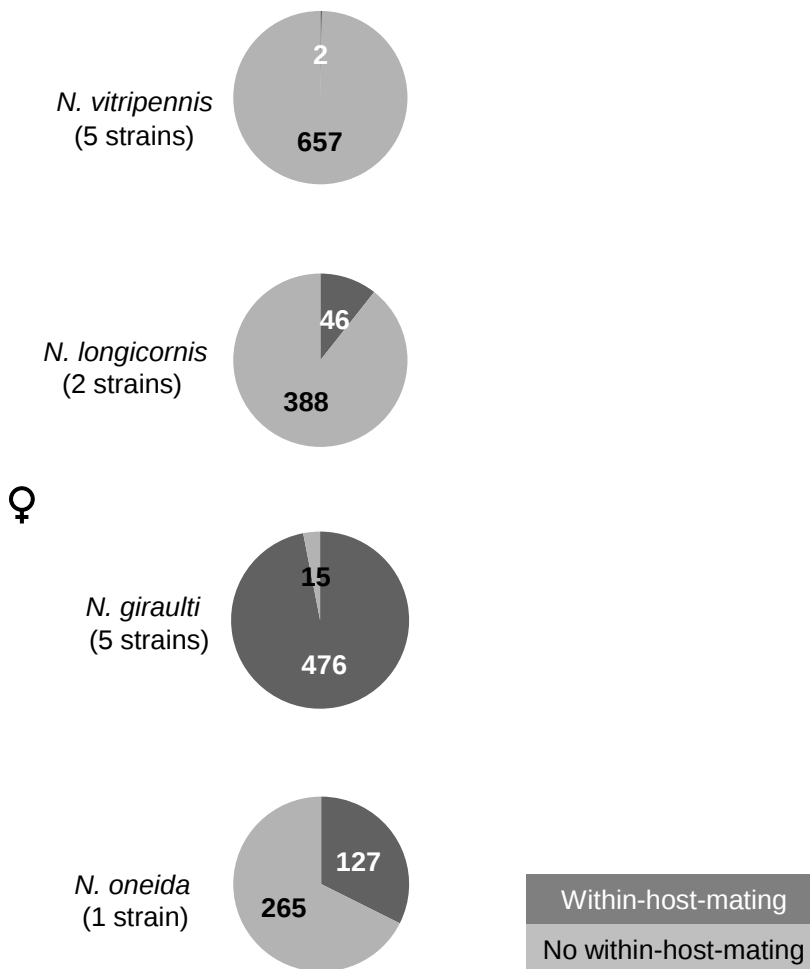


Figure 5. Interspecific differences in within-host-mating. For all four *Nasonia* species, the proportion of females is shown that mated either inside or outside the host. Observations are performed using a variety of strains; the number of strains used per species is indicated below the female species name. Values in the pie slices indicate the number of females.

In a study into the possible role of sexual conflict, Geuverink et al. (2009) compared copulation durations of conspecific and heterospecific crosses to test for possible mating incompatibilities (e.g., morphological or behavioural) between lines of *N. vitripennis* and *N.*

longicornis that were cured from their *Wolbachia* infection. In both species, heterospecific crosses did not differ in duration from conspecific crosses which suggests that sperm transfer is not mitigated in heterospecific crosses. This was further substantiated by occurrence of sperm in the spermatheca of heterospecifically mated females immediately after copulation. Moreover, the results indicated that the speed at which sperm is transferred and the amount of transferred sperm did not differ much between conspecific and heterospecific crosses. Furthermore, the survival of sperm in the course of time was not affected either, as females that had their reproductive period delayed by denying them access to hosts for the first ten days of their life, still appeared to have sperm in good condition in their spermatheca, regardless of whether they were heterospecifically or conspecifically mated. These data indicate that post-mating incompatibilities between heterospecific sperm and the spermatheca environment appear to play no role in *Nasonia*, at least between the two tested species *N. vitripennis* and *N. longicornis*.

Geuverink et al. (2009) used crosses between *N. vitripennis* and *N. giraulti* to determine whether female acceptance is influenced by the species of the first male partner. A clear effect of male species on female re-mating rate was found, both *N. vitripennis* and *N. giraulti* females mated more frequently after having first mated a heterospecific male. However, this increase was only significant for *N. giraulti* females, which may in part be due to the higher conspecific re-mating rate of *N. vitripennis* females under lab conditions. This shows that heterospecific males are less efficient in decreasing female receptivity. At present it is not possible to differentiate between the mechanisms that could reduce receptivity, such as differences in male postcopulatory behaviour or physiological effects of sperm in the reproductive tract of the female. Interestingly, *N. giraulti* males have longer post-copulatory courtship than *N. vitripennis* males, which may reduce the re-mating rate of *giraulti* mated females.

REINFORCEMENT

Whether interspecific mating can lead to reinforcement of premating behaviour is still controversial, even though there is both theoretical support (e.g., Butlin 1987, Liou and Price 1994, Servedio and Noor 2003) and empirical evidence for its role in premating isolation (e.g., Blair 1955, Noor 1995, Saetre et al. 1997, Rundle and Schluter 1998, Jaenike et al. 2006). A prediction of the reinforcement hypothesis is that sympatric populations will have higher levels of mate discrimination than allopatric populations of the same taxa (Jaenike et al. 2006). We compared female mate discrimination of *N. vitripennis* strains occurring in microsympatry with *N. giraulti* (from North America) with that of allopatric strains (from Europe). The opposite comparison, involving mate discrimination of allopatric *N. giraulti* females towards *N. vitripennis* males, is not possible, because *N. giraulti* has to date not been found in areas where *N. vitripennis* is absent.

We used a no-choice experiment in which females of four allopatric (European) and four sympatric (North American) *N. vitripennis* strains were tested against males of an *N. giraulti* strain (North American) using crosses to males of their own strain as control. Females of the allopatric *N. vitripennis* strains accepted *N. giraulti* males in 2-26% of cases compared to 0-13% of cases of sympatric strains (Figure 6), which is not significantly different (GLMM: df=1, Chi=1.36, p=0.2441). The control experiments of the conspecific design showed an invariably

high mating rate (near 100%) for both *N. vitripennis* and *N. giraulti*. Although female mate acceptance did not differ between the allopatric and sympatric strains of *N. vitripennis*, a trend exists towards longer duration of courtship bouts that *N. giraulti* males needed to induce receptivity in sympatric females compared to allopatric females (GLMM: $df=1$, $\chi^2=3.47$, $p=0.063$, Figure 7, only courtship bouts that resulted in receptive females were taken into account). This is in line with the prediction of the reinforcement theory. In nature, males may give up courting sooner and pursue another female, since more than one mating partner may be present on a patch. For latency time no significant difference between allopatric and sympatric females was found (GLMM: $df=1$, $\chi^2=2.87$, $p=0.0902$, Figure 7), although there is a trend towards higher latency time in the sympatric design, in line with the reinforcement hypothesis.

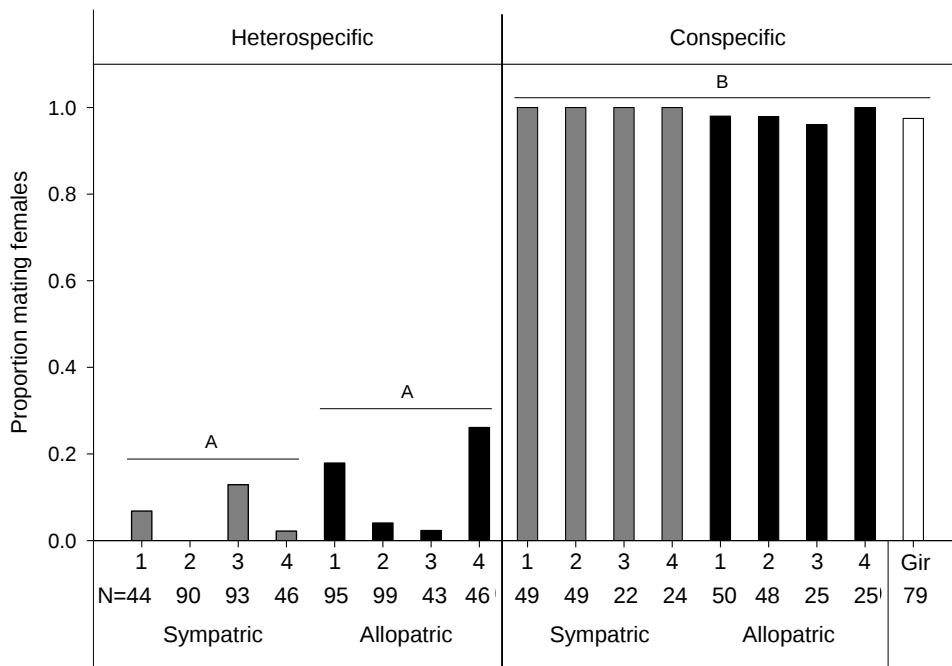


Figure 6. Proportion of *N. vitripennis* females that mated in heterospecific (*N. giraulti* male, left panel) and conspecific (*N. vitripennis* male, right panel) no-choice experiments. Females from sympatric strains are indicated by grey bars; females from allopatric strains by black bars. The right most white bar represents the proportion mating females in a conspecific no-choice experiment with *N. giraulti* males. Sample sizes are shown between brackets on the x-axis. GLMM indicated no significant difference in proportion mating females between allopatric and sympatric strains in both heterospecific and conspecific no-choice experiments ($\chi^2=4.56$, $df=2$, $p=0.102$, significance indicated by capital letters in the plot).

Another prediction of the reinforcement hypothesis is that females that occur sympatrically will evolve reproductive character displacement resulting in stronger discrimination against males of their own species from allopatric populations, as for example found by Jaenike et al. (2006) in two *Drosophila* species. We found no evidence supporting this prediction in our study as all conspecific designs, including those between females and males of different continents yielded high acceptance rates (over 96% successful matings in both directions). Furthermore, there was no significant difference in courtship duration within the conspecific design between

allopatric and sympatric strains (GLMM: $df=1$, $\chi^2=0.44$, $p=0.507$, Figure 7). This indicates that the observed elevated courtship duration of the sympatric heterospecific design is not a strain effect, but may reflect early reinforcement. We conclude that reinforcement has at most slightly strengthened prezygotic isolation between these two species through courtship duration but not female mate acceptance.

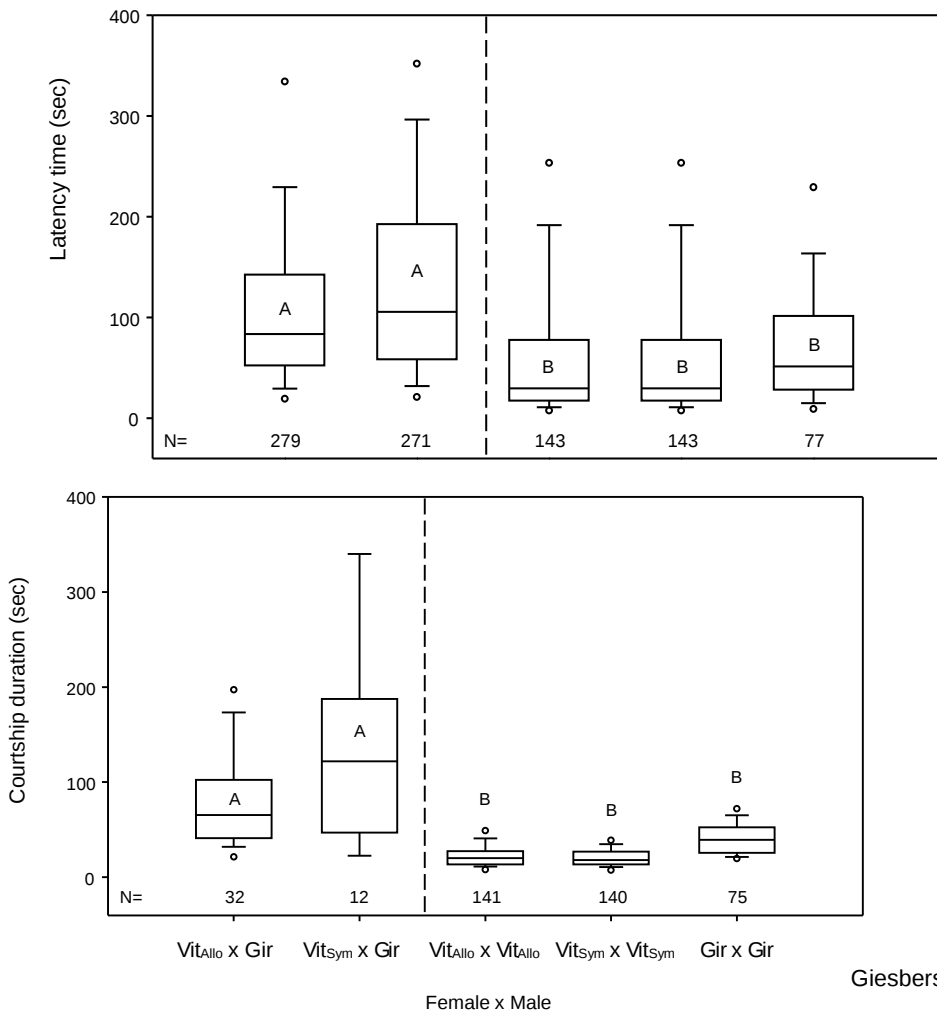


Figure 7. Latency time and courtship duration of five different designs in no-choice experiments. The two box and whiskers in the left panel show results of the heterospecific design with *N. vitripennis* (Vit) females from allopatric (Allo) and sympatric (Sym) strains. The three box and whiskers in the right panel refer to conspecific controls of both species. The two *N. giraulti* (Gir 1 and Gir 2) strains are pooled. There is no difference in latency time between allopatric (Vit_{Allo} x Gir) and sympatric (Vit_{Sym} x Gir) heterospecific design. All conspecific designs differed significantly from the heterospecific designs. A pairwise comparison of courtship duration was significant for the heterospecific design, but not for the conspecific control design. Significance at $\alpha=0.05$ level is indicated by capital letters in the plot. The top and bottom of the boxes give the 25th and 75th percentile; the line in the middle of the box is the median. The ends of the whiskers represent the 2nd and 98th percentile. Outliers are indicated by open circles. Sample sizes are given at the bottom of the plot. Statistical details are given in the text.

Bordenstein et al. (2000) found that *N. longicornis* females discriminate stronger against *N. vitripennis* males than *vice versa*, which they believe is due to an increased exposure to hybridization of *N. longicornis* because its distribution range is embedded within that of *N. vitripennis* (see Figure 2). In addition, *N. vitripennis* is a generalist parasitoid and appears to have higher population densities than *N. longicornis*, which likely results in higher incidences of microsympatry. These authors also found some evidence for reinforcement within *N. vitripennis* as one allopatric strain of Minnesota showed higher receptivity to *N. longicornis* (88% mating) than sympatric females (40-66% mating). As reinforcement in *N. giraulti* may have acted upon the evolution of within-host-mating, the strongest effect of reinforcement on female interspecific mate discrimination may actually be found in *N. oneida*. This species was only recognized as a separate species because of its strong mating discrimination against its sympatric congeners *N. vitripennis* and *N. giraulti*. No investigations of reinforcement have been performed with *N. oneida* as of yet.

One possible reason for not finding strong evidence of reinforcement in *N. vitripennis* is that not enough time has passed since its sympatry was established. Following results of van Opijnen et al. (2005), Raychoudhury et al. (2010b) found evidence from molecular studies that current *N. vitripennis* populations in North America result from a recent (re) immigration event from Europe. Thus, the courtship behaviour of *N. vitripennis* might not yet have completely adapted to encounters with *N. giraulti*. However, Giesbers et al. (in prep) have shown that interspecific mate discrimination of *N. giraulti* can decrease in response to artificial selection within a few generations. Other studies have also shown that traits that are important for mating success, such as courtship song in *D. melanogaster*, have relatively high heritability (Hedrick 1988, Ritchie and Kyriacou 1996). The temporal aspects of selection processes involving mate discrimination may differ considerably between laboratory settings and natural conditions. This requires further investigation of the behavioural interactions in microsympatric *Nasonia* populations. Finally, reinforcement of prezygotic isolation may actually have acted on different aspects of the mating behaviour in *Nasonia* than female mate discrimination, such as within-host-mating (Drapeau and Werren 1999) or chemical communication (see below).

In conclusion, *Nasonia vitripennis* strains that occur in sympatry with *N. giraulti* seem to require longer courtship but otherwise do not show a strong increase in female mate discrimination against *N. giraulti* males. Reinforcement may rather have acted on other aspects of the mating system in *Nasonia*, e.g., within-host-mating. The high aggression of *N. vitripennis* males, the tendency to be the first to mount the female in interspecific contests and the sneaking behaviour of *N. vitripennis* males potentially make heterospecific matings between *N. vitripennis* females and *N. giraulti* males rare in nature.

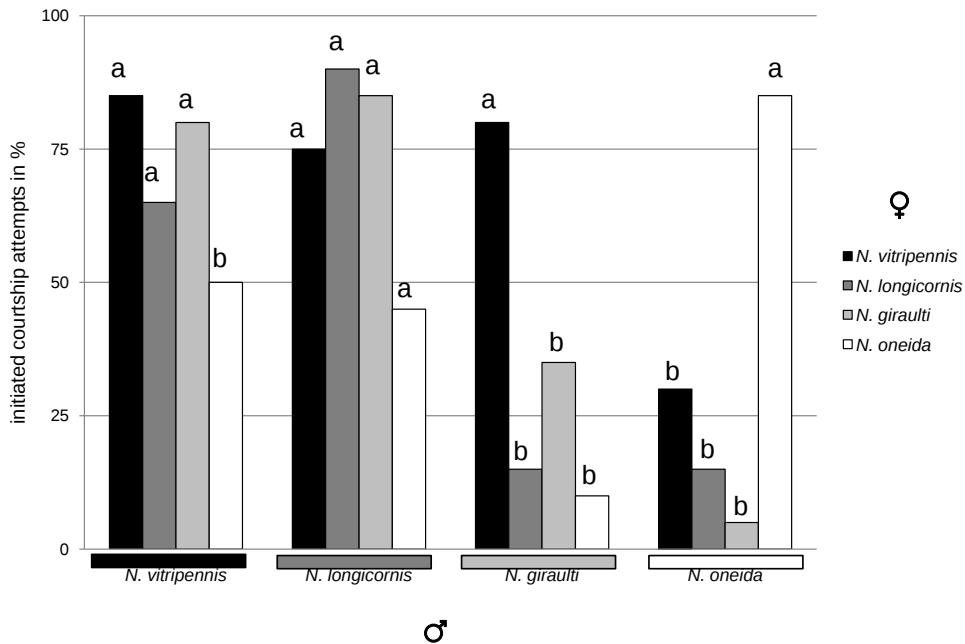
CHEMICAL COMMUNICATION AND THEIR ROLE IN SPECIES DISCRIMINATION

Insects in general have exploited chemical signaling as their primary mode of communication (Greenfield 2002). Two main types of chemical cues and signals are commonly differentiated: long-range and close-range (Thornhill and Alcock 1983, Blomquist and Vogt 2003, Guarino et al. 2008). Concerning the latter type, cuticular hydrocarbons (CHC) are particularly common as close-range cues in a wide variety of insect communication systems

(Blomquist and Bagnères 2010). CHC are the prevalent fraction of the lipid layer on the outer segment of the insect cuticle (Lockey 1985, Howard and Blomquist 2005), and one of their main functions is the prevention of desiccation (Hadley 1981). As for their role in insect communication, CHC have been shown to function as both major sexual (Simmons et al. 2003; Peterson et al. 2007; Oppelt et al. 2009) and interspecific recognition cues (Bagnères et al. 1991, Takahashi and Gassa 1995, Singer 1998). CHC have been exceptionally well studied in the genus *Drosophila* with respect to species-specific CHC variation and CHC-based mate and species discrimination (Buckley et al. 1997, Ferveur 1997, 2005). Concerning CHC studies in *Nasonia*, Carlson et al. (1999) were the first to investigate the CHC profile of *N. vitripennis* and found clear sex-specific differences between adult males and females. Furthermore, Steiner et al. (2006) found experimental evidence that female CHC profiles function as sex pheromones in *N. vitripennis*, initiating courtship behaviour in conspecific males. A comparative CHC study between two *Nasonia* species was done by Raychoudhury et al. (2010a), who established that CHC profiles were divergent enough to separate *N. oneida* from its most closely related sister species, *N. giraulti*, as well as distinguish between their sexes.

Buellesbach et al. (in prep) extended the CHC profile comparison and analyzed qualitative and quantitative CHC differences between all four *Nasonia* species, and the respective sexes in each species. In this study, CHC were found to be sufficiently divergent to unambiguously separate the *Nasonia* genus according to sex and species (Buellesbach et al. submitted). Sex-specificity constitutes a general pre-requisite for CHC to function as cues in sexual communication, while species-specificity hints at their potential role in premating isolation. Pursuing this lead further, Buellesbach et al. (in prep) also performed behavioural assays focussing on the species-specificity of female CHC as mating cues. Female CHC were found to function as species-specific sexual cues for the males in most, but not all, *Nasonia* species.

We performed additional assays presented here to test for the significance of species-specific differences in female CHC profiles for assortative mating mediated by the males. In these assays, freeze-killed females were used as dummies to minimize other cues apart from CHC and those dummies were presented to con- and heterospecific males. We tested whether the males initiated courtship behaviour or not (Figure 8) and whether subsequent copulation attempts took place (Figure 9). The conspecific pairing was used as a reference for the heterospecific pairings. As a control, we assessed male courtship and copulation behaviour on freeze-killed female dummies from which the CHCs had been removed. No courtship or copulation attempts were initiated on these CHC-deprived female dummies at all (data not shown). This clearly hints at the significance of female CHC as sexual cues for males to initiate courtship and/or copulation attempts in *Nasonia*. Concerning courtship behaviour, a striking pattern of discrimination against heterospecific females most likely based on their CHC profile can be seen in *N. oneida* males, who court conspecific female dummies significantly more than all investigated heterospecific female dummies (Figure 8). In contrast, *N. vitripennis* did not discriminate con- from heterospecific female dummies at all in courting (Figure 8). *N. longicornis* only showed pronounced discriminatory behaviour regarding the initiation of courtship against *N. oneida*. Unexpectedly, *N. giraulti* males were apparently more attracted to heterospecific female *N. vitripennis* CHC profiles than to their own conspecific female CHC profile when considering their courtship attempts (Figure 9). One possible explanation for this is that their habit of within-host-mating has not resulted in any differentiation in chemical recognition.

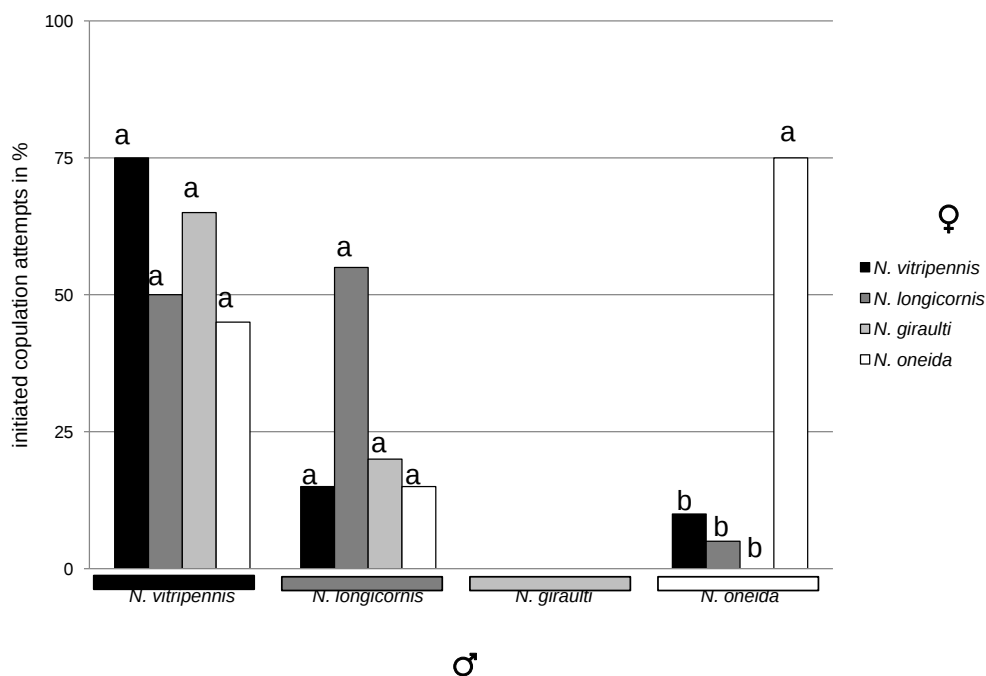


Giesbers et al.

Figure 8. Courtship attempts of males from representative strains of all four *Nasonia* species towards conspecific and heterospecific female dummies in all 16 possible combinations. The number of courtship attempts in the respective conspecific combination was used as a reference against which the numbers of courtship attempts in the heterospecific combinations were tested for each respective male strain. Absolute numbers of courtship attempts for each male strain were tested with Bonferroni-corrected Fisher's exact tests, significant differences ($p < 0.0167$) are indicated by different letters. 20 Trials were conducted in each combination. Strains used are: *N. longicornis* (IV7R2), *N. vitripennis* (NY07/18), *N. giraulti* (NGVA II), *N. oneida* (NONY 11/36).

These differences in attraction were even more pronounced when regarding the copulation attempts (Figure 9). *N. oneida* and *N. vitripennis* generally show the same pattern for copulation attempts as for courtship attempts (see Figure 8) with a pronounced discrimination in the latter and no discrimination at all against heterospecific female dummies in the former case, respectively. *N. longicornis* males show no significant differences in number of copulation attempts against heterospecific vs. conspecific female dummies, although a trend towards more copulation attempts with conspecific female dummies is apparent (see Figure 9). *N. giraulti* males constitute the most curious case, since they did not attempt to copulate with the freeze-killed female dummies at all, independent of whether they were con- or heterospecific (Figure 9). This indicates that female CHC profiles are not sufficient as sexual cues for *N. giraulti* males to initiate copulation.

Overall, our experiments indicate that CHC profiles are important as sexual cues for males of all investigated *Nasonia* species. However, their species-specificity in initiating both courtship and copulation attempts by the males greatly varies according to the investigated species pairs, from clear male discriminatory behaviour against heterospecific female profiles (as displayed in *N. oneida*) to none at all (as displayed in *N. vitripennis*). *N. giraulti* males apparently require additional cues to discriminate con- from heterospecific females for both initiating courtship and copulation attempts on female dummies.



Giesbers et al.

Figure 9. Copulation attempts of males from representative strains of all four *Nasonia* species towards conspecific and heterospecific female dummies in all 16 possible combinations. The number of copulation attempts in the respective conspecific combination was used as a reference against which the numbers of copulation attempts in the heterospecific combinations were tested for each respective male strain. Absolute numbers of copulation attempts for each male strain were tested with Bonferroni-corrected Fisher's exact tests, significant differences ($p < 0.0167$) are indicated by different letters. 20 Trials were conducted in each combination. Strains used are: *N. longicornis* (IV7R2), *N. vitripennis* (NY07/18), *N. giraulti* (NGVA II), *N. oneida* (NONY 11/36).

In conclusion, clear indications were found for a role of female CHC profiles in *Nasonia* as sexual cues potentially involved in prezygotic isolation, but further study is needed to shed more light on both the exact mechanism of female-mediated CHC signalling as well as the particular CHC compounds functioning as the major cues for the males. Taking a different angle to investigate the basis for species-specific CHC variation in *Nasonia*, Niehuis et al. (2011) investigated the genetic background of CHC differences between *N. vitripennis* and *N. giraulti*. Using hybrid males and taking advantage of their haploidy, they found a large number of QTL (>100) governing species-specific CHC variation, distributed over all five chromosomes. The results revealed interesting associations to orthologous genes known to be involved in CHC biosynthesis in *Drosophila* as well as QTL hinting at hitherto unknown genes potentially involved in the biosynthesis of CHC compound classes (Niehuis et al. 2011). It will be an important next step to further analyze and unravel this unexpectedly complicated genetic architecture of CHC variation between the *Nasonia* species. This should ultimately lead to a better understanding of how CHC divergence is genetically maintained and biosynthetically governed.

Concerning non-CHC based, long-range chemical communication, Ruther et al. (2007, 2008) were the first to isolate a male specific sex pheromone from *Nasonia vitripennis*. Males

produce a mixture of (4R, 5R)- and (4R, 5S)-5-hydroxy-4-decanolide (HDL) from glands in their abdomen which they deposit by abdomen dipping to attract virgin females. The pheromone can be classified as long-range as females perceive it at distances of up to 5 centimeters and it remains attractive for up to 2 hours (Steiner and Ruther 2009). HDL is only attractive to virgin females; after copulation females immediately become insensitive to it (Ruther et al. 2007). In a later study using sperm depleted males, Ruther et al. (2009) could also show that virgin females' olfactory responses were positively correlated with pheromone dose, proving the mate assessment function of HDL.

In addition to the HDL pheromone, males also release a short range aphrodisiac from their mandibular glands, which appears to be deposited directly onto the female antennae during their headnods in the courtship sequence. Van den Assem et al. (1980) were the first to show the existence of this aphrodisiac by sealing the male mouthparts with glue, resulting in females that did not become receptive upon being courted by manipulated males. Ruther et al. (2011) used the same technique to show that virgin females do not lose their attraction to HDL if they have not received this short range aphrodisiac, further supporting its role in inducing female receptivity. However, the exact chemical composition of this aphrodisiac remains elusive as it appears to be difficult to isolate and analyse, and decades of intensive research did not get any closer to its chemical detection and characterization so far (Ruther, Schmitt and Werren, pers. comm.)

CONCLUSION

We have presented and discussed the current knowledge of prezygotic hybridization barriers in the *Nasonia* species complex. Although it is evident that the strengths of these barriers vary, depending on the species pair considered, it is much harder to determine how these differences potentially have evolved. It is likely that most differences came into being after postmating reproductive isolation was established. The reason is that species-specific infections with *Wolbachia* bacteria causing cytoplasmic incompatibility between the different species are considered as the most likely cause of the speciation events in the *Nasonia* genus (Bordenstein et al. 2001). *N. oneida* appears to be a clear example of sympatric speciation as it originated within the distribution range of *N. vitripennis* and *N. giraulti*. Some geographic differentiation in mate discrimination within species is apparent in *N. vitripennis*, although lack of such evidence from the other species may be due to a lower intensity of investigation.

A conspicuous characteristic of the *Nasonia* genus is its strong differences in degrees of prezygotic isolation between species. These differences are often asymmetric in that females of one species are either more or less discriminative against males of the other species. The degrees of premating isolation are generally consistent with the divergence time of the species. However, while *N. oneida* is genetically very similar to *N. giraulti*, females show very strong mating discrimination against *N. giraulti*, which indicates a rapid evolution of mate discrimination behaviour. Within-host-mating may constitute another mechanism to prevent interspecific matings in nature. Consistent with this notion, *N. giraulti* exhibits high levels of within-host-mating in combination with low levels of interspecific mate discrimination.

Chemical communication is likely another important factor in prezygotic isolation between the *Nasonia* species. Chemical signals have been shown to play important roles in the *Nasonia*

mating patterns. Sexual cues and/or signals act on at least three different levels during the *Nasonia* mating procedure, i.e., the male-produced long-range pheromone, female CHC and the male mouthpart aphrodisiac. There is clear evidence from behavioural assays that at least one of these chemical cues or signals are not only involved in intraspecific sexual selection, but also play a role in interspecific sexual isolation. Differences in male courtship displays, and possibly courtship song, likely play a modulating role.

Differentiation in mate choice among strains within species suggests that considerable levels of genetic variation are present in natural populations. This notion is supported by high heritability of mate discrimination behaviour and a fast response to selection in the laboratory. There is, however, as yet no strong evidence for a role of reinforcement in the *Nasonia* species complex (Bordenstein et al. 2000, this chapter). Several explanations can be proposed, such as that the species have diverged too recently for reinforcement to become established, or that other behaviours or traits may have been selected to prevent hybridization (e.g., within host mating, host choice). Although there is considerable overlap in the distribution ranges of the four species in North America, where species can even occur microsympatrically (e.g., in the same fly host), we still know little about behavioural interactions under natural conditions. Nevertheless, the tractability and genomic knowledge of the *Nasonia* genus makes it a promising system for further study of the evolution of prezygotic isolation and speciation.

ACKNOWLEDGMENTS

We are grateful to Rienk Apperloo, Jan Geijer, Bernd Grillenberger, Sjoerd Hobma, Ido Pen, Saleta Perez Vila, Joop Prijs, Corinna Bächle and Franziska Kurz who all contributed to the data presented in this chapter. This work was funded by grants ALW 833.02.003 and ALW 817.01.011 of the Netherlands Organisation for Scientific Research to LWB and the Excellence Initiative of the German Research Foundation (GSC-4, Spemann Graduate School) to JB and TS.

REFERENCES

- Arbuthnott D. 2009. The genetic architecture of insect courtship behavior and premating isolation. *Heredity*. 103:15-22.
- Bagnères A. G., Killian A., Clement J. L., Lange C. 1991. Interspecific recognition among termites of the genus *reticulitermes*: Evidence for a role for the cuticular hydrocarbons. *J Chem Ecol*. 17:2397-2420.
- Barrass R. 1960. The effect of age on the performance of an innate behaviour pattern in *Mormoniella vitripennis*. *Behaviour* 15:210-218.
- Barrass R. 1976. Inhibitory effects of courtship in the wasp *Nasonia vitripennis* and a new interpretation of the biological significance of courtship in insects. *Physiol Entomol*. 1:229-234.
- Beukeboom L., Desplan C. 2003. *Nasonia*. *Curr Biol*. 13:860.

- Beukeboom L. W., van den Assem J. 2001. Courtship and mating behaviour of interspecific *Nasonia* hybrids (hymenoptera, pteromalidae): A grandfather effect. *Behav Genet.* 31:167-177.
- Beukeboom L. W., van den Assem J. 2002. Courtship displays of introgressed, interspecific *Nasonia* males: Further investigations into the Grandfather effect. *Behaviour* 139:1029-1042.
- Blair W. F. 1955. Mating call and stage of speciation in the *Microhylla olivacea*-*M. carolinensis* complex. *Evolution* 9:469-480.
- Blomquist G. J., Bagnères A. G. 2010. *Insect Hydrocarbons: Biology, Biochemistry, and Chemical Ecology*. New York, USA: Cambridge University Press.
- Blomquist G. J., Vogt R. G. 2003. *Insect Pheromone Biochemistry and Molecular Biology*. London: Elsevier Academic.
- Bordenstein S. R., Drapeau M. D., Werren J. H. 2000. Intraspecific variation in sexual isolation in the jewel wasp *Nasonia*. *Evolution* 54:567-573.
- Bordenstein S. R., O'Hara F. P., Werren J. H. 2001. *Wolbachia*-induced incompatibility precedes other hybrid incompatibilities in *Nasonia*. *Nature* 409:707-710.
- Bordenstein S. R., Uy J. J., Werren J. H. 2003. Host genotype determines cytoplasmic incompatibility type in the haplodiploid genus *Nasonia*. *Genetics* 164:223-233.
- Bordenstein S. R., Werren J. H. 1998. Effects of A and B *Wolbachia* and host genotype on interspecies cytoplasmic incompatibility in *Nasonia*. *Genetics* 148:1833-1844.
- Bordenstein S. R., Werren J. H. 2007. Bidirectional incompatibility among divergent *Wolbachia* and incompatibility level differences among closely related *Wolbachia* in *Nasonia*. *Heredity* 99:278-287.
- Breeuwer J. A., Werren J. H. 1990. Microorganisms associated with chromosome destruction and reproductive isolation between two insect species. *Nature* 346:558-560.
- Breeuwer J. A. J., Werren J. H. 1995. Hybrid breakdown between two haplodiploid species: The role of nuclear and cytoplasmic genes. *Evolution* 49:705-717.
- Brown W. L., Wilson E. O. 1956. Character displacement. *Syst Zool.* 7:49-64.
- Buckley S. H., Tregenza T., Butlin R. K. 1997. Speciation and signal trait genetics. *Trends Ecol Evol.* 12:299-301.
- Burton-Chellew M. N., Sykes E. M., Shuker D. M., West S. A. 2007. The cost of mating and the relationship between body size and fitness in males of the parasitoid wasp *Nasonia vitripennis*. *Evol Ecol Res.* 9:1-14.
- Butlin R. 1987. Speciation by reinforcement. *Trends Ecol Evol.* 2:8-13.
- Campbell H. D., Schimansky T., Claudianos C., Ozsarac N., Kasprzak A. B., Cotsell J. N., Young I. G., de Couet H. G., Miklos G. L. 1993. The *Drosophila melanogaster* flightless-I gene involved in gastrulation and muscle degeneration encodes gelsolin-like and leucine-rich repeat domains and is conserved in *Caenorhabditis elegans* and humans. *Proc Natl Acad Sci USA.* 90:11386-11390.
- Carlson D. A., Geden C. J., Bernier U. R. 1999. Identification of pupal exuviae of *Nasonia vitripennis* and *Muscidifurax raptorellus* parasitoids using cuticular hydrocarbons. *Biol Control* 15:97-105.
- Charnoy E. L., Skinner S. W. 1984. Evolution of host selection and clutch size in parasitoid wasps. *Florida Entomol.* 67:5-21.
- Clark M. E., O'Hara F. P., Chawla A., Werren J. H. 2010. Behavioral and spermatogenic hybrid male breakdown in *Nasonia*. *Heredity* 104: 289-301.

- Coyne J. A., Kim S. Y., Chang A. S., Lachaise D., Elwyn S. 2002. Sexual isolation between two sibling species with overlapping ranges: *Drosophila santomea* and *Drosophila yakuba*. *Evolution* 56:2424-2434.
- Coyne J. A., Orr H. A. 1997. Patterns of speciation in *Drosophila*, revisited. *Evolution* 51:295-303.
- Coyne J. A., Orr H. A. 1998. The evolutionary genetics of speciation. *Phil Trans R Soc Lond B Biol Sci.* 353:287-305.
- Coyne J. A., Orr H. A. 2004. *Speciation*. Sunderland, MA: Sinauer Associates Inc.
- da Costa Lima. 1928. Notas sobre a biologia do *Telenomus fariai* Lima, parasito dos ovos de *Triatoma*. *Memorias do Instituto Oswaldo Cruz* [Notes on the biology of *Telenomus fariai* Lima, parasite of *Triatoma* eggs. *Memories of the Oswaldo Cruz Institute*]. 21: 201-217.
- Darling D. C., Werren J. H. 1990. Biosystematics of *Nasonia* (hymenoptera: Pteromalidae): Two new species reared from birds' nests in North America. *Ann Entom Soc Am.* 83:352-370.
- Dobzhansky T. 1935. Maternal effect as a cause of the difference between the reciprocal crosses in *Drosophila pseudoobscura*. *Proc Natl Acad Sci USA.* 21:443-446.
- Dobzhansky T. 1940. Speciation as a stage in evolutionary divergence. *Amer Nat.* 74:312-321.
- Drapeau M., Werren J. H. 1999. Differences in mating behavior and sex ratio between three sibling species of *Nasonia*. *Evol Ecol Res.* 1:223-234.
- Ferveur J. F. 1997. The pheromonal role of cuticular hydrocarbons in *Drosophila melanogaster*. *Bioessays* 19:353-358.
- Ferveur J. F. 2005. Cuticular hydrocarbons: Their evolution and roles in *Drosophila* pheromonal communication. *Behav Genet.* 35:279-295.
- Gadau J., Page R. E., Werren J. H. 1999. Mapping of hybrid incompatibility loci in *Nasonia*. *Genetics* 153:1731-1741.
- Geuverink E., Gerritsma S., Pannebakker B. A., Beukeboom L. W. 2009. A role for sexual conflict in the evolution of reproductive traits in *Nasonia* wasps? *Animal Biol.* 59:417-434.
- Gow J. L., Peichel C. L., Taylor E. B. 2006. Contrasting hybridization rates between sympatric three-spined sticklebacks highlight the fragility of reproductive barriers between evolutionarily young species. *Mol Ecol.* 15: 739-752.
- Grant P. 1972. Convergent and divergent character displacement. *Biol J Linn Soc.* 4:39-68.
- Greenfield M. D. 2002. *Signalers and Receivers: Mechanisms and Evolution of Arthropod Communication*. New York, USA: Oxford University Press.
- Grillenberger B. K., Gadau J., Bijlsma R., van de Zande L., Beukeboom L. W. 2009a. Female dispersal and isolation-by-distance of *Nasonia vitripennis* populations in a local mate competition context. *Entomol Exp Appl.* 132: 147-154.
- Grillenberger B. K., Koevoets T., Burton-Chellew M. N., Sykes E. M., Shuker D. M., van de Zande L., Bijlsma R., Gadau J., Beukeboom L. W. 2008. Genetic structure of natural *Nasonia vitripennis* populations: Validating assumptions of sex-ratio theory. *Mol Ecol.* 17:2854-2864.
- Grillenberger B. K., van de Zande L., Bijlsma R., Gadau J., Beukeboom L. W. 2009b. Reproductive strategies under multiparasitism in natural populations of the parasitoid wasp *Nasonia* (hymenoptera). *J Evol Biol.* 22:460-470.

- Guarino S., de Pasquale C., Peri E., Alonzo G., Colazza S. 2008. Role of volatile and contact pheromones in the mating behavior of *Bagrada hilaris* (Heteroptera: Pentatomidae). *Eur J Entomol.* 105:613-617.
- Hadley N. F. 1981. Cuticular lipids of terrestrial plants and arthropods: A comparison of their structure, composition, and waterproofing function. *Biol Rev.* 56:23-47.
- Hamilton W. D. 1967. Extraordinary sex ratios. A sex-ratio theory for sex linkage and inbreeding has new implications in cytogenetics and entomology. *Science* 156:477-488.
- Hedrick A. V. 1988. Female choice and the heritability of attractive male traits: An empirical study. *Am Nat.* 132:267-276.
- Holmes H. B. 1972. Genetic evidence for fewer progeny and higher percent males when *Nasonia vitripennis* oviposits in previously parasitized hosts. *Entomophaga* 17:79-88.
- Howard R. W., Blomquist G. J. 2005. Ecological, behavioral, and biochemical aspects of insect hydrocarbons. *Annu Rev Entomol.* 50:371-393.
- Jachmann F., van den Assem J. 1993. The interaction of external and internal factors in the courtship of parasitic wasps (Hym., Pteromalidae). *Behaviour* 125:1-19.
- Jachmann F., van den Assem J. 1996. A causal ethological analysis of the courtship behaviour of an insect (the parasitic wasp *Nasonia vitripennis*, Hym., Pteromalidae). *Behaviour* 133:1051-1075.
- Jaenike J., Dyer K. A., Cornish C., Minhas M. S. 2006. Asymmetrical reinforcement and *Wolbachia* infection in *Drosophila*. *PLoS Biol.* 4:e325.
- King B. H., Skinner S. W. 1991. Proximal mechanisms of the sex-ratio and clutch size responses of the wasp *Nasonia vitripennis* to parasitized hosts. *Anim Behav.* 42:23-32.
- Koevoets T., Niehuis O., van de Zande L., Beukeboom L. W. 2012. Hybrid incompatibilities in the parasitic wasp genus *Nasonia*: Negative effects of hemizygosity and the identification of transmission ratio distortion loci. *Heredity* 108:302-311.
- Kondrashov A. S., Shpak M. 1998. On the origin of species by means of assortative mating. *Proc Biol Sci.* 265:2273-2278.
- Kyriacou C. P., Hall J. C. 1982. The function of courtship song rhythms in *Drosophila*. *Anim Behav* 30:794-801.
- Leonard J. E., Boake C. R. B. 2006. Site-dependent aggression and mating behaviour in three species of *Nasonia* (hymenoptera: Pteromalidae). *Anim Behav.* 71:641-647.
- Li Y.-F., Wen S.-Y., Ritchie M. G. 2012. Copulatory song in three species of the *Drosophilamontium* subgroup extends copulation and shows unusual genetic control. *Anim Behav.* 83:233-238.
- Liou L. W., Price T. D. 1994. Speciation by reinforcement of premating isolation. *Evolution* 48:1451-1459.
- Mallet J. 1998. Tropical ecology in miniature. *Trends Ecol Evol.* 13:377.
- Mayr E. 1942. *Systematics and the Origin of Species*. New York: Columbia Univ. Press.
- Mayr E. 1963. *Animal Species and Evolution*. Cambridge, MA: Harvard Univ. Press.
- Niehuis O., Bullesbach J., Judson A. K., Schmitt T., Gadau J. 2011. Genetics of cuticular hydrocarbon differences between males of the parasitoid wasps *Nasonia giraulti* and *Nasonia vitripennis*. *Heredity* 107:61-70.
- Niehuis O., Judson A. K., Gadau J. 2008. Cytonuclear genic incompatibilities cause increased mortality in male F2 hybrids of *Nasonia giraulti* and *N. vitripennis*. *Genetics* 178:413-426.
- Noor M. A. 1995. Reinforcement in speciation. *Trends Ecol Evol.* 10:492.

- Oppelt A., Heinze J. 2009. Mating is associated with immediate changes of the hydrocarbon profile of *Leptothorax gredleri* ant queens. *J Insect Physiol.* 55:624-628.
- Orr H. A., Masly J. P., Presgraves D. C. 2004. Speciation genes. *Curr Opin Genet Dev.* 14: 675-679.
- Peterson M. A., Dobler S., Larson E. L., Juarez D., Schlarbaum T., Monsen K. J., Francke W. 2007. Profiles of cuticular hydrocarbons mediate male mate choice and sexual isolation between hybridising *Chrysochus* (coleoptera: Chrysomelidae). *Chemoecology* 17:87-96.
- Pultz M. A., Leaf D. S. 2003. The jewel wasp *Nasonia*: Querying the genome with haplo-diploid genetics. *Genesis* 35:185-191.
- Raychoudhury R., Desjardins C. A., Buellesbach J., Loehlin D. W., Grillenberger B. K., Beukeboom L., Schmitt T., Werren J. H. 2010a. Behavioral and genetic characteristics of a new species of *Nasonia*. *Heredity* 104:278-288.
- Raychoudhury R., Grillenberger B. K., Gadau J., Bijlsma R., van de Zande L., Werren J. H., Beukeboom L. W. 2010b. Phylogeography of *Nasonia vitripennis* (Hymenoptera) indicates a mitochondrial-*Wolbachia* sweep in North America. *Heredity* 104:318-326.
- Ritchie M., G., Kyriacou C. P. 1996. Artificial selection for a courtship signal in *Drosophila melanogaster*. *Anim Behav.* 52:603-611.
- Ritchie M. G., Gleason J. M. 1995. Rapid evolution of courtship song pattern in *Drosophila willistoni* sibling species. *J Evol Biol.* 8:463-479.
- Robertson H. M., Gadau J., Wanner K. W. 2010. The insect chemoreceptor superfamily of the parasitoid jewel wasp *Nasonia vitripennis*. *Insect Mol Biol.* 19 Suppl 1:121-136.
- Rundle H. D., Schluter D. 1998. Reinforcement of stickleback mate preferences: Sympatry breeds contempt. *Evolution* 52:200-208.
- Ruther J., Matschke M., Garbe L. A., Steiner S. 2009. Quantity matters: Male sex pheromone signals mate quality in the parasitic wasp *Nasonia vitripennis*. *Proc Biol Sci.* 276:3303-3310.
- Ruther J., Stahl L. M., Steiner S., Garbe L. A., Tolasch T. 2007. A male sex pheromone in a parasitic wasp and control of the behavioral response by the female's mating status. *J Exp Biol.* 210:2163-2169.
- Ruther J., Steiner S., Garbe L. A. 2008. 4-methylquinazoline is a minor component of the male sex pheromone in *Nasonia vitripennis*. *J Chem Ecol.* 34:99-102.
- Ruther J., Thal K., Steiner S. 2011. Pheromone communication in *Nasonia vitripennis*: Abdominal sex attractant mediates site fidelity of releasing males. *J Chem Ecol.* 37:161-165.
- Saetre G., Moum T., Bures S., Kral M., Adamjan M., Moreno J. 1997. A sexually selected character displacement in flycatchers reinforces premating isolation. *Nature* 387:589-592.
- Servedio M., Noor M. 2003. The role of reinforcement in speciation: Theory and data. *Annu Rev Ecol Evol Syst.* 34:339-364.
- Shaw A., Lee Y., Stout D., Vacquier V. D. 1994. The species-specificity and structure of abalone sperm lysin. *Semin Dev Biol.* 5:209-215.
- Simmons L. W., Alcock J., Reeder A. 2003. The role of cuticular hydrocarbons in male attraction and repulsion by female dawson's burrowing bee, *Amegilla dawsoni*. *Anim Behav.* 66:677-685.
- Singer T. L. 1998. Roles of hydrocarbons in the recognition systems of insects. *Am Zool.* 38:394-405.

- Steiner S., Hermann N., Ruther J. 2006. Characterization of a female-produced courtship pheromone in the parasitoid *Nasonia vitripennis*. *J Chem Ecol.* 32:1687-1702.
- Steiner S., Ruther J. 2009. Mechanism and behavioral context of male sex pheromone release in *Nasonia vitripennis*. *J Chem Ecol.* 35:416-421.
- Suzuki Y., Hiehata K. 1985. Mating systems and sex ratios in the egg parasitoids, *Trichogramma dendrolini* and *T. papilionis* (Hymenoptera: Trichogrammatidae). *Anim Behav.* 33:1223-1227.
- Takahashi S., Gassa A. 1995. Roles of cuticular hydrocarbons in intra- and interspecific recognition behavior of two rhinotermitidae species. *J Chem Ecol.* 21:1837-1845.
- Thornhill R., Alcock J. 1983. *The Evolution of Insect Mating Systems*. Cambridge, USA: Harvard University Press.
- Tram U., Fredrick K., Werren J. H., Sullivan W. 2006. Paternal chromosome segregation during the first mitotic division determines *Wolbachia*-induced cytoplasmic incompatibility phenotype. *J Cell Sci.* 119:3655-3663.
- Tram U., Sullivan W. 2002. Role of delayed nuclear envelope breakdown and mitosis in *Wolbachia*-induced cytoplasmic incompatibility. *Science* 296: 1124-1126.
- van den Assem J., Gijswijt M. J., Nübel B. K. 1980. Observations on courtship - and mating strategies in a few species of parasitic wasps (Chalcidoidea). *Neth J Zool.* 30:208-227.
- van den Assem J., Jachmann F. 1999. Changes in male perseverance in courtship and female readiness to mate in a strain of the parasitic wasp *Nasonia vitripennis* over a period of 20+ years. *Neth J Zool.* 49: 125-137.
- van den Assem J., Visser J. 1976. Aspects of sexual receptivity in female *Nasonia vitripennis* (Hym., Pteromalidae). *Biol Behav.* 1:37-56.
- van den Assem J., Werren J. H. 1994. A comparison of the courtship and mating behavior of three species of *Nasonia* (Hymenoptera: Pteromalidae). *J Insect Behav.* 7:53-65.
- van den Assem J. 1986. Mating behaviour in parasitic wasps. In: Waage J, Greathead D, editors. *Insect Parasitoids*. London: Academic Press. p. 137-167.
- van Opijnen T., Baudry E., Baldo L., Bartos J., Werren J. H. 2005. Genetic variability in the three genomes of *Nasonia*: Nuclear, mitochondrial and *Wolbachia*. *Insect Mol Biol.* 14:653-663.
- Velthuis B. J., Yang W. C., van Opijnen T., Werren J. H. 2005. Genetics of female mate discrimination of heterospecific males in *Nasonia* (Hymenoptera, Pteromalidae). *Anim Behav.* 69:1107-1120.
- Werren J. H., Loehlin D. W. 2009. The parasitoid wasp *Nasonia*: An emerging model system with haploid male genetics. *Cold Spring Harb Protoc.* 10.
- Werren J. H., Richards S., Desjardins C. A. et al. (153 co-authors). 2010. Functional and evolutionary insights from the genomes of three parasitoid *Nasonia* species. *Science* 327:343-348.
- Werren J. H. 1980. Sex ratio adaptations to local mate competition in a parasitic wasp. *Science* 208:1157-1159.
- Werren J. H. 1983. Sex ratio evolution under local mate competition in a parasitic wasp. *Evolution* 37:116-124.
- Werren J. H. 1984. Brood size and sex-ratio regulation in the parasitic wasp *Nasonia vitripennis* (Walker) (Hymenoptera: Pteromalidae). *Neth J Zool.* 34:123-143.
- Whiting A. R. 1967. The biology of the parasitic wasp *Mormoniella vitripennis* [= *Nasonia brevicornis*] (walker). *Quart Rev Biol.* 42:333-406.

Chapter 8

**WHERE TO LOOK FOR SPECIATION
GENES WHEN DIVERGENCE IS DRIVEN
BY POSTMATING, PREZYGOTIC ISOLATION**

Jeremy L. Marshall*

Department of Entomology, Kansas State University, Manhattan, KS, US

ABSTRACT

Not all genes are equally important during the process of speciation. This premise underlies a basic question in evolutionary biology – “Does divergence at any one gene, or set of genes, play a particularly important role in speciation?” The answer to this question may appear to be no for some forms of reproductive isolation, but there may indeed be ‘kinds of genes’ that routinely play a role in speciation when divergence is driven by postmating, prezygotic traits. Here, I outline the kinds of molecular pathways and interactions that underlie postmating, prezygotic phenotypes which ultimately points to the kinds of genes where we should look for species-specificity. Interestingly, it is only when we consider the entire system of interacting sex proteomes, cell structure, membrane dynamics, and physiological pathways that a picture of where species-specific interactions likely occur becomes clear. While this approach points to several kinds of pathways and gene types, a notable finding is that cell membrane receptors (like G-protein coupled receptors and receptor tyrosine kinases) that line the inside of the female reproductive tract and trigger post-copulation cell signaling should be considered the kinds of genes that routinely contribute to reproductive isolation and speciation when divergence is driven by postmating, prezygotic phenotypes.

* Corresponding Author’s Email: cricket@ksu.edu (Jeremy L. Marshall, Department of Entomology, Kansas State University, Manhattan, KS 66506. Tel: (785) 532-5588).

INTRODUCTION

Sexual reproduction is widespread in animals and often involves internal insemination. This leads to the direct interaction of male sperm and ejaculate proteins with proteins from the female reproductive tract. Consequently, these interacting proteomes shape reproductive physiology and determine fertilization success – ultimately, influencing fitness.

Given that reproductive proteins play such a fundamental role in fitness, it is not surprising that they have been the focus of so much study. Over the last few decades, it has become clear that many reproductive proteins in both males and females evolve rapidly (e.g., Swanson et al. 2001, 2004; Swanson and Vacquier 2002a,b; Clark et al. 2006; Ramm et al. 2009; Dorus et al. 2010; Marshall et al. 2011). Interestingly, over the same time period, evolutionary biologists have demonstrated the importance of postmating, prezygotic phenotypes in isolating closely related species (reviewed in Howard 1999; more recent studies include Brown and Eady 2001; Simmons 2001; Marshall 2004; Geyer and Palumbi 2005; Dean and Nachman 2009; Levesque et al. 2010; Sweigart 2010). Taken together, these findings lead to the hypothesis that the rapid evolution of male and female reproductive proteins, along with their interactions, underlies much of the postmating, prezygotic reproductive isolation seen in nature (although few studies have directly tested this hypothesis; Turner and Hoekstra 2008).

As with all forms of reproductive isolation, there is a search for generalities in the processes and genetic mechanisms that underlie their function. Over the many decades of searching for such patterns, several important questions have arisen, including, “Does divergence at any one gene, or set of genes, play a particularly important role in speciation?” For postmating, prezygotic reproductive isolation, we know the genes of interest are some subset of male and female reproductive genes. But several questions remain, including which male and female reproductive genes are linked to reproductive isolation, are the important genes all the genes that have been shown to be under positive selection, and finally, are there kinds or sets of genes that are expected to routinely contribute to reproductive isolation? Empirical evidence will ultimately answer these questions; however, is there an *a priori* expectation of which reproductive genes are more likely to contribute to reproductive isolation and speciation? The objective of this article is to combine what we know about interacting sex proteomes, cell structure, membrane dynamics, and physiological pathways to predict the kinds of genes where we should look for species-specificity. In the end, these should not only be the types of genes that exhibit species-specific variation and are driven by positive selection, but that control postmating, prezygotic phenotypes which isolate closely-related species.

A SYSTEMS APPROACH TO POSTMATING, PREZYGOTIC PHENOTYPES

What System Underlies Postmating, Prezygotic Phenotypes?

The genomic and post-genomic era has provided new insights into the mechanisms underlying the evolution of reproductive isolation (e.g., Noor and Feder 2006; Kulathinal et al. 2009; Butlin 2010; Presgraves 2010). However, more than any one answer, what these studies have done is drive home the point that variation within genetic pathways underlies phenotypic variation – in essence, genes only make sense in light of the other genes they interact with

[apologies to Dobzhansky (1964) for borrowing his phrasing]. Countless transcriptomic studies have shown that differences in gene expression are associated with phenotypic variation (reviewed in Huestis and Marshall 2009) and RNAi/mutant allele studies have shown that knocking down one gene can dramatically alter patterns of gene and phenotypic expression (e.g., Finkel and Holbrook 2000; Bonaldi et al. 2008). So, if genetic pathways are important, then what does this say about how we should study the genetics of speciation?

To begin, it would seem we must first consider how genes interact. The answer to “how do genes interact” lies, in large part, in the products of genes, i.e., proteins, and more specifically protein-protein interactions. These interactions take place within, between, and outside of cells and can involve proteins from the same cell, different tissues or even different individuals (e.g., the interaction of male ejaculates and female reproductive tracts). A critical component of protein-protein interactions is posttranslational modification (e.g., phosphorylation, methylation, and glycosylation, among about 200 others; reviewed in Jensen 2006) as they can activate protein function, modify function, and even serve as master switches that turn on genetic pathways.

If protein-protein interactions and posttranslational modifications within pathways are important players in the genetics of reproductive isolation, then how would this fact influence how we think about the long-standing questions of the number, kind, and variation of genes driving the evolution of reproductive isolation? In thinking about this, we must recognize that pathways have physical structure that guides their operation (Alberts et al. 2007). Specifically, there is cell structure that dictates that molecules from other cells, tissues, or individuals must interact with or pass through the plasma membrane to effect change within the cell (Hancock 2005). This physical structure has implications for the genetics of reproductive isolation and thus, on these long-standing questions (e.g., Dorus et al. 2010).

The implications of this physical structure are particularly relevant for forms of postmating, prezygotic isolation – as much of this type of reproductive isolation boils down to the dynamics of signals and receivers (Endler and Basolo 1998; Boughman 2002; Sultan 2007). Whether it is sperm-eggs interactions or ejaculate-female reproductive tract interactions, the molecular mechanism underlying the signal-receiver system will shape the evolution of reproductive isolation. At the molecular level, these signal-receiver systems are controlled by ligands and receptors, with receptors (i.e., the receiver) occurring on the plasma membrane and ligands (i.e., the signal) being the extracellular molecules from other cells or individuals that bind to receptors and cause the activation of one or more genetic pathways (usually through phosphorylation; Hancock 2005). Given this, if we consider the question, “does divergence at any one gene, or set of genes, play a particularly important role in speciation?”, then we would predict an important role for receptor and ligand genes that activate genetic pathways when speciation is driven by postmating, prezygotic barriers. Furthermore, receptor-ligand systems also predict the number, effect size, and variation of genes that should underlie these types of postmating, prezygotic barriers – that is, such barriers should be controlled by a few genes (those of receptors and ligands) of large effect that exhibit structural variation.

Cell Signaling and Phosphorylation as Mechanisms Underlying Postmating, Prezygotic Phenotypes

The postmating, prezygotic phenotype that isolates a pair of species can occur at any point following successful copulation. Given this, the key question becomes at what point does a heterospecific ejaculate fail in the female reproductive tract? There are numerous reasons why such an ejaculate might fail, including being expelled from the female reproductive tract (e.g., Otronen and Siva-Jothy 1991), being attacked by the female immune system (Johansson et al. 2004), or not fully activating post-copulation physiologies (Neubaum and Wolfner 1999; McGraw et al. 2004; Yang et al. 2009). Indeed, ample research has shown that the ejaculate proteome of a male interacts with the reproductive-tract proteome of a female to initiate a wide range of post-copulation physiologies (e.g., Neubaum and Wolfner 1999; Ram and Wolfner 2007; Pilpel et al. 2008; Yapici et al. 2008; Yang et al. 2009; Heifetz et al. 2010; Vargas et al. 2010). Therefore, the failure of heterospecific ejaculates to fully activate post-copulation physiologies is driven by the heterospecific ejaculate not possessing the correct genotype or genotypes to interact properly with the female reproductive tract.

Within the female reproductive tract, the activation of particular post-copulation physiologies starts at the plasma membrane. Whether activator/ligand molecules are allowed to freely flow across the membrane through channels or bind to receptors on the membrane surface, the dynamics of cell physiology start at the membrane (see Hancock 2005). A critical component of this activation is cell signaling – whereby a single message can be received at the cell surface and then transferred to the rest of the cell, including the nucleus, via turning on a particular genetic pathway that “transfers” the message to the appropriate target in the cell (Alberts et al. 2007). In brief, there are three major types of transmembrane receptors that facilitate cell signaling: ion-channel-coupled receptors, G-protein coupled receptors (GPCRs), and enzyme-coupled receptors (e.g., receptor tyrosine kinases, RTKs, are receptors that directly phosphorylate their target when activated; Alberts et al. 2007). The latter two types of receptors (i.e., GPCRs and RTKs), along with a handful of “minor” receptors (e.g., integrins), are relevant to phosphorylation (see Figure 1).

The work of phosphorylation is done by kinases (depicted by the shaded objects in Figure 1), which are proteins that add one or more phosphate groups to the amino-acid side chains of a target protein. This process triggers conformational changes in the target protein that functionally turns it on or off. In short, the activation of a kinase can be viewed as turning on a master switch that either starts or stops a physiological process. For example, a specific GPCR receptor can be activated by a ligand, leading to the activation of a G-protein which can then turn on a specific kinase like GSK3 or one or more of a set of kinases like PKA, PKC, or PI3K (Figure 1). While these phosphorylation pathways occur in all cells, there is a rich history of studies investigating the role of phosphorylation in reproductive cells and physiology (e.g., Sato et al. 1998; Visconti and Kopf 1998; Muhrad and Ward 2002; Urner and Sakkas 2003; Simonet et al. 2004; Baker et al. 2006; Horner et al. 2006).

Overall, it is these male by female interactions that underlie successful phosphorylation of reproductive proteins and could diverge between species. For instance, species-specific divergence in one male-ligand — female-receptor pair could lead to dysfunctional phosphorylation of target proteins in the female reproductive tract (or a target cell outside the reproductive tract) following a heterospecific mating. In this case, the lack of proper binding of the male ligand to the female receptor would render the specific phosphorylation pathway

inactive and thus not turn on the “master switch” leading to a *necessary* physiological function. This proposed mechanism is all the more plausible when considering that transmembrane receptors like GPCRs and RTKs are often under positive selection (e.g., Nimura and Nei 2005; Salzburger et al. 2007; Steiger et al. 2009) and their evolutionary histories are replete with gene duplication events and subfunctionalization (Brody and Cravchik 2000; Harmar 2001; Bjarnadottir et al. 2006; Grassot et al. 2006) – thus increasing the likelihood that protein-protein mismatches will evolve.

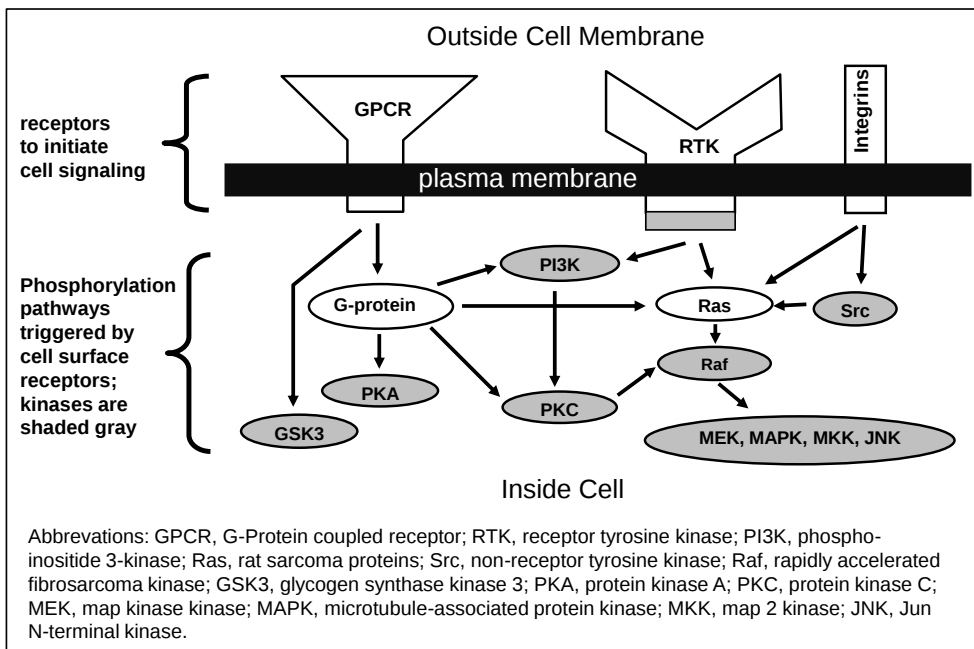


Figure 1. Cell membrane receptors and kinases that could underlie postmating, prezygotic phenotypes. Receptors are in the plasma membrane and kinases (shaded gray) are inside the cell.

Using Phosphorylation Studies to Identify Ligand and Receptor Genes Underlying Postmating, Prezygotic Reproductive Isolation

In general, the study of phosphorylation patterns can be used as a tool to identify species-specific, receptor-ligand interactions that lead to postmating, prezygotic reproductive isolation. The key is to evaluate patterns of phosphorylation in the tissue or tissues where species-specific interactions might occur. The easiest example to consider is when a species-specific, receptor-ligand interaction takes place within the female reproductive tract. In this scenario, you can compare patterns of phosphorylation within the female reproductive tract between females that have successfully copulated with either a con- or heterospecific male. To do so, you would use 2D protein gels (not 2D-DIGE, i.e., two dimensional differential in-gel electrophoresis) that stain for total protein (e.g., usually fluorescently labeled green prior to running the gel) and phosphoproteins (i.e., proteins that have been phosphorylated; stained for after running the gel and usually in red for easy comparisons; e.g., Westermeyer et al. 2008). You would run one 2D gel on the reproductive tracts from conspecific copulations and a separate 2D gel on ones from

heterospecific copulations. As an added control, it is useful to run a similar 2D gel on reproductive tracts from virgin females.

Once these gels are run and stained, you can compare them to determine if patterns of phosphorylation in the female reproductive tract are different between con- and heterospecific copulations compared with the virgin-female control. If you find that particular phosphorylation only occurs after a conspecific copulation (but not following a heterospecific copulation or within a virgin female), then you can explore the link between that pattern and the postmating, prezygotic phenotype that isolates the species. To do this, you would first use MS/MS to identify the differentially phosphorylated proteins. Once these proteins have been identified you can assess the function of these proteins (via blast and literature searches) and determine if they could underlie the specific postmating, prezygotic phenotype. If so, then you can look at the peptide data from your MS/MS protein identifications to determine which amino acids were phosphorylated. You can then “blast” the sequences from the phosphorylated peptides in a database like the NetPhos 2.0 Server (www.cbs.dtu.dk/services/NetPhos/) to identify which kinase likely did the phosphorylation. Once you know the kinases involved a little literature review can point you toward the receptors that are involved.

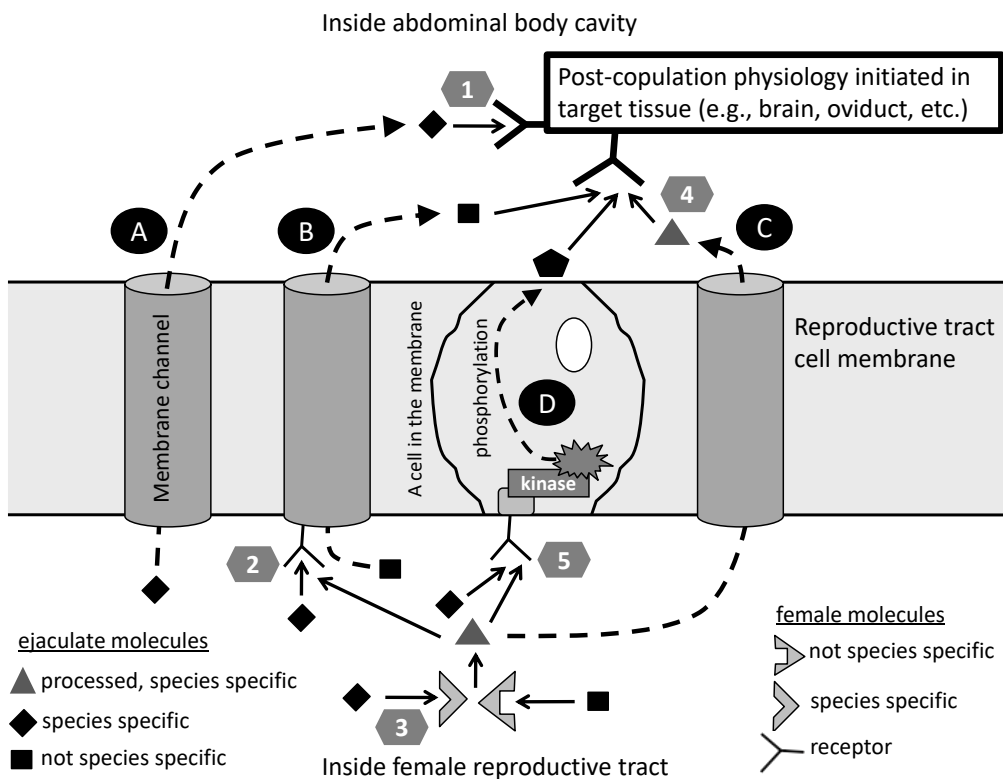


Figure 2. The mechanisms (A, B, C, and D) and potential species-specific (1, 2, 3, 4, and 5) components of postmating, prezygotic isolation along the cell membrane of the female reproductive tract.

At this point, you will either need a fully sequenced genome or a reproductive tract EST database to screen for the candidate receptors. Indeed, even if you have a fully sequenced genome, a tissue-specific EST library is needed to identify candidate receptors, as there may be hundreds or thousands of receptors coded for in the genome with different ones being expressed in different tissues. With the candidate receptor(s) in hand, functional analyses (e.g., RNAi or knockout lines) can be used to determine if the candidate receptor influences the target postmating, prezygotic phenotype. If the receptor does influence the phenotype of interest, then you can try to identify the ligand that activates the receptor – a process that will warrant discussions with your local biochemist or cell physiologist.

Furthermore, once a receptor or receptor-ligand pair is identified, you can sequence the underlying gene to determine if there is species-specific variation and whether or not positive selection has influenced the occurrence of that variation. Also, if species-specific alleles for the receptor and ligand are identified, then functional assays can be conducted to determine if conspecific receptor-ligand pairs bind and, or function better than heterospecific pairs. In the end, if the above approach yields a female receptor that is linked to the postmating, prezygotic phenotype that isolates species, is under positive selection, controls a portion of phosphorylation that only occurs following a conspecific copulation, and interacts with the male ligand in a species-specific fashion, then a strong case can be made that the particular receptor-ligand pair contributes to reproductive isolation (and potentially speciation if the phenotype is among the only traits that isolate the focal species). Although a few “tools” like reproductive tract EST databases and the ability to do RNAi are needed for this approach, it can be conducted in almost any system, thus enabling researchers to explore the genetics of postmating, prezygotic isolation in non-model systems.

PATHWAYS LEADING TO POSTMATING, PREZYGOTIC REPRODUCTIVE ISOLATION

At its core, most postmating, prezygotic reproductive isolation must be driven by incompatible or dysfunctional interactions between the reproductive proteomes of heterospecifics. Understanding the dynamics of male-female proteome-proteome interactions and how they lead to reproductive isolation is a complex endeavor as there are hundreds, if not thousands, of ways in which such interactions could fail. My goals here are to outline some of the major mechanisms that should be considered, where species-specificity could occur, and ways in which these dynamics can be studied regardless of the organismal system.

Membrane Channels and Reproductive Isolation Caused by Protein-Protein Interactions Outside the Female Reproductive Tract

Researchers have known for some time that male reproductive proteins can leave the female reproductive tract and affect female physiology (e.g., Lung and Wolfner 1999; Gillott 2003; Lay et al. 2004). The mechanism of action in these cases can be relatively simple as the environment of the female reproductive tract is usually altered following copulation (e.g., Engelmann 1970; Gillott 2003). For example, the pH or salinity may change resulting in the

opening of cell membrane or tissue channels that allow molecules (including male reproductive proteins) in the female reproductive tract to leave and travel to other places in the female body. Given that the mechanism of passing through the female reproductive tract can be considered to be passive (or non-selective) in this case, it is reasonable to ask, “where might a species-specific interaction occur that leads to reproductive isolation?”

The answer to this question lies in understanding where these male sex molecules are going once they leave the female reproductive tract (see Figure 2, mechanism “A”). These male molecules could be targeting any number of tissues including the ovaries, oviducts, or brain. However, regardless of the target tissue, the male molecules could interact with receptors on the surface of the target tissue in a species-specific fashion (Figure 2, see the #1). A successful, conspecific interaction could then trigger cell signaling that leads to the initiation of post-copulation physiology. On the other hand, molecules from a heterospecific male would fail to activate the receptor thus not eliciting the post-copulation physiology.

Of the mechanisms discussed here, this is among the most difficult to assess experimentally. The difficulty, in this case, lies in determining which target tissue or tissues outside the female reproductive tract should be evaluated. A possible solution is to simply evaluate a wide range of tissues for differences in post-copulation phosphorylation (as outlined above). Consequently, there may be several different candidate receptors to test that require tissue-specific knockdown to evaluate their link to the postmating, prezygotic phenotype that isolates species. While this approach is certainly more time consuming, it is not overly expensive (i.e., a few thousand US dollars per receptor) and still workable for most systems.

Species-Specific, Receptor-Gated Membrane Channels

Mechanism “B” (Figure 2) assumes that a species-specific interaction between a male ejaculate molecule and a female receptor (see #2 in Figure 2) controls the opening of a membrane channel that subsequently allows non-species specific molecules to flow across the female reproductive tract membrane and eventually bind with a receptor activating a post-copulation physiology. As with mechanism “A”, identifying the interacting proteins can be difficult. For example, this mechanism could yield patterns of post-copulation phosphorylation similar to those from mechanism “A”; yet unlike mechanism “A”, a study of the receptors in the affected tissue would yield no evidence of a species-specific interaction linked to the postmating, prezygotic phenotype. Instead, a functional analysis of membrane channel receptors, especially those that are uniquely expressed in the female reproductive tract, would be a reasonable place to start. Moreover, if such a receptor were found to be linked to the postmating, prezygotic phenotype, then the next step would be to simply sequence the receptor gene to assess species-specific variation. A finding of species-specific variation, especially variation driven by positive selection, would warrant further study as a mechanism of reproductive isolation.

Species-specific, Male-female Protein Interactions within the Female Reproductive Tract

A common mechanism of species-specific interactions involves the direct interaction of male ejaculate proteins and female reproductive tract proteins (see #3 in Figure 2). These interactions could involve a wide range of proteins and peptides. For example, species-specific cleavage of a male pro-hormone by a female protease may generate the ligand that activates a membrane channel (see mechanism “B” in Figure 2) or a phosphorylation pathway (Figure 2), or passes through a membrane channel to interact with a receptor outside the female reproductive tract (mechanism “C” in Figure 2). The hormone ligand in this scenario does not have to exhibit species-specific variation, it simply has to be cleaved in a species-specific fashion. On the other hand, the pro-hormone—protease interaction does not itself have to be a species-specific interaction, as the resulting ligand could contain variation that results in species-specificity when it interacts with a receptor inside (see #2 and #5 in Figure 2) or outside the female reproductive tract (see #4 in Figure 2).

These types of male-female proteome interactions can be evaluated with comparative proteomic techniques like 2D-DIGE (e.g., Westermeier et al. 2008). With this technique, two or three samples can be compared on the same protein gel with each sample being labeled with a different fluorescent dye. So, if we compare a female reproductive tract following a conspecific copulation with one from a heterospecific copulation, then it would be possible to determine if a single or set of “new” protein spot showed up on the gel following the conspecific copulation, but not the heterospecific one – which would provide evidence for a species-specific interaction. You would also have two controls in this case with at least one of them run on the above gel as a third sample. The two controls would be (1) comparing virgin female reproductive tracts from the two species and (2) comparing male ejaculates from the two species with each control being run on a separate 2D-DIGE gel (also each species would be labeled with a different dye). The control for the above post-copulation gel would be a combined sample from one of these sex-specific controls. The power of this design is that you can separate the effects of species divergence from species-specific “processing events” that only occur after a conspecific copulation. As outlined above, these uniquely processed proteins can be identified with MS/MS and tested further. Of all the mechanisms outlined here, this is likely the easiest to conduct and can be a very powerful way to begin to explore the genetic basis of postmating, prezygotic phenotypes (see Marshall et al. 2011).

Mechanisms Triggering Phosphorylation within the Female Reproductive Tract

While the above pathways are possible, previous research points to processes such as cell signaling and phosphorylation in the female reproductive tract as being key players in initiating post-copulation physiologies. In particular, transcriptomic studies comparing the reproductive tracts from virgin females to those from mated females indicate that a large number of genes are turned on after mating (e.g., Wolfner 2002; McGraw et al. 2004, 2008; Mack et al. 2006). Based on cell structure, the activation of gene expression in female reproductive tissues post-copulation must be triggered by cell signaling – specifically through the action of male ligands

binding to transmembrane receptors in the female reproductive tract that in turn activate the phosphorylation pathways leading to gene expression (reviewed in Hancock 2005). This finding in combination with research showing that phosphorylation plays a key role in reproductive physiology (e.g., Visconti and Kopf 1998; Simonet et al. 2004; Escobar-Restrepo et al. 2007), indicates that the successful initiation of phosphorylation within the female tract is a fruitful area to search for species-specificity.

As shown in Figure 2, there are several pathways to generate species-specificity in the receptor-ligand interaction that triggers phosphorylation (see the “inside female reproductive tract” portion of mechanism “D” in Figure 2). Indeed, the ligand can come directly from the male and interact with the cell membrane receptor in a species-specific fashion (see #5 in Figure 2). However, as outlined in the “Species-specific, male-female protein interactions within the female reproductive tract” section above, the male ligand could be processed in the female reproductive tract prior to interacting with the female receptor. In this case, species-specificity could lie in the processing event or the product of the processing event. Regardless of the pathway, it is not difficult to see how heterospecific copulations could yield a ligand, or set of ligands, that fail to induce post-copulation phosphorylation and thus, results in reproductive isolation.

CONCLUSION

In all, the above is meant to paint a very particular picture about how basic biology, cell structure and function, and organismal physiology combine to reveal that certain kinds of genes are more likely than others to be involved in postmating, prezygotic reproductive isolation. While there are certainly accents or fine-scale elements that may be missing from the above picture, I hope the canvass reveals the major themes and motivates increased research on the genetics of postmating, prezygotic isolation. One brush stroke that I may not have emphasized enough is the ubiquity of this approach. We have been doing these approaches in the cricket genus *Allonemobius*, which is not a genetic model system. The proteomic and biochemical approaches outlined here, for the most part, can be used in almost any system – thus making it possible for a relatively complete picture of the genetics of postmating, prezygotic isolation to be developed if enough researchers add a brush stroke from their respective organismal system.

ACKNOWLEDGMENTS

I thank my students and colleagues, past and present, for their help in developing this approach to the genetics of postmating, prezygotic isolation. As I have stated in numerous seminars, I am not a biochemist, so I am particularly thankful for Drs. John Tomich and Brenda Oppert at Kansas State University for teaching me and guiding me through the process of comparative proteomics. The ideas for this work were developed from my research funded by NSF (DEB-0746316). This is contribution number 12-376-B from the Kansas Agricultural Experiment Station.

REFERENCES

- Alberts B., Johnson A., Lewis J., Raff M., Roberts K., Walter P. 2007. *Molecular Biology of the Cell*. 5th Ed. London: Garland Science.
- Baker M. A., Hetherington L., Aitken R. 2006. Identification of SRC as a key PKA-stimulated tyrosine kinase involved in the capacitation-associated hyperactivation of murine spermatozoa. *J. Cell Science* 119: 3182-3192.
- Bjarnadóttir T. K., Gloriam D. E., Hellstrand S. H., Kristiansson H., Fredriksson R., Schiöth H. B. 2006. Comprehensive repertoire and phylogenetic analysis of the G protein-coupled receptors in human and mouse. *Genomics* 88:263-273.
- Bonaldi T., Straub T., Cox J., Kumar C., Becker P. B., Mann M. 2008. Combined use of RNAi and quantitative proteomics to study gene function in *Drosophila*. *Molecular Cell* 31:762-772.
- Boughman J. W. 2002. How sensory drive can promote speciation. *TREE* 17:571-577.
- Brody T., Cravchik A. 2000. *Drosophila melanogaster* G Protein-coupled Receptors. *J. Cell Biol.* 150:F83-F88.
- Brown V., Eady P. E. 2001. Functional incompatibility between the fertilization systems of two allopatric populations of *Callosobruchus maculatus* (Coleoptera: Bruchidae). *Evolution* 55:2257-2262.
- Butlin R. K. 2010. Population genomics and speciation. *Genetica* 138:409-418.
- Clark N. L., Aagaard J. E., Swanson W. J. 2006. Evolution of reproductive proteins from animals and plants. *Reproduction* 131:11-22.
- Dean M. D., Nachman M. W. 2009. Faster fertilization rate in conspecific versus heterospecific matings in house mice. *Evolution* 63:20-28.
- Dobzhansky T. 1964. Biology, molecular and organismic. *American Zoologist* 4:443-452.
- Dorus S., Wasbrough E. R., Busby J., Wilkin E. C., Karr T. L. 2010. Sperm proteomics reveals intensified selection on mouse sperm membrane and acrosome genes. *Mol Biol Evol* 27: 1235-1246.
- Endler J. A., Basolo A. L. 1998. Sensory ecology, receiver biases and sexual selection. *TREE* 13:415-420.
- Engelmann F. 1970. *The physiology of insect reproduction*. New York: Pergamon Press.
- Escobar-Restrepo J.-M., Huck N., Kessler S., Gagliardini V., Gheyselinck J., Yang W. C., Grossniklaus U. 2007. The FERONIA receptor-like kinase mediates male-female interactions during pollen tube reception. *Science* 317:656-660.
- Finkel T., Holbrook N. J. 2000. Oxidants, oxidative stress and the biology of ageing. *Nature* 408: 239-247.
- Geyer L., Palumbi S. R. 2005. Conspecific sperm precedence in two species of tropical sea urchins. *Evolution* 59:97-105.
- Gillott C. 2003. Male accessory gland secretions: modulators of female reproductive physiology and behavior. *Annual Reviews of Entomology* 48:163-184.
- Grassot J., Gouy M., Perrière G., Mouchiroud G. 2006. Origin and molecular evolution of receptor tyrosine kinases with immunoglobulin-like domains. *Mol. Biol. Evol.* 23:1232-1241.
- Hancock, J. 2005. *Cell Signalling*. New York: Oxford University Press.
- Harmar A. J. 2001. Family-B G-protein-coupled receptors. *Genome Biol.* 2:reviews3013.1-10.

- Heifetz Y., Lung O., Frongillo E. A. Jr., Wolfner M. F. 2000. The *Drosophila* seminal fluid protein Acp26Aa stimulates release of oocytes by the ovary. *Curr. Biol.* 10:99-102.
- Horner C. L. B., Dohi A., Nguyen Q., Dillon G. H., Singh M. 2006. ERK/MAPK pathway regulates GABA_A receptors. *J. Neurobiology* 66:1467-1474.
- Howard D. J. 1999. Conspecific sperm and pollen precedence and speciation. *Ann. Rev. Ecol. Syst.* 30:109-132.
- Huestis D. L., Marshall J. L. 2009. From gene expression to phenotype in insects: non-microarray approaches for transcriptome analysis. *Bioscience* 59:373-384.
- Jensen O. N. 2006. Interpreting the protein language using proteomics. *Nature Reviews Molecular Cell Biology* 7:391-403.
- Johansson M., Bromfield J. J., Jasper M. J., Robertson S. A. 2004. Semen activates the female immune response during early pregnancy in mice. *Immunology* 112:290-300.
- Kulathinal R. J., Stevison L. S., Noor M. A. F. 2009. The genomics of speciation in *Drosophila*: diversity, divergence, and introgression estimated using low-coverage genome sequencing. *PLoS Genet* 5:e1000550.
- Lay M., Loher W., Hartmann R. 2004. Pathways and destination of some male gland secretions in female *Locusta migratoria migratorioides* (R & F) after insemination. *Achieves of Insect Biochemistry and Physiology* 55:1-25.
- Levesque L., Brouwers B., Sundararajan V., Civetta A. 2010. Third chromosome candidate genes for conspecific sperm precedence between *D. simulans* and *D. mauritiana*. *BMC Genetics* 11:21.
- Lung O., Wolfner M. F. 1999. *Drosophila* seminal fluid proteins enter the circulatory system of the mated female fly by crossing the posterior vaginal wall. *Insect Bioch. Mol. Biol.* 29:1043-1052.
- Mack P. D., Kapelnikov A., Heifetz Y., Bender M. 2006. Mating-responsive genes in reproductive tissues of female *Drosophila melanogaster*. *PNAS* 103:10358-10363.
- Marshall J. L. 2004. The *Allonemobius-Wolbachia* host-endosymbiont system: evidence for rapid speciation and against reproductive isolation driven by cytoplasmic incompatibility. *Evolution* 58:2409-2425.
- Marshall J., Huestis D. L., Garcia C., Hiromasa Y., Wheeler S., Noh S., Tomich J. M., Howard D. J. 2011. Comparative proteomics uncovers the signature of natural selection acting on the ejaculate proteomes of two cricket species isolated by postmating, prezygotic phenotypes. *Mol. Biol. Evol.* 28:423-435.
- McGraw L. A., Gibson G., Clark A. G., Wolfner M. F. 2004. Genes regulated by mating, sperm, or seminal proteins in mated female *Drosophila melanogaster*. *Curr. Biol.* 14:1509-1514.
- McGraw L. A., Clark A. G., Wolfner M. F. 2008. Post-mating gene expression profiles of female *Drosophila melanogaster* in response to time and to four male accessory gland proteins. *Genetics* 179:1395-1408.
- Muhlrad P. J., Ward S. 2002. Spermiogenesis Initiation in *Caenorhabditis elegans* involves a casein kinase 1 encoded by the *spe-6* gene. *Genetics* 161:143-155.
- Neubaum D. M., Wolfner M. F. 1999. Mated *Drosophila melanogaster* females require a seminal fluid protein, Acp36DE, to store sperm efficiently. *Genetics* 153:845-857.
- Niimura Y., Nei M. 2005. Evolutionary dynamics of olfactory receptor genes in fishes and tetrapods. *PNAS* 102: 6039-6044.
- Noor M. A. F., Feder J. L. 2006. Speciation genetics: Evolving approaches. *Nature Reviews Genetics* 7:851-861.

- Otronen M., Siva-Jothy M. T. 1991. The effect of postcopulatory male behaviour on ejaculate distribution within the female sperm storage organs of the fly, *Dryomyza anilis* (Diptera: Dryomyzidae) *Behav. Ecol. Socio.* 29:133-137.
- Pilpel N., Nezer I., Applebaum S. W., Heifetz Y. 2008. Mating-increases trypsin in female *Drosophila* hemolymph. *Insect Biochem Mol Biol* 38:320330.
- Presgraves D. C. 2010. The molecular evolutionary basis of species formation. *Nature Reviews Genetics* 11:175-180.
- Ram K. R., Wolfner M. F. 2007. Sustained post-mating response in *Drosophila melanogaster* requires multiple seminal fluid proteins. *PLoS Genet.* 3:e238.
- Ramm S. A., McDonald L., Hurst J. L., Beynon R. J., Stockley P. 2009. Comparative proteomics reveals evidence for evolutionary diversification of rodent seminal fluid and its functional significance in sperm competition. *Mol. Biol. Evol.* 26:189-198.
- Salzburger W., Braasch I., Meyer A. 2007. Adaptive sequence evolution in a color gene involved in the formation of the characteristic egg-dummies of male haplochromine cichlid fishes. *BMC Biology* 5:51.
- Sato K., Iwasaki T., Tamaki I., Aoto M., Tokmakov A. A., Fukami Y. 1998. Involvement of protein-tyrosine phosphorylation and dephosphorylation in sperm-induced *Xenopus* egg activation. *FEBS Lett.* 424:1-2.
- Simmons L. W. 2001. *Sperm competition and its evolutionary consequences in insects*. Princeton: Princeton University Press.
- Simonet G., Poels J., Claeys I., van Loy T., Franssens V., de Loof A., Vanden Broeck J. 2004. Neuroendocrinological and molecular aspects of insect reproduction. *J. Neuroendocrinology* 16:649-659.
- Steiger S. S., Kuryshv V. Y., Stensmyr M. C., Kempnaers B., Mueller J. C. 2009. A comparison of reptilian and avian olfactory receptor gene repertoires: species-specific expansion of group γ genes in birds. *BMC Genomics* 10:446.
- Sultan S. E. 2007. Development in context: the timely emergence of eco-devo. *TREE* 22:575-582.
- Swanson W. J., Clark A. G., Waldrip-Dali H. M., Wolfner M. F., Aquadro C. F. 2001. Evolutionary EST analysis identifies rapidly evolving male reproductive proteins in *Drosophila*. *PNAS* 98:7375-7379.
- Swanson W. J., Wong A., Wolfner M. F., Aquadro C. F. 2004. Evolutionary expressed sequence tag analysis of *Drosophila* female reproductive tracts identifies genes subjected to positive selection. *Genetics* 168:1457-1465.
- Swanson W. J., Vacquier V. D. 2002a. The rapid evolution of reproductive proteins. *Nat. Rev. Genet.* 3:137-144.
- Swanson W. J., Vacquier V. D. 2002b. Reproductive protein evolution. *Ann. Rev. Ecol. Syst.* 33:161-179.
- Sweigart A. L. 2010. The genetics of postmating, prezygotic reproductive isolation between *Drosophila virilis* and *D. americana*. *Genetics* 184:401-410.
- Turner L. M., Hoekstra H. E. 2008. Causes and consequences of the evolution of reproductive proteins. *Int. J. Dev. Biol.* 52:769-780.
- Urner F., Sakkas D. 2003. Protein phosphorylation in mammalian spermatozoa. *Reproduction* 125:17-26.

- Vargas M. A., Luo N., Yamaguchi A., and Kapahi P. 2010. A role for S6 kinase and serotonin in postmating dietary switch and balance of nutrients in *D. melanogaster*. *Curr. Biol.* 20:1006-1011.
- Visconti P. E., Kopf G. S. 1998. Regulation of protein phosphorylation during sperm capacitation. *Biology of Reproduction* 59:1-6.
- Westermeyer R., Naven T., Hopker H.-R. 2008. *Proteomics in practice: A guide to successful experimental design*. 2nd ed. Weinheim (Germany), Wiley-VCH.
- Wolfner M. F. 2002. The gifts that keep on giving: physiological functions and evolutionary dynamics of male seminal proteins in *Drosophila*. *Heredity* 88:85-93.
- Yang C. H., Rumpf S., Xiang Y., Gordon M. D., Song W., Jan L. Y. and Jan Y. N. 2009. Control of the postmating behavioral switch in *Drosophila* females by internal sensory neurons. *Neuron* 61:519-526.
- Yapici N., Kim Y. J., Ribeiro C., Dickson B. J. 2008. A receptor that mediates the post-mating switch in *Drosophila* reproductive behaviour. *Nature* 451:33-37.

Chapter 9

THE MOLECULAR AND EVOLUTIONARY BASIS OF HYBRID STERILITY: FROM *ODYSSEUS* TO *OVERDRIVE*

Nitin Phadnis^{1,*} and Harmit S. Malik^{2,3}

¹Department of Biology, University of Utah, Salt Lake City, UT, US

²Howard Hughes Medical Institute, Chevy Chase, MD, US

³Division of Basic Sciences, Fred Hutchinson Cancer Research Center,
Seattle, WA, US

ABSTRACT

Speciation, the process by which one species splits into two, involves the evolution of reproductive barriers such as the sterility or inviability of hybrids between previously interbreeding populations. One of the earliest intrinsic barriers to gene flow to evolve between geographically isolated populations is the sterility of hybrids of the heterogametic sex. The Dobzhansky-Muller model describes how hybrid incompatibilities that underlie intrinsic postzygotic reproductive barriers such as hybrid sterility may evolve. A major goal in speciation research is to identify the genes that underlie hybrid incompatibilities. The identification of such genes opens the door to understanding the molecular pathways which, when tinkered by evolution within species, lead to sterility in hybrids, and promises to reveal the biological forces that drive the evolution of hybrid sterility genes. While very few hybrid sterility genes have been identified so far, the idea that the evolution of DNA-protein interfaces driven by intragenomic conflict may cause hybrid sterility is gaining wider acceptance. Here we describe how the molecular and evolutionary insights from two hybrid sterility genes – *Odysseus* and *Overdrive* – have illuminated the role of heterochromatin in the molecular basis of hybrid sterility and the role of genetic conflict as a driving force in speciation.

* Corresponding Author's Email: nitin.phadnis@gmail.com (Nitin Phadnis, Department of Biology, 257 South 1400 East, University of Utah, Salt Lake City, UT 84112; Tel: (801) 581-6517; Fax: (801)581-4668).

INTRODUCTION

In his book *"On the origin of species"*, Darwin compiled overwhelming evidence for two key phenomena. First, species adapt. Second, single species can split into two species. A combination of these two processes, according to Darwin, is able to provide a natural explanation for the amazing diversity of life. One of the great and enduring successes of Darwin lay in his explanation of the first process. Species adapt through the mechanism called "natural selection" in which genetic variation within a population is constantly subjected to selection sieves, through which alleles conferring fitter traits pass through to subsequent generations while those conferring weaker traits do not. The second process, however, vexed Darwin till the end. Speciation, the process by which one species splits into two, involves the evolution of reproductive barriers such as the sterility or inviability of hybrids between previously interbreeding populations. The question of how natural selection could possibly permit the evolution of apparently severe deleterious traits, such as sterility or inviability, proved much harder to explain. Darwin could find no satisfactory solution to this fundamental problem and, therefore, called it the *"mystery of mysteries"*.

This *"mystery of mysteries"* remains one of the important and unanswered questions in biological research: how do two species evolve from one? Speciation research became more tractable when Ernst Mayr helped define speciation as the evolution of reproductive isolation. Mayr's Biological Species Concept states that *"species are truly or potentially interbreeding natural populations that are reproductively isolated from other such groups"* (Mayr 1942). Thus, genetic studies of speciation could be recast into a more tractable problem: the genetic basis of reproductive isolation. Theodosius Dobzhansky made further progress in the study of speciation by classifying different reproductive isolating barriers into discrete, manageable categories (Dobzhansky 1937). Thus, "prezygotic" isolating barriers (e.g., behavioral isolation) act before a hybrid zygote is formed, whereas "postzygotic" isolating barriers (hybrid sterility or inviability) act after a hybrid zygote is formed.

As Darwin recognized, postzygotic isolating barriers are a seemingly counterintuitive phenomenon. Why would natural selection permit, let alone favor, the onset of genetic barriers that diminish the prospect of successful reproduction? A significant step towards the theoretical resolution of this paradox emerged in the early 1900's. Bateson, Dobzhansky and Muller independently suggested a genetic model of the evolution of postzygotic isolation that addressed this problem. In particular, they showed how hybrid sterility or inviability could evolve unopposed by natural selection. A simple two-locus version of the Dobzhansky-Muller model (Figure 1) considers a species that is split into two geographically isolated populations. Initially, both populations are genetically identical, having an *AA* genotype at one locus and *BB* at another. An allele *a*, compatible with both *A* and *B*, appears and is fixed in one population (*aaBB*). Similarly, an allele *b* appears and is fixed in the other population (*AAbb*). The key point of the model is that, while the newly emerged *a* and *b* alleles function properly in their respective genetic backgrounds, natural selection has not tested *a* and *b* together as they have never appeared together in the same genome. Thus, when hybrids between the two populations are formed, there is no guarantee that *a* and *b* will function properly in a common genome. The possible result of a negative epistatic interaction between *a* and *b*, referred to as a Dobzhansky-Muller incompatibility, is sterility or inviability of hybrids.

The Dobzhansky-Muller model enjoys considerable empirical support and is now well established as the appropriate theoretical framework for understanding the evolution of intrinsic postzygotic isolation (Coyne, Orr 2004). This has helped advance speciation research from classical genetic studies to identifying individual genes that cause hybrid incompatibilities. The identification of genes that cause hybrid sterility or inviability is an indispensable step towards a deeper understanding of speciation, and provides one of the most promising – but challenging – problems in evolutionary genetics. Several hybrid inviability genes have been identified so far, with most of the successes coming from crosses between the model genetic system *Drosophila melanogaster* and its sister species *D. simulans*. The melanogaster species subgroup consists of four species of *Drosophila* that are believed to have diverged approximately 2.5 million years ago (Figure 2). This group of species has attracted a lot of attention in the field of speciation because it includes *Drosophila melanogaster*, which is the model organism of choice for many genetic and genomic studies. Crosses between *D. melanogaster* and the other three species yield only inviable or sterile progeny in both sexes. These hybrid progeny can be inviable, depending on the direction of the cross (i.e., which species is maternal) and this has provided a basis to genetically dissect the causes of hybrid inviability. The isolation of naturally occurring strains that rescue hybrid inviability (Watanabe 1979; Hutter, Ashburner 1987; Sawamura, Taira, Watanabe 1993; Sawamura, Watanabe, Yamamoto 1993) and the availability of the formidable genetic toolkit in *D. melanogaster* has facilitated the rapid discovery of hybrid inviability genes (Presgraves 2010; Maheshwari, Barbash 2011).

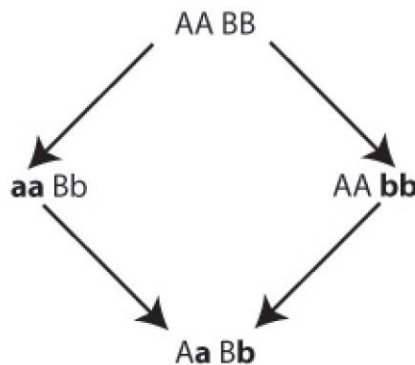


Figure 1. The Dobzhansky-Muller model.

In contrast to hybrid inviability, which often arises later during the accumulation of reproductive isolating barriers, hybrid male sterility is one of the earliest evolving postzygotic barriers between geographically isolated populations. The identification of the hybrid male sterility genes in young species pairs is, therefore, essential for insights into the evolution of reproductive barriers relevant during the earliest stages of speciation. Unfortunately, far fewer hybrid sterility genes are known thus far than ones that cause hybrid inviability. One major reason for this discrepancy is that the powerful tools and rich history of mutations in *D. melanogaster* have not been as useful from the perspective of the study of hybrid sterility due to its older divergence from its sister species (with the exception of *JYalpha*, which is a case of gene transposition (Masly et al. 2006)). Evolutionary geneticists have, instead, focused on identifying hybrid sterility genes in crosses involving young, non-model species pairs.

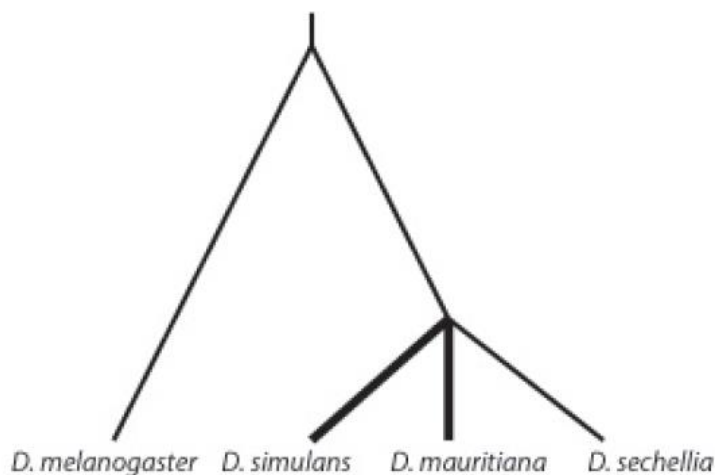


Figure 2. The melanogaster species subgroup of *Drosophila*.

Recent work on hybrid sterility genes in two young *Drosophila* species pairs have begun providing a deeper understanding of the phenomenon of hybrid sterility at the molecular level and of the particular biological forces that drive the evolution of hybrid sterility genes. First, the case of *Odysseus*, which causes hybrid sterility between *Drosophila simulans* and *D. mauritiana*, has highlighted the role of rapidly evolving heterochromatin and heterochromatin-binding proteins in the evolution of hybrid male sterility. Second, the case of *Overdrive*, which causes hybrid male sterility between two subspecies *Drosophila pseudoobscura* USA and Bogota, has provided one of the first direct lines of evidence for the role of genetic conflict involving segregation distorters in the evolution of hybrid sterility. These insights from *Odysseus* and *Overdrive* have been largely complementary. *Odysseus* is one of the best-characterized hybrid sterility genes at the molecular level, with few hints about the biological processes that may have driven its evolution. *Overdrive* is one of the best-understood hybrid sterility genes with regards the evolutionary forces that drove the changes at this gene, but little is known about its action at the molecular level. Here, we review the genetic, evolutionary and molecular insights provided by these two hybrid sterility genes – *Odysseus* and *Overdrive* – and outline future research prospects.

ODYSSEUS

The melanogaster species subgroup consists of four species of *Drosophila* that are believed to have diverged approximately 2.5 million years ago (Figure 2). The study of hybrid sterility has focused on the three sibling species, *D. simulans* (a cosmopolitan species found worldwide), *D. mauritiana* and *D. sechellia* (both island species). These three species are believed to have diverged only about 250,000 years ago (Kliman et al. 2000). Every possible cross between these three species yields fertile females and sterile males. Because the females are fertile, they can be backcrossed to either parental species to yield introgression lines, in which genetic material from one species can be brought into a genomic background that is otherwise entirely from another species. This allows a genetic investigation of the factors that may cause hybrid male sterility. For instance, in 1993, this genetic introgression strategy

revealed the basis for a single chromosomal location, 16D, which when introduced from *D. mauritiana*'s X chromosome into the corresponding position of the *D. simulans* X chromosome caused complete male sterility (Perez et al. 1993). This genetic factor was named *Odysseus* ("who, in the well known epic, was the major figure hidden in the Trojan horse that in the end caused complete destruction of the foreign land it was brought into" (Perez et al. 1993)). Further genetic discussion suggested that a single gene of *D. mauritiana* origin at 16D along with a few other neighboring genes were required for the manifestation of complete male sterility (Figure 3).

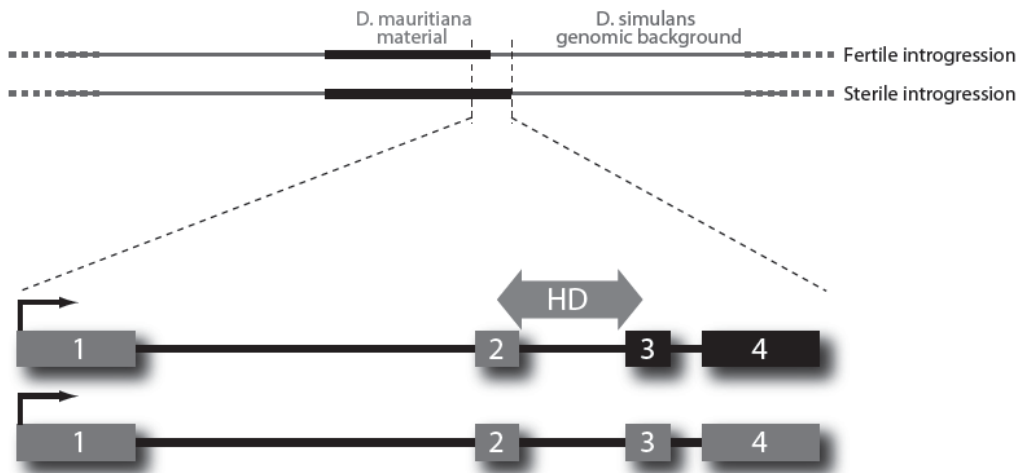


Figure 3. Schematic of *D. mauritiana* introgressions into the *D. simulans* genomic background. The introgression that introduces the whole of *OdsH* from *D. mauritiana* is sterile.

Odysseus was the first animal hybrid sterility gene identified (Ting et al. 1998). It was found to encode a homeobox containing protein (hence the name *Odysseus site homeobox gene*, or simply *OdsH*). Homeoboxes are typically DNA-binding modules found in transcription factors and are highly conserved because of their important DNA-binding function. Intriguingly, the homeobox of *OdsH* had undergone rapid evolution, much more so than the rest of the protein-coding gene and even faster than neighboring intronic sequence. Further investigation revealed that the *OdsH* gene represented a gene-duplication from the *unc4* transcription factor, which is found in all metazoans (an evolutionary history spanning 700 million years) and is very slow to evolve. *OdsH* has evolved rapidly even in 250,000 years of *Drosophila* evolution, with 16 replacement changes having taken place in the 60 codon homeodomain, a remarkably high rate of evolution. Furthermore, whereas *unc-4* is expressed primarily in neural tissue, *OdsH* is exclusively expressed in testes, consistent with its role in potentially causing male hybrid sterility. Exactly what this role might be within species and how it might cause male hybrid sterility is still unknown.

The finding that *OdsH* is related to a known transcription factor (*unc-4*) and the finding that the DNA-binding determinant has evolved rapidly led to the proposal of a 'transcriptional incompatibility' model (Nei, Zhang 1998). Under this hypothesis, incompatibilities between the *OdsH* binding-specificity and sequence evolution of cis-acting sequence at the promoter/enhancer sequences of male meiotic (euchromatic) genes provide the underlying cause of hybrid sterility. This model is attractive because its been well documented that there

is significant transcriptional misregulation in hybrids, particularly in gonadal tissues (Meiklejohn et al. 2003; Michalak, Noor 2003; Michalak, Noor 2004). However, despite the initial attractiveness of a transcription-mediated sterility model, it raised several concerns.

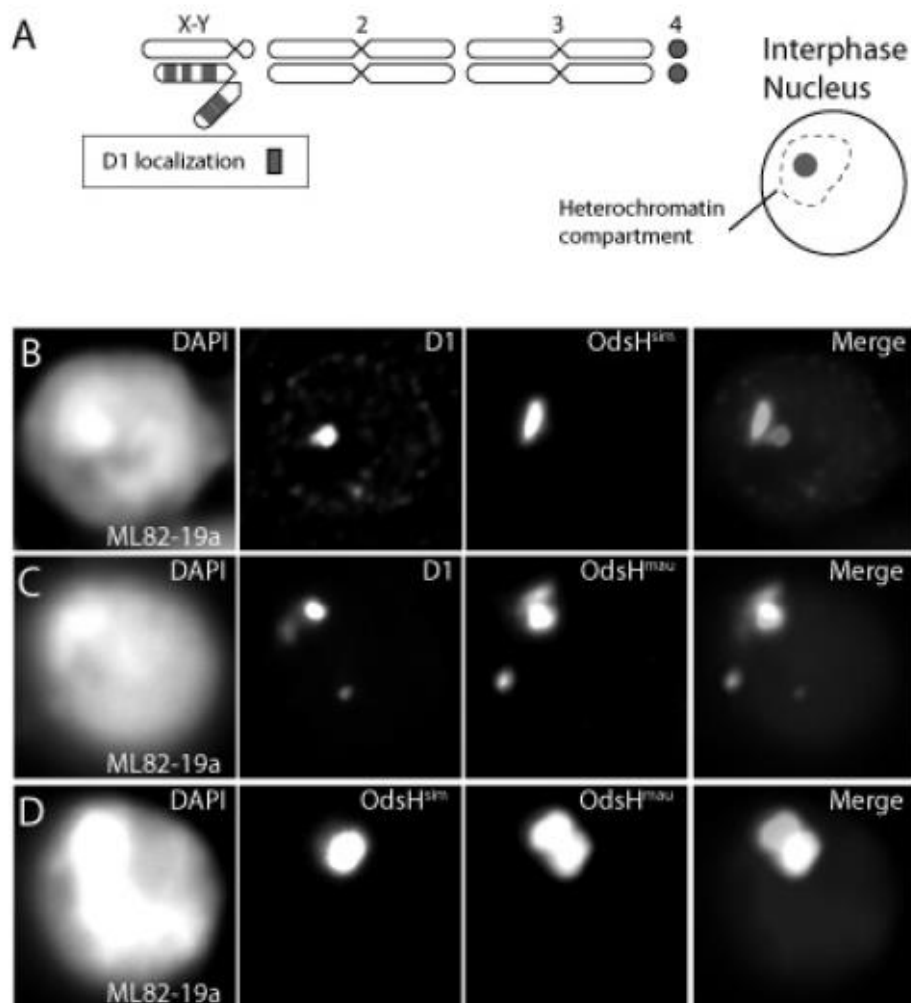


Figure 4. Cytological localization of OdsH protein in *D. simulans* cell lines reveals a satellite DNA-binding function.

First, it was unclear why the protein: DNA interface of a gene important in male meiosis would tolerate rapid evolution. Indeed, one might expect such changes would be so detrimental that they would not have survived natural selection. More dramatically, it is even harder to favor gene regulation models that could account for the rapid evolution of this protein: DNA interface. Although there is now precedent for positive selection shaping at least some transcription factors and cis-regulatory elements (Landry et al. 2005; Haygood et al. 2007; Chen et al. 2012), progress at elucidating the molecular function of *OdsH* and the biological basis by which it causes male sterility was slow despite the identification of *OdsH* and the proposal of

the ‘transcriptional incompatibility’ model. This is especially surprising because for many years, *OdsH* was the only known hybrid sterility gene identified in *any* animal taxa.

An alternative ‘genetic conflict’ model posited that *OdsH* encodes a protein that binds to evolutionarily labile components of the genome, such as satellite DNAs found at centromeric or heterochromatic regions. In contrast to the ‘transcriptional incompatibility’ model, hypotheses of segregation distortion or centromere-drive can readily account for the rapid evolution of satellite-DNA sequences, and subsequently the rapid evolution of satellite-DNA binding proteins. Under this scenario, conflict between satellite DNA sequences and proteins that bind them provide the impetus for divergence of hybrid sterility genes. This scenario would predict that *OdsH* encodes a satellite-DNA binding protein. This model would further predict that *OdsH* evolution in *D. mauritiana* versus *D. simulans* should result in altered localization of OdsH binding in the *D. simulans* genome.

We previously tested the idea that *OdsH* encodes a satellite-DNA binding protein by expressing *OdsH^{sim}* fused to a FLAG epitope in a male *D. simulans* embryonic cell culture line (Bayes, Malik 2009). The punctate localization pattern of *OdsH^{sim}* was reminiscent of chromatin-binding proteins that localize to specific pericentric repetitive DNA sequences, in particular the AT-rich satellite binding protein D1. Antibody detection of the D1 protein, which is also known to localize exclusively to repetitive sequences on the *D. simulans* Y and 4th chromosomes (Figure 4A), provided a cytological ‘marker’ for comparing OdsH localization (Aulner et al. 2002). Indeed, immunostaining of D1 showed a distinct signal just adjacent to *OdsH^{sim}* localization (Figure 4B). In contrast, expression of an *OdsH^{mau}* fusion protein showed partial overlap with the D1 signal in *D. simulans* cells (Figure 4C). Co-expression *OdsH^{sim}* and *OdsH^{mau}* revealed that the two proteins share localization to a common site, but that *OdsH^{mau}* has an additional localization preference (Figure 4D). These experiments suggested that both *OdsH^{sim}* and *OdsH^{mau}* both encode satellite-binding proteins, which have altered localization specificities in the *D. simulans* genome.

Transgenic *D. simulans* fly lines that expressed N-terminal fusion proteins of 3xFLAG-*OdsH^{sim}* and Venus (Yellow Fluorescent Protein)-*OdsH^{mau}* allowed further study of the localization specificities of the *OdsH^{sim}* and *OdsH^{mau}* proteins in a system that allowed better chromosomal resolution. Visualization of localization patterns of the OdsH proteins in *D. simulans* larval neuroblast cells, whose chromosomes are often condensed by virtue of undergoing frequent cell division, provided an opportunity to map protein localization to individual chromosomes. These studies revealed that *OdsH^{sim}* localized to the X and 4th chromosome in larval neuroblasts, confirming the association of OdsH with repeat-rich regions of the *D. simulans* genome. In significant contrast, *OdsH^{mau}* shows gross localization to the *D. simulans* Y chromosome, which is also gene-poor and repeat-rich, while sharing localization with *OdsH^{sim}* to the pericentric region of the X chromosome and the 4th chromosome. Furthermore, the sites of OdsH localization displayed altered chromatin morphology and, in particular, the 4th and Y chromosomes show chromatin decondensation in both pure species and hybrid individuals.

To study the endogenously expressed *OdsH* function at a molecular level, we raised an antibody to the OdsH protein for cytological characterization of the OdsH protein in testes. Whole-mount immunochemistry of OdsH in *D. simulans* testes revealed that OdsH localization is punctate (2 dots representing the 4th and X chromosome) and is restricted to spermatocytes that have completed their 4 sequential mitotic divisions and are now increasing in cell volume as the chromosomes prepare for meiotic division. In the fertile introgression, OdsH staining is

also punctate within the pre-meiotic zone of the male testis (Figure 5A) similar to the wild-type, 'pure-species *D. simulans* case. In contrast, in the sterile introgression line, three dots including a new, intense staining on a third dot are observed (more easily visualized in Figure 5B, inset), which is presumed to be the *D. simulans* Y chromosome based on the earlier cytological characterization. Furthermore, the OdsH staining extends much beyond the testes in pure-species of fertile introgression *D. simulans* males. This indicates a dramatic expansion of the pre-meiotic zone in the sterile introgression testis, implying the activation of a meiotic checkpoint likely as a result of chromosomal abnormalities resulting from negative epistasis between *OdsH^{mau}* and the *D. simulans* genome.

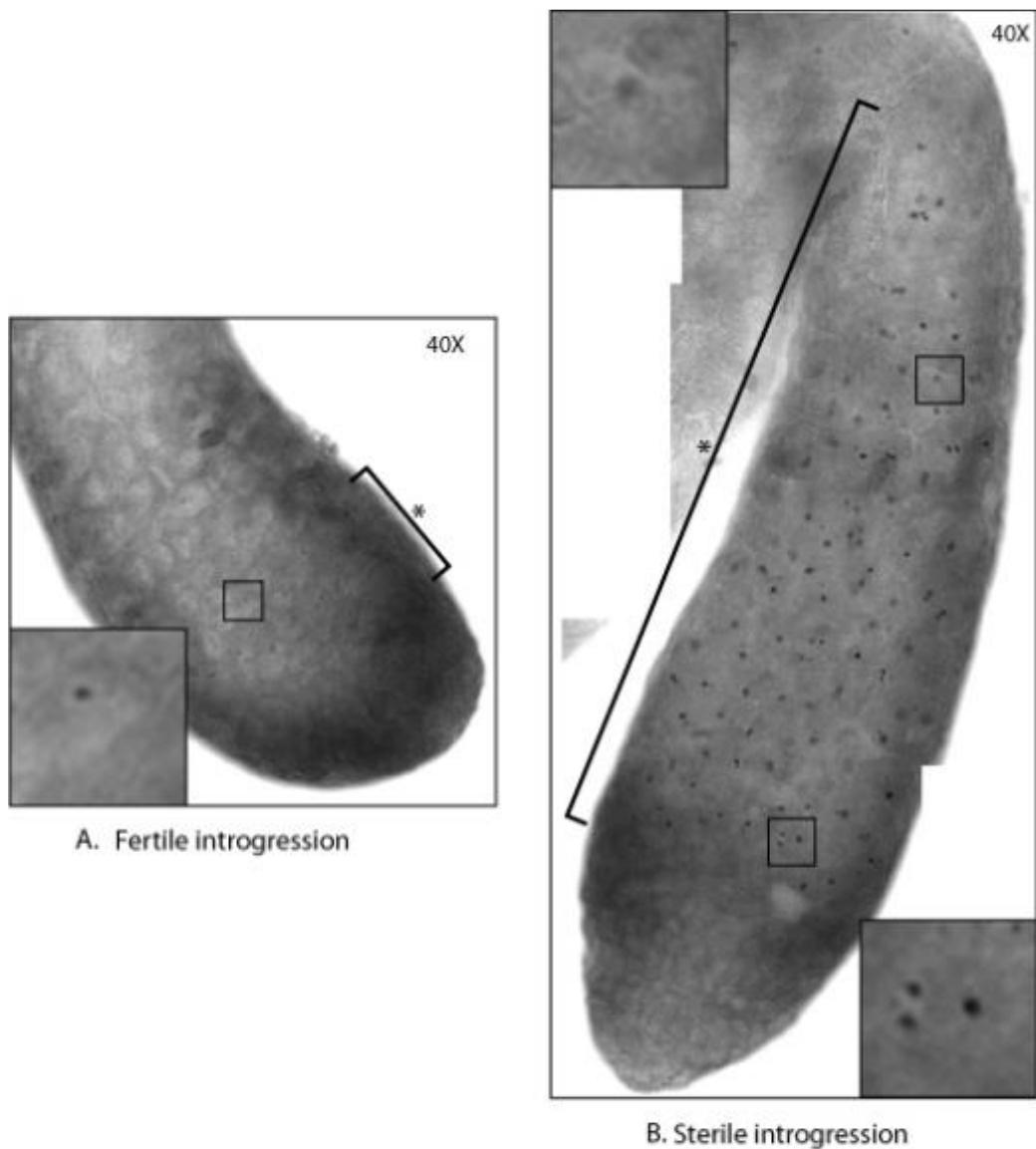


Figure 5. OdsH antibody staining in whole mount testes of (A) fertile and (B) sterile introgressions.

Together, these studies clearly established that in contrast to the prevailing model, aberrant satellite-DNA binding underlies hybrid sterility. Moreover, the lack of endogenous *OdsH* expression in *D. mauritiana* testes further established that sterility does not result from the impairment of some necessary *OdsH* function, in transcription or otherwise. Instead, it clearly established a gain-of-function of *OdsH^{mau}* protein aberrantly binding and decondensing the *D. simulans* Y chromosome as the primary cause behind hybrid sterility. This study represented one of the first detailed characterizations of the molecular basis of hybrid male sterility caused by a genic incompatibility. It made clear that the emerging pattern of DNA-binding genes in hybrid incompatibilities may have little to do with transcriptional regulation and, instead, established the role of rapidly evolving heterochromatin and heterochromatin-binding proteins in the evolution of intrinsic reproductive isolation.

OVERDRIVE

Drosophila pseudoobscura pseudoobscura (hereafter referred to as the USA subspecies) is distributed across western North America, and is most famous as the subject of Dobzhansky's classic research on the fitness effects of chromosome rearrangements (Dobzhansky 1937). *Drosophila pseudoobscura bogotana* (hereafter referred to as Bogota) is found in regions of high elevation near Bogota, Colombia (Dobzhansky 1963) and is geographically separated from the USA subspecies by more than 2000km (Figure 6). The study of the genetic basis of hybrid male sterility in the Bogota-USA hybridization has a long history. The presence of *D. pseudoobscura* near Bogota, Colombia was known since DOBZHANSKY (1963), but these were thought to be a part of the main species range of *D. pseudoobscura*. However, hybrid male sterility between these taxa was documented almost a decade later (PRAKASH 1972). We now know that the USA and Bogota subspecies are young allopatric subspecies, estimated to have diverged between 150,000 and 230,000 years (Schaeffer, Miller 1991; Wang, Wakeley, Hey 1997). There is little prezygotic isolation between these subspecies (Noor 1995). The USA and Bogota subspecies are among the youngest studied hybridizations and provide a rare opportunity to understand the earliest steps in speciation.

Crosses between Bogota females and USA males produce F₁ hybrid males that are nearly completely sterile, whereas F₁ hybrid females and hybrids of both sexes from the reciprocal cross are completely fertile (Prakash 1972). Maternal effect was initially thought to play a large role in hybrid sterility (DOBZHANSKY 1974; ORR 1989a; ORR 1989b), but was later shown to be unimportant (Orr and Irving 2001). The large, essential role of the right arm of the Bogota X chromosome (XR) on hybrid sterility was first established through the use of the SR chromosome and the molecular marker *Esterase-5* (PRAKASH 1972). This result was largely forgotten until the *sepia* region on Bogota XR was shown to be essential for hybrid male sterility (ORR and IRVING 2001). Hybrid sterility appears to involve a single complex genetic incompatibility in which all loci are essential for the manifestation of full sterility. An interaction of Bogota alleles at loci on the right and left arms of the X chromosome (XR and XL, respectively) with dominant USA alleles on the second and third autosomes is necessary to cause hybrid sterility (Orr, Irving 2001). None of the important regions shows much effect individually on hybrid sterility because all interacting partners must be simultaneously present for the full expression of hybrid sterility. Because these genes are essential for sterility, they

represent the first Dobzhansky-Muller incompatibility to evolve between these populations. That is, these genes do not represent a mere additional layer of incompatibility on top of others; they are important *during* speciation.

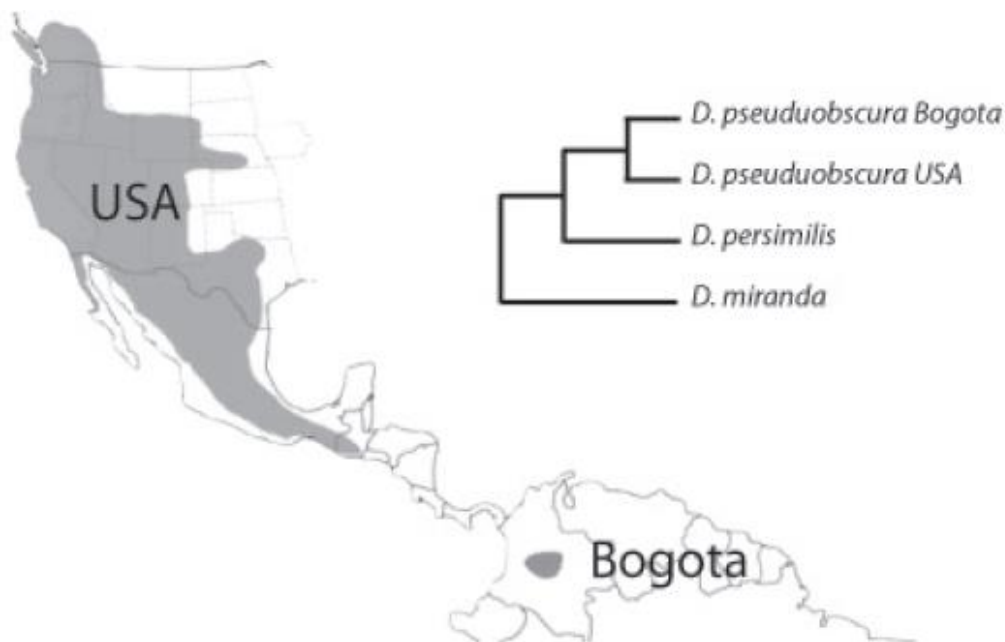


Figure 6. Bogota and USA are young allopatric populations of *Drosophila pseudoobscura*.

This point is especially important because theory has shown that hybrid incompatibilities that underlie reproductive barriers should accumulate between taxa faster than linearly with divergence, a pattern known as the “snowball effect” (Orr 1995; Matute et al. 2010; Moyle, Nakazato 2010). This rapid accumulation of genic incompatibilities continues even after reproductive isolation is complete. Therefore, in older taxa, this process can confound the interpretation of the number of genes required to cause intrinsic reproductive isolation. Despite the identification of a few individual genes that are involved in hybrid incompatibilities, we do not know how many partner genes typically interact within a *single* incompatibility. Also, although a single hybrid incompatibility can, in principle, cause postzygotic isolation, it is difficult to estimate how many independent incompatibilities typically separate species. Finally, it is difficult to say which incompatibilities were important during speciation *versus* those that accumulated after the attainment of complete reproductive isolation. Mapping and identifying the interacting partner genes that form Dobzhansky-Muller incompatibilities in evolutionarily *young* hybridizations is, therefore, critical to understanding the genetic architecture of speciation.

Even though the sterility of hybrids between the *Drosophila pseudoobscura* Bogota and USA subspecies has been known for over 30 years, a study surprisingly found that the “sterile” F₁ hybrid males become very weakly fertile when aged and produce progeny that are almost all daughters (Figure 7) (Orr, Irving 2005). This sex-ratio distortion is not caused by hybrid inviability but by an overrepresentation of X-bearing sperm among the functional gametes: X-bearing sperm from F₁ hybrid males fertilize many more eggs than do Y-bearing sperm (Orr,

Irving 2005). The precise mechanism of segregation distortion is not well understood and may involve true meiotic drive (which acts during meiosis) or gamete killing (which acts after meiosis). In either case, the “cheating” *X* chromosome enjoys an evolutionary advantage. The genetic basis of segregation distortion in these hybrids is similar to that of hybrid sterility: Bogota alleles at loci on the right and left arms of the *X* chromosome (*XR* and *XL*, respectively) interact with dominant USA alleles on the second and third autosomes. Early genetic mapping had implicated the same region on Bogota *XR* in both hybrid phenomena (sterility and segregation distortion). The most important question was whether the two phenomena of segregation distortion and hybrid male sterility were related to each other. If so, this would provide evidence consistent with the long hypothesized role of genetic conflict involving segregation distorters in the evolution of hybrid sterility (Frank 1991; Hurst, Pomiankowski 1991). *Drosophila pseudoobscura* USA and Bogota provide a powerful system for answering not only questions about the genetic architecture of hybrid incompatibilities, but also the biological forces that drive the evolution of hybrid incompatibility genes.

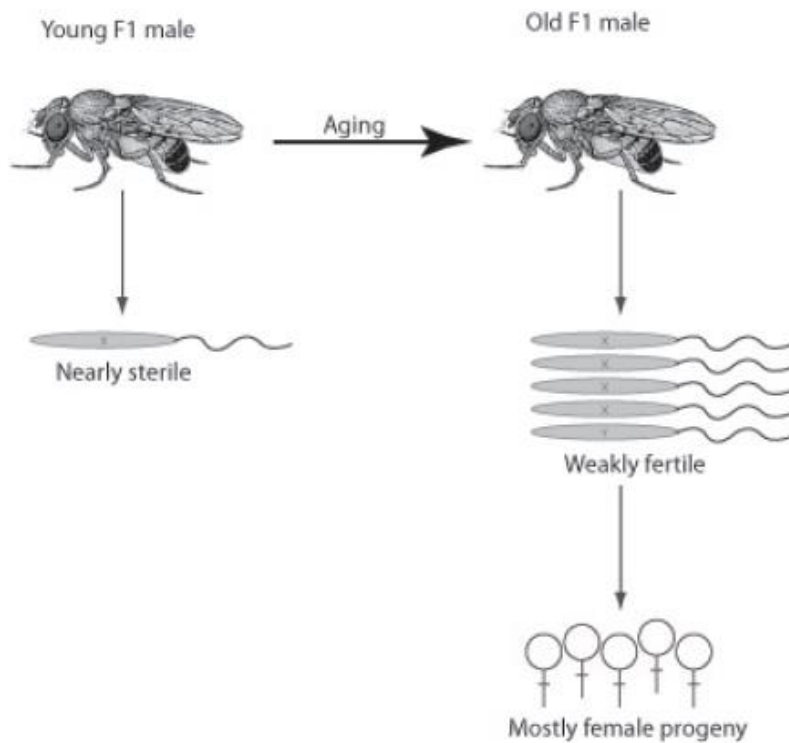


Figure 7. “Sterile” hybrid males become weakly fertile when aged and produce mostly daughters.

To explore the relationship between segregation distortion and hybrid sterility, we fine mapped the genes responsible for sterility and segregation distortion within the *sepia* (*se*) region on *XR* (Phadnis, Orr 2009). 175 independent introgression lines were generated in which the USA *sepia* region was moved into an otherwise pure Bogota background. The resulting introgression lines, backcrossed for 28 generations, have genomes that derive almost entirely from Bogota except for a small chromosomal region near *sepia*. When females from these introgression lines are crossed to USA males, the resulting hybrid males are genetically

identical to F₁ hybrid males except for the small chromosomal region introgressed from USA. All 175 lines yielded hybrid males that were both fertile and produced normal (~50:50) sex ratios by virtue of now possessing USA alleles instead of the sterility-associated Bogota alleles. This defined a region that harbors the hybrid sterility and hybrid segregation distortion gene(s) (Figure 8).

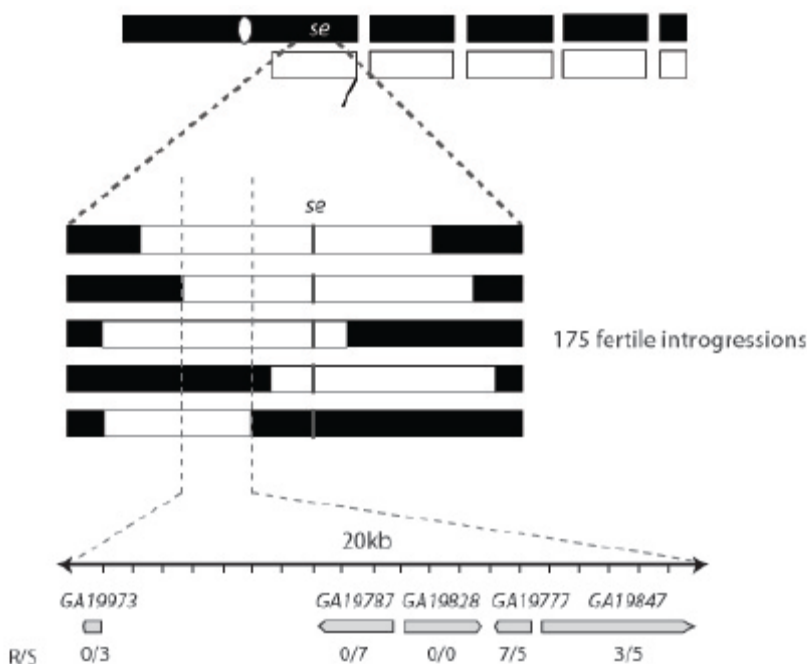


Figure 8. Introgression mapping identifies the genetic region responsible for both hybrid sterility and segregation distortion.

To further refine the genetic mapping, Bogota females heterozygous for an introgression were crossed to USA males and *se*⁺ males were tested for ability to produce progeny. One such fertile *se*⁺ male was found; this male also produced progeny in an even sex ratio ('rescuing' the hybrid sterility phenotype). Genotyping this line localized the gene(s) causing hybrid sterility and hybrid segregation distortion to a 20kb region that includes only five predicted loci (Figure 8). This entire 20kb region from the Bogota subspecies was sequenced and compared to the homologous USA sequence from the *D. pseudoobscura* (USA) genome (Richards et al. 2005). Three of the five predicted genes—*GA23450*, *GA19787* and *GA19828*—showed no non-synonymous differences, while *GA23843* showed three non-synonymous differences. The predicted gene *GA19777*, however, showed eight non-synonymous changes, a surprising number given its small coding region (591 bp, excluding one 74 bp intron). Given that genes causing reproductive isolation often evolve rapidly (Presgraves 2010), *GA19777* (later named *Overdrive*), which features a DNA-binding motif, represented the best candidate for the cause of hybrid sterility and/or hybrid segregation distortion.

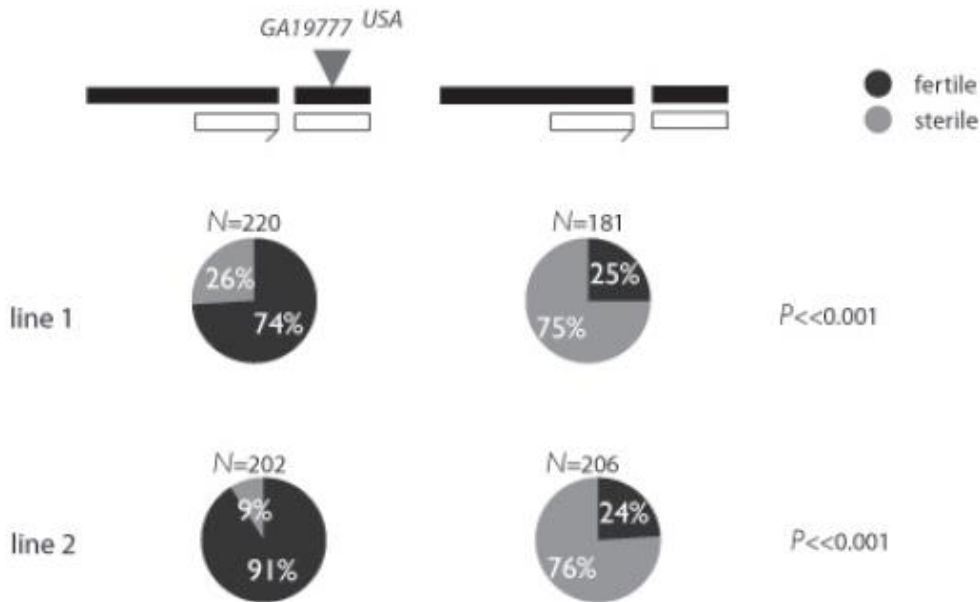


Figure 9. The USA allele of *Overdrive* (*GA19777*) rescues hybrid fertility.

To confirm whether *Overdrive* (*Ovd*) caused these hybrid phenotypes, transgenic experiments were performed to attempt to rescue the fertility of F_1 hybrid males with a transgenic copy of the (fertile) USA allele of *Ovd* (hereafter *Ovd*^{USA}). Males that inherit both transgenic *Ovd*^{USA} and endogenous *Ovd*^{Bog} are far more fertile and produce progeny significantly more often than those that inherit only the endogenous *Ovd*^{Bog} (Figure 9). Surprisingly, *Ovd*^{USA} transgenic hybrids, which rescue fertility weakly, still produce almost all daughters. Thus, while *Ovd*^{USA} transgenes rescue the hybrid sterility effect, they do not suppress segregation distortion. However, in the opposite direction, transgenes of *Ovd*^{Bog} resulted a strikingly female-biased sex ratio in the appropriate genetic background, whereas control males that inherit only the endogenous *Ovd*^{USA} produce normal sex ratios, clearly implicating the *Ovd*^{Bog} allele as playing a role in hybrid segregation distortion (Figure 10). Thus, the same allele of *Ovd*^{Bog} causes both hybrid sterility and segregation distortion. *GA19777* was named *Overdrive* (*Ovd*) because of its role in both segregation distortion (“meiotic drive”) and hybrid sterility. Interestingly, the effects of *Ovd* transgenes on hybrid sterility and segregation distortion are observed even in *trans*. For example, *Ovd*^{Bog} transgenes inserted on the autosomes can cause segregation distortion of the X chromosome in hybrids.

Ovd is predicted to encode a polypeptide with a single MADF DNA-binding domain near its C-terminus. As expected given its phenotype, RT-PCR revealed that *Ovd* is expressed in the testes of both pure subspecies males and in sterile F_1 hybrid males. Evolutionary analyses showed that all seven non-synonymous changes and all five synonymous changes fixed between Bogota and USA *Ovd* alleles occurred in the Bogota lineage — the lineage whose X chromosome causes hybrid sterility and hybrid segregation distortion. This significant enrichment of substitutions indicates accelerated evolutionary changes in the Bogota lineage, consistent with a history of segregation distortion at the *Ovd* locus in Bogota. The most profound aspect of the *Ovd* study was the demonstration that the same gene, *Ovd*, causes both

F₁ segregation distortion and hybrid sterility in evolutionarily young taxa that obey Haldane’s rule. This study provided the first direct evidence that genetic conflict is an important force in the evolution of postzygotic isolation.

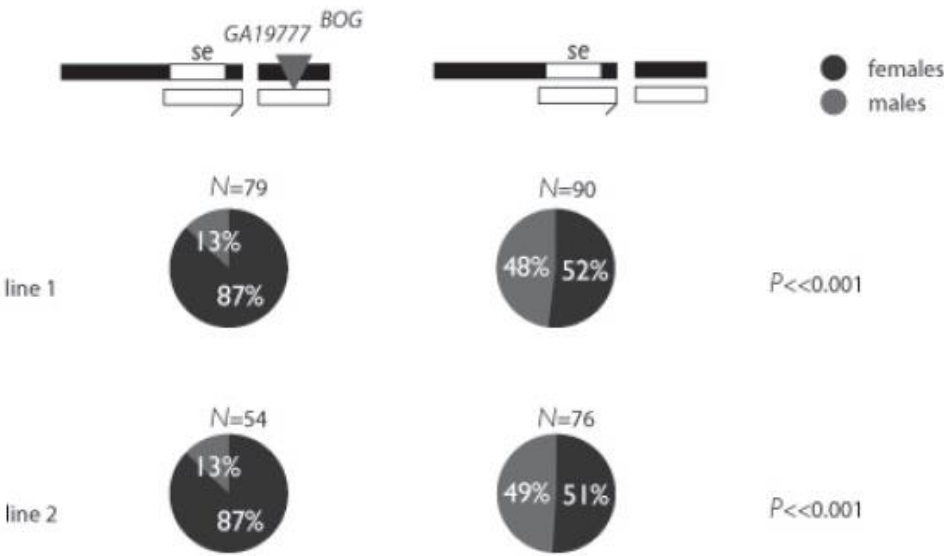


Figure 10. The Bogota allele of *Overdrive* (*GA19777*) causes segregation distortion.

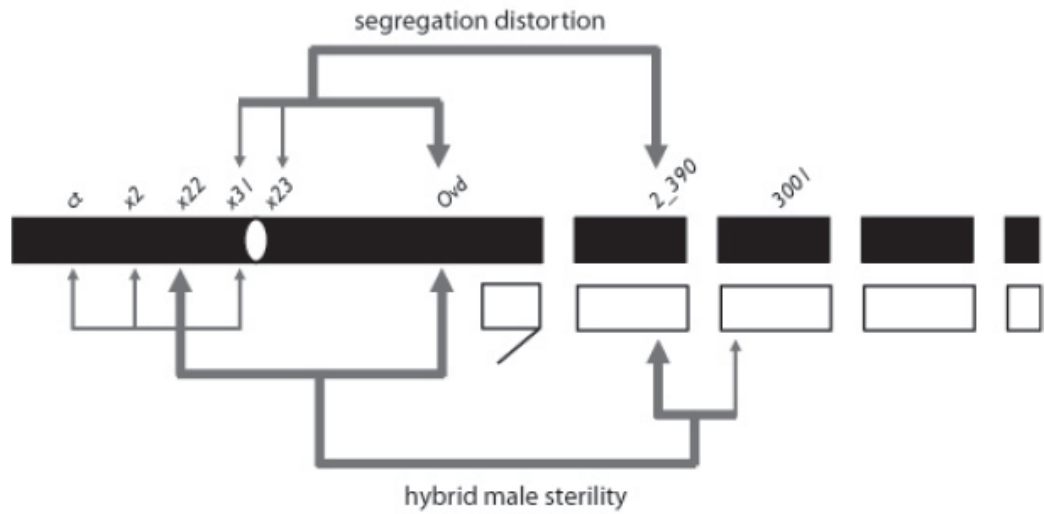


Figure 11. The genetic architecture of hybrid sterility and segregation distortion in Bogota-USA hybrids.

While *Ovd* is necessary to cause hybrid male sterility or segregation distortion, it is not sufficient by itself to cause either phenomenon. That is, it requires “correct” alleles at several interacting loci to yield complete hybrid sterility or segregation distortion. How many such interacting loci are involved in the hybrid incompatibility responsible for male sterility? This

question remained unanswered even in the context of loci on the Bogota *X* – the best characterized chromosome in this hybridization. Questions also remained about the genetic architecture of segregation distortion in hybrid males and its suppression in pure Bogota males. More important, are all loci involved in hybrid sterility also involved in segregation distortion, as in the case of *Ovd*?

To address these questions, we performed genome-wide mapping of all loci involved in hybrid male sterility and segregation distortion (Phadnis 2011). Separate crosses were performed to map the factors on the *X* chromosome and the autosomes. The crosses for mapping the factors on the *X* chromosome controlled for the effects of *Ovd* to maximize mapping power for the loci that interact with *Ovd*. The crosses for mapping the factors the autosomes used a *sepia* introgression line analogous to a ‘fertility rescue mutation’ to maximize mapping power for dominantly acting loci that interact with *Ovd*. These analyses reveal a coherent picture that implicates an interaction of *Ovd* with only two major effect loci (one on Bogota *XL* and another on the second chromosome) and with a few minor modifier loci) in a single dominant hybrid incompatibility as the genetic basis of hybrid sterility (Figure 11). This is one of the most comprehensively described genetic network of Dobzhansky-Muller incompatibilities in speciation, and lays the foundation for fine mapping experiments to identify *all* the genes that interact with *Ovd* to cause sterility in hybrid F1 males between the Bogota and USA subspecies of *D. pseudoobscura*. Since postzygotic isolation between Bogota and USA involves a single incompatibility with a modest number of interacting loci, identification and characterization of these genes will provide important insights into the molecular nature of postzygotic isolation between species.

The genetic architecture of segregation distortion and its suppression is even simpler: *Ovd* interacts with only two loci on the Bogota *X*-chromosome to cause segregation distortion, which is nearly completely suppressed by a single locus on the Bogota second chromosome. Most important, the loci that cause hybrid males sterility and segregation distortion are partially – but not completely – overlapping. Of particular interest is the *X22* locus, which appears to be required for hybrid male sterility but not for segregation distortion. If the effects of *X22* on segregation distortion and hybrid sterility prove to be separable, then the *X22* locus can be represented as the genetic switch that converts a segregation distortion system *within* species into a hybrid sterility system *between* species. Identifying the gene that underlies this tipping point from segregation distortion to hybrid sterility holds the promise of providing insight into the events that lead to reproductive isolation between incipient species and explaining with greater clarity how segregation distortion can set the stage for speciation.

CONCLUSION

The studies of *Odysseus* and *Overdrive* provide largely complementary insights into the molecular and evolutionary basis of hybrid sterility. The cell biological study of *Odysseus* – the first hybrid sterility gene to be discovered – has provided one of the first hints into the molecular nature of hybrid sterility and into the mechanisms of hybrid dysfunction. In particular, localization studies of alternative alleles of *OdsH* in cell culture and in flies have revealed how evolutionary gains of function in the context of DNA-binding specificities can drive functional divergence of genes that cause hybrid incompatibility. Further, the chromatin

decondensation defects caused by *OdsH* provide a molecular view into the mechanisms of hybrid dysfunction and highlight the role of the dynamic evolution of heterochromatin in intrinsic postzygotic isolation.

While *Odysseus* has provided a strong study case for the functional and molecular aspects of hybrid sterility, questions about the genetic architecture of the hybrid incompatibility that *OdsH* participates in and the biological forces that drove its rapid evolution remain open. First, it is unclear how many genes *OdsH* interacts with to cause hybrid sterility. *OdsH* alone is insufficient to cause complete hybrid sterility and needs to interact with other *X*-linked genes included in the introgression from *D. mauritiana*. The sterile introgression that harbors *OdsH* is present in an otherwise *D. simulans* background. Therefore, there may be additional *D. simulans* components that may also be involved in this incompatibility. This also makes it difficult to say whether the *OdsH* incompatibility is recessive (contributes to the sterility of F2-like hybrids) or dominant (contributes to F1 hybrid male sterility). Second, the findings of *OdsH* chromosomal localization, the apparent antagonism of heterochromatin by chromatin decondensation, and protein expression patterns in the testes lead to possible genetic conflict scenarios in causing male hybrid sterility. One scenario posits that the *X*-linked *OdsH* protein may participate in segregation distortion in males, recognizing a *Y*-satellite to cause decondensation and, thereby, the failure of *Y*-bearing sperm. Due to the deleterious effects of this *X*-drive (reduced fertility, skewed sex ratios), autosomal and *Y*-linked suppressors would arise in each population. However, introgression of *OdsH^{mau}* into the *D. simulans* genome (as seen in hybrids) may unleash hidden distortion that had been successfully suppressed in *D. simulans* (Frank 1990; Hurst, Pomiankowski 1991). While few clues currently explain the exact biological processes that drove the rapid evolution between *D. mauritiana* and *D. simulans*, further characterization of the sites of localization of *OdsH* and the meiotic defects in the sterile hybrid testes may help differentiate alternative models.

In contrast to *Odysseus*, the biological force that drove the evolution of *Overdrive* is more obvious. The idea that genetic conflict between segregation distorters and its suppressors may drive the evolution of hybrid sterility between populations is very intuitive, but the lack of empirical evidence hindered its mainstream acceptance. The demonstration that *Ovd* causes both segregation distortion and hybrid male sterility in the *D. pseudoobscura* USA-Bogota hybrids finally connected the two phenomena in the context of speciation. The simplest versions of the theory that segregation distortion can drive the sterility of inter-species hybrids often invoked at least two independent segregation distortion systems in two populations. For example, *X*-chromosome distortion suppressed in Bogota and *Y*-chromosome distorters suppressed in USA can both get de-repressed in their F1 hybrids. The Bogota *X*-chromosome distorter may kill all *Y*-bearing sperm and the USA *Y*-chromosome distorter may kill all *X*-bearing sperm, thus rendering the hybrid sterile. We can now reject this simplest version of the theory in the case of the USA-Bogota hybrids: while the Bogota *X*-chromosome participates in segregation distortion, there is no evidence of segregation distortion carried out by the USA *Y*-chromosome. How might a single segregation distortion system cause the complete sterility of hybrids? One plausible explanation is that, in hybrids, the strength of the *X*-chromosome distortion may be beyond levels ever experienced during the repeated bouts of the arms race between segregation distorter and suppressor alleles in the evolutionary history of the Bogota lineage. This may result in the killing of not only the sensitive *Y*-bearing sperm, but also the normally more resistant *X*-bearing sperm, leading to the near complete sterility of the hybrid males.

A comprehensive picture of the genetic architecture of hybrid sterility and segregation distortion may help address the above question from a molecular viewpoint. Our genome-wide analyses revealed that the single hybrid incompatibility that isolates USA and Bogota consists of only a small number of interacting components, making it possible to identify the complete complement of genes essential for reproductive isolation (Phadnis 2011). The loci essential for hybrid sterility nearly overlap with those that are involved in segregation distortion and its suppression. Identifying the apparent exception to this overlap – the X22 locus on the Bogota X-chromosome, which appears to be essential for hybrid sterility but not for segregation distortion – may hold the key to addressing how a single segregation distortion system within a population may cause complete sterility in hybrids.

While the role of *Odysseus* in binding heterochromatin is now demonstrated clearly, the molecular function of *Overdrive* and the mechanism by which it causes segregation distortion remains an open line of inquiry. Because *Ovd* is involved in specifically killing Y-bearing sperm, it appears likely that it may function through binding to heterochromatin, perhaps Y-specific heterochromatin, in a mechanism that is analogous to that of *Odysseus*. Ongoing functional studies to characterize the function of *Ovd* within species and its mechanism of hybrid sterility between species may reveal whether *Odysseus* and *Overdrive* are fundamentally similar in their causes of evolution and their mechanisms of causing reproductive isolation.

Given the relative paucity of genes directly implicated in hybrid sterility in animals, the molecular insights obtained from *Odysseus* and the evolutionary genetic insights obtained from *Overdrive* will not only help elucidate each other's underlying biology but may also reveal potentially common themes associated with hybrid sterility in other systems.

REFERENCES

- Aulner, N, C Monod, G Mandicourt, D Jullien, O Cuvier, A Sall, S Janssen, UK Laemmli, E Kas. 2002. The AT-hook protein D1 is essential for *Drosophila melanogaster* development and is implicated in position-effect variegation. *Mol Cell Biol* 22:1218-1232.
- Bayes, JJ, HS Malik. 2009. Altered Heterochromatin Binding by a Hybrid Sterility Protein in *Drosophila* Sibling Species. *Science* 326:1538-1541.
- Chen, S, X Ni, BH Krinsky, YE Zhang, MD Vibranovski, KP White, M Long. 2012. Reshaping of global gene expression networks and sex-biased gene expression by integration of a young gene. *Embo Journal* 31:2798-2809.
- Coyne, JA, HA Orr. 2004. *Speciation*. Sunderland, MA: Sinauer Associates.
- Dobzhansky, T. 1937. *Genetics and the origin of species*. New York,: Columbia Univ. Press.
- Dobzhansky, T, Hunter, A. S., Pavlovsky, B., Wallace, B. 1963. Genetics of natural populations. XXXI. Genetics of an isolated marginal population of *Drosophila pseudoobscura*. *Genetics* 48:91-103.
- Frank, SA. 1990. Divergence of meiotic drive-suppression systems as an explanation for sex-biased hybrid sterility and inviability. *Evolution* 45:262-267.
- Frank, SA. 1991. Divergence of Meiotic Drive-Suppression Systems as an Explanation for Sex-Biased Hybrid Sterility and Inviability. *Evolution* 45:262-267.

- Haygood, R, O Fedrigo, B Hanson, KD Yokoyama, GA Wray. 2007. Promoter regions of many neural- and nutrition-related genes have experienced positive selection during human evolution. *Nat Genet* 39:1140-1144.
- Hurst, LD, A Pomiankowski. 1991. Causes of sex ratio bias may account for unisexual sterility in hybrids: a new explanation of Haldane's rule and related phenomena. *Genetics* 128:841-858.
- Hutter, P, M Ashburner. 1987. Genetic rescue of inviable hybrids between *Drosophila melanogaster* and its sibling species. *Nature* 327:331-333.
- Kliman, RM, P Andolfatto, JA Coyne, F Depaulis, M Kreitman, AJ Berry, J McCarter, J Wakeley, J Hey. 2000. The population genetics of the origin and divergence of the *Drosophila simulans* complex species. *Genetics* 156:1913-1931.
- Landry, CR, PJ Wittkopp, CH Taubes, JM Ranz, AG Clark, DL Hartl. 2005. Compensatory cis-trans evolution and the dysregulation of gene expression in interspecific hybrids of *Drosophila*. *Genetics* 171:1813-1822.
- Maheshwari, S, DA Barbash. 2011. The Genetics of Hybrid Incompatibilities. *Annual Review Genetics*, Vol 45 45:331-355.
- Masly, JP, CD Jones, MAF Noor, J Locke, HA Orr. 2006. Gene transposition as a cause of hybrid sterility in *Drosophila*. *Science* 313:1448-1450.
- Matute, DR, IA Butler, DA Turissini, JA Coyne. 2010. A test of the snowball theory for the rate of evolution of hybrid incompatibilities. *Science* 329:1518-1521.
- Mayr, E. 1942. *Systematics and the origin of species from the viewpoint of a zoologist*. New York,: Columbia University Press.
- Meiklejohn, CD, J Parsch, JM Ranz, DL Hartl. 2003. Rapid evolution of male-biased gene expression in *Drosophila*. *Proc Natl Acad Sci U S A* 100:9894-9899.
- Michalak, P, MA Noor. 2003. Genome-wide patterns of expression in *Drosophila* pure species and hybrid males. *Mol Biol Evol* 20:1070-1076.
- Michalak, P, MA Noor. 2004. Association of misexpression with sterility in hybrids of *Drosophila simulans* and *D. mauritiana*. *J Mol Evol* 59:277-282.
- Moyle, LC, T Nakazato. 2010. Hybrid incompatibility "snowballs" between *Solanum* species. *Science* 329:1521-1523.
- Nei, M, J Zhang. 1998. Molecular origin of species. *Science* 282:1428-1429.
- Noor, MA. 1995. Incipient sexual isolation in *Drosophila pseudoobscura bogotana*. *Pan-Pacific Entomologist* 71:125-129.
- Orr, HA. 1995. The population genetics of speciation - the evolution of hybrid incompatibilities. *Genetics* 139:1805-1803.
- Orr, HA, S Irving. 2001. Complex epistasis and the genetic basis of hybrid sterility in the *Drosophila pseudoobscura* Bogota-USA hybridization. *Genetics* 158:1089-1100.
- Orr, HA, S Irving. 2005. Segregation distortion in hybrids between the Bogota and USA subspecies of *Drosophila pseudoobscura*. *Genetics* 169:671-682.
- Perez, DE, CI Wu, NA Johnson, ML Wu. 1993. Genetics of reproductive isolation in the *Drosophila simulans* clade: DNA marker-assisted mapping and characterization of a hybrid-male sterility gene, *Odysseus* (Ods). *Genetics* 134:261-275.
- Phadnis, N. 2011. Genetic Architecture of Male Sterility and Segregation Distortion in *Drosophila pseudoobscura* Bogota-USA Hybrids. *Genetics* 189:1001-U1428.
- Phadnis, N, HA Orr. 2009. A single gene causes both male sterility and segregation distortion in *Drosophila* hybrids. *Science* 323:376-379.

- Prakash, S. 1972. Origin of reproductive isolation in the absence of apparent genic differentiation in a geographic isolate of *Drosophila pseudoobscura*. *Genetics* 72:143-155.
- Presgraves, DC. 2010. The molecular evolutionary basis of species formation. *Nat Rev Genet* 11:175-180.
- Richards, S, Y Liu, BR Bettencourt, et al. 2005. Comparative genome sequencing of *Drosophila pseudoobscura*: Chromosomal, gene, and cis-element evolution. *Genome Research* 15:1-18.
- Sawamura, K, T Taira, TK Watanabe. 1993. Hybrid Lethal Systems in the *Drosophila-Melanogaster* Species Complex .1. The Maternal-Hybrid-Rescue (Mhr) Gene of *Drosophila-Simulans*. *Genetics* 133:299-305.
- Sawamura, K, TK Watanabe, MT Yamamoto. 1993. Hybrid Lethal Systems in the *Drosophila-Melanogaster* Species Complex. *Genetica* 88:175-185.
- Schaeffer, SW, EL Miller. 1991. Nucleotide-sequence analysis of ADH genes estimates the time of geographic isolation of the Bogota population of *Drosophila pseudoobscura*. *Proceedings of the National Academy of Sciences of the United States of America* 88:6097-6101.
- Ting, CT, SC Tsauro, ML Wu, CI Wu. 1998. A rapidly evolving homeobox at the site of a hybrid sterility gene. *Science* 282:1501-1504.
- Wang, RL, J Wakeley, J Hey. 1997. Gene flow and natural selection in the origin of *Drosophila pseudoobscura* and close relatives. *Genetics* 147:1091-1106.
- Watanabe, TK. 1979. A gene that rescues the lethal hybrids between *Drosophila melanogaster* and *D. simulans*. *Jpn J. Genet.* 54:325–331.

Chapter 10

THE ROLE OF TRANSPOSABLE ELEMENTS IN SPECIATION

Eric T. Watson and Jeffery P. Demuth*

Department of Biology, The University of Texas at Arlington, Arlington, TX, US

ABSTRACT

With innovative DNA sequencing technologies has come a new appreciation for the content of animal and plant genomes. Overwhelmingly, a picture has been painted in which mobile and repetitive elements dominate the genomic landscape. Mobile genetic elements have recently been shown to contribute coding and regulatory sequences during their proliferation, leading to functional and regulatory novelty as well as element-mediated rearrangements coinciding with speciation events. Additionally, dormant elements occasionally erupt in bouts of excision and transposition in interspecific hybrids, resulting in a suite of maladaptive traits. The potential for mobile elements as key players in the evolution and diversification of genomes and species is immense yet in many respects transposable elements still remain the “dark matter” of the genome. This is particularly true of their role in speciation, and in order to fully appreciate their role, much work is still needed. In this chapter, we investigate the evidence for transposable elements as drivers of diversification and speciation.

INTRODUCTION

Transposable elements (TEs) are a diverse group of genetic sequences that share the common ability to move within a genome. The broadest distinction among TEs classifies them based on whether or not transposition involves an RNA intermediate (Wicker et al., 2007). Retrotransposons (class I TEs) transpose through a replicative “copy and paste” mechanism, which involves the production of a processed mRNA transcript that becomes re-inserted into the host genome after being reverse transcribed into complementary DNA by an element-

* Corresponding Author's Email: jpdemuth@uta.edu (Jeff Demuth, UTA- Dept. of Biol., 500 S. Nedderman Dr., Arlington TX 76019; Tel: (817) 272-2653; Fax (817) 272-2855.

encoded reverse transcriptase. In contrast, DNA transposons (class II TEs) rely on a non-replicative “cut and paste” mechanism, involving a diversity of element-encoded enzymes, such as transposase, C-integrase, and tyrosine recombinase. Both class I and class II TEs are prone to proliferative bursts and exist as either autonomous elements, which encode the proteins necessary to catalyze their own transposition; or non-autonomous elements, which use the replication machinery of autonomous TEs.

TEs are often considered “parasitic” and “selfish” due to their ability to invade a genome despite imposing a fitness cost to their hosts in the process. By inserting themselves into coding or regulatory regions, TEs often have deleterious consequences, but their impact varies among taxa. For instance, TE mobilization is associated with over 100 human diseases (Goodier and Kazazian 2008; Belancio, Hedges, and Deininger) but this represents a relatively low percentage of pathogenic mutations (~0.3%; (Callinan and Batzer 2006; Kazazian 1998). Whereas, in mouse and *Drosophila*, TE insertions constitute a much larger proportion of deleterious mutations (10% and 50% respectively; Maksakova et al. 2006; Finnegan 1992). The variation in how deleterious TEs are likely reflects an interaction between host genome defense and variation in the composition of TE types (Eickbush and Furano 2002).

While TEs are perhaps best known for their capacity to disrupt host gene function, speculation about their potential evolutionary benefits as drivers of diversity have been around almost since their discovery in the 1940s (McClintock 1950; 1956; Britten and Davidson 1971). However, it was not until the genomic era that the ubiquity, abundance and diversity of TEs has been fully appreciated (Kidwell and Lisch 2001; Fedoroff 2012). For example, variation in TE abundance between the genomes of different species reveals lineage specific dynamics in the composition and abundance of different classes of TEs. Recent large-scale sequencing projects and advances in the screening of genomic sequences for hallmarks of repetitive elements reveal that TEs comprise 10-90% of plant nuclear genomes (Li et al. 2004; Haas et al. 2005), up to 20% of fungal genomes (Wicker et al. 2007), 9% of the chicken genome (International Chicken Genome Sequencing Consortium 2004), 15-22% of the *Drosophila melanogaster* genome (Kapitonov and Jurka 2003; Biemont and Vieira 2005), 6-17% of the *Tribolium castaneum* genome (Wang et al. 2008), and 50-66% of the human genome (International Human Genome Sequencing Consortium 2001; Koning et al. 2011). However, given that identifying TEs and their copy number from genomic sequence is fraught with computational difficulties, some of these values likely remain underestimates (e.g., Koning et al. 2011). Across eukaryotes, TEs are a much better predictor of genome size than protein coding gene content (Feschotte and Pritham 2007a). In addition to their abundance, a handful of well-annotated genomes illustrate that the TEs present in eukaryotes are very diverse (Wicker et al. 2007; Mandal and Kazazian 2008; Venner et al. 2009; Jurka et al. 2011; Chénais et al. 2012).

The diversity of genetic novelty infused by TE invasion and proliferation has been a fundamental argument for their role in driving organismal diversity (reviewed in Fedoroff 1999; Kazazian 2004; Biémont and Vieira 2006; Oliver and Greene 2009; 2011) and there are a growing number of specific examples where TE domestication has contributed coding and/or regulatory sequences implicated in adaptive novelty (Brandt et al. 2005; Feschotte and Pritham 2007b; Feschotte 2008; Butter et al. 2010). Furthermore, the burst dynamics of TEs wherein they invade, rapidly proliferate, and then are silenced by host defense has been hypothesized to explain broad evolutionary patterns from geological time scales to contemporary biodiversity (Zeh et al. 2009; Oliver and Greene 2011; Table 1).

Table 1. Mobile genetic elements and speciation in geologic time

TE event/Species history	Reference
Reduced L1 and SINE accumulation during radiation of African apes (14–15 Mya).	Intl Hum Gen Seq Consort (2002)
Expansion of L1 subfamilies parallels intense speciation in <i>Rattus sensu stricto</i> .	Verneau <i>et al.</i> (1998)
Lx family amplification is concomitant to the radiation of murine mammals.	Pascale <i>et al.</i> (1990)
Rapid speciation in the genus <i>Taterillus</i> (gerbil) occurred along with intense activity of TEs in nascent lineages.	Dobigny <i>et al.</i> (2004)
Intense DNA elements transposition during the <i>Myotis</i> radiation.	Ray <i>et al.</i> (2008)
DNA transposon bursts parallel speciation events in pseudo tetraploid salmonids and occurred after genome duplication.	de Boer <i>et al.</i> (2007)
Acquisition and consequent transposition of an endogenous retrovirus (ERV) element and lineage specific enrichment of TEs in <i>Entamoeba histolytica</i> .	Lorenzi <i>et al.</i> (2008)
Repeated bouts of Haplochromine-specific SINE insertions followed by extensive radiations found in all inhabited lakes.	Shedlock <i>et al.</i> (2004)
Peak L2 and MIR activity coincides with marsupial-placental split 120-150 MYA.	Kim <i>et al.</i> 2004
Peak L1 activity corresponds to the eutherian radiation 100MYA.	Kim <i>et al.</i> 2004
Unprecedented LTR activity on the Y-chromosome corresponds to the K-T ecological disturbance 70 MYA	Kim <i>et al.</i> 2004
Alu and young L1 activity is restricted to the radiation of Old World and New World monkeys 40 and 25 MYA, respectively.	Kim <i>et al.</i> 2004
Diverse species exhibit different numbers of TE families between subpopulations with relatively low amounts of sequence divergence (0-561 families at < 1% and 5-1093 families at < 5% divergence)	Jurka <i>et al.</i> 2011
Tourist-like MITE, miniature Ping (mPing), present in 14 copies in <i>Oryza indica</i> and 70 copies in <i>O. japonica</i> .	Jiang <i>et al.</i> 2003
<i>Ty3/gypsy</i> -like retrotransposon expansion in hybrid species <i>H. anomalus</i> , <i>H. deserticola</i> , and <i>H. paradoxus</i> occurred near the time of their origin 0.5 to 1 Mya.	Ungerer <i>et al.</i> 2009

Adapted from: Rebollo *et al.* 2010 and Kim *et al.* 2004.

Transposable element activity in the following examples is concordant with phylogenetic activity of their hosts. Additional examples describe intraspecific diversity in numbers of TE families.

Despite evidence suggesting a role for TEs as drivers of organismal diversity, it does not necessarily follow that TEs are important to the process of speciation. In sexually reproducing species, speciation is the process of converting segregating variation within a species to fixed differences between species through the evolution of reproductive isolation (Dobzhansky 1937; Mayr 1942). Darwin's theory of natural selection, while providing an elegant mechanism for adaptation, did not fully explain how speciation *per se* occurs. The logical difficulty that arises is that selection should weed out variants causing reduced fitness in conspecific matings (Orr

1996). For this reason, even if we assume that TEs are broadly important to biodiversity, it is also worthwhile to address whether they might play a special role in generating isolation.

Thirty years ago Rose and Doolittle (1983) authored a paper in *Science* titled “Molecular Biological Mechanisms of Speciation” in which they examined the empirical evidence for repetitive DNA’s role in the formation of reproductive isolation. It marked an early synthetic effort to relate TEs to the process of speciation. In their review, the mechanisms by which TEs might facilitate the origin of species were subdivided into three general categories: I) Genomic Disease, II) Mechanical Genome Incompatibility, and III) Genome Resetting. Based primarily on what was then recent data demonstrating hybrid dysgenesis between P and M strains of *Drosophila melanogaster*, Rose and Doolittle concluded that Genomic Disease, while seemingly least plausible, had more empirical support than the other two categories. We have learned much in the intervening three decades and in this chapter we will revisit the categories of Rose and Doolittle as a convenient conceptual scaffold for understanding the variety of ways TEs might directly, or indirectly, cause reproductive incompatibility.

MECHANISM I: GENOMIC DISEASE

The view that TEs are genomic parasites lends itself naturally to the hypothesis that antagonistic coevolution between disease (the TEs) and immunity (the host’s genomic defenses) might promote speciation. Much like how B and T cells of the vertebrate adaptive immune system recognize and remember specific pathogens, a host’s need to suppress the deleterious effects of selfishly proliferating TEs may drive specificity of host genomic defenses against their particular complement of TEs (Ironically, TE domestication contributes to the V (D)J recombination system that makes adaptive immunity to pathogens possible; Market and Papavasiliou 2003; Zhou et al. 2004; Kapitonov and Jurka 2005; re- viewed in Litman et al. 2010). Consequently, populations evolving in allopatry that are exposed to different TE pressures may diverge in genomic defense as well.

There are a few ways this genomic disease model could result in reduced fitness for F_1 hybrids. First, since F_1 s are haploid for both parental genomes, to the extent that defense mechanisms are haploinsufficient, TEs from both parental types may escape suppression. Second, if the contribution to defense is inherited primarily from one parent but not the other, TEs from the non-contributing parent might be freed from suppression. Uniparental inheritance of defense is particularly intriguing because it predicts asymmetries in hybrid breakdown. For instance, it is straightforward to imagine that uniparental defense could result in “Darwin’s corollary”, which observes that hybridization barriers are often asymmetric (i.e., the degree of isolation depends on which population is the maternal vs paternal parent; see Turelli and Moyle 2007). If a population that has adapted to invasion by a novel TE hybridizes with a naïve population, hybrids in only one direction will suffer. Additionally, since sex-limited chromosomes (Y or W) tend to accumulate TEs disproportionately relative to the rest of the genome (Charlesworth et al. 1994; Abe et al. 2000; Bachtrog 2005; Charlesworth et al. 2005) the heterogametic sex may be more likely to suffer in the F_1 . For example, if XX mothers contribute to defense, XY hybrid sons might be affected disproportionately because they will not be guarded against TEs originating from their father’s Y chromosome. The expected pattern

of heterogametic F_1 hybrids suffering disproportionately is consistent with a “rule of speciation”, Haldane’s rule (Haldane 1922).

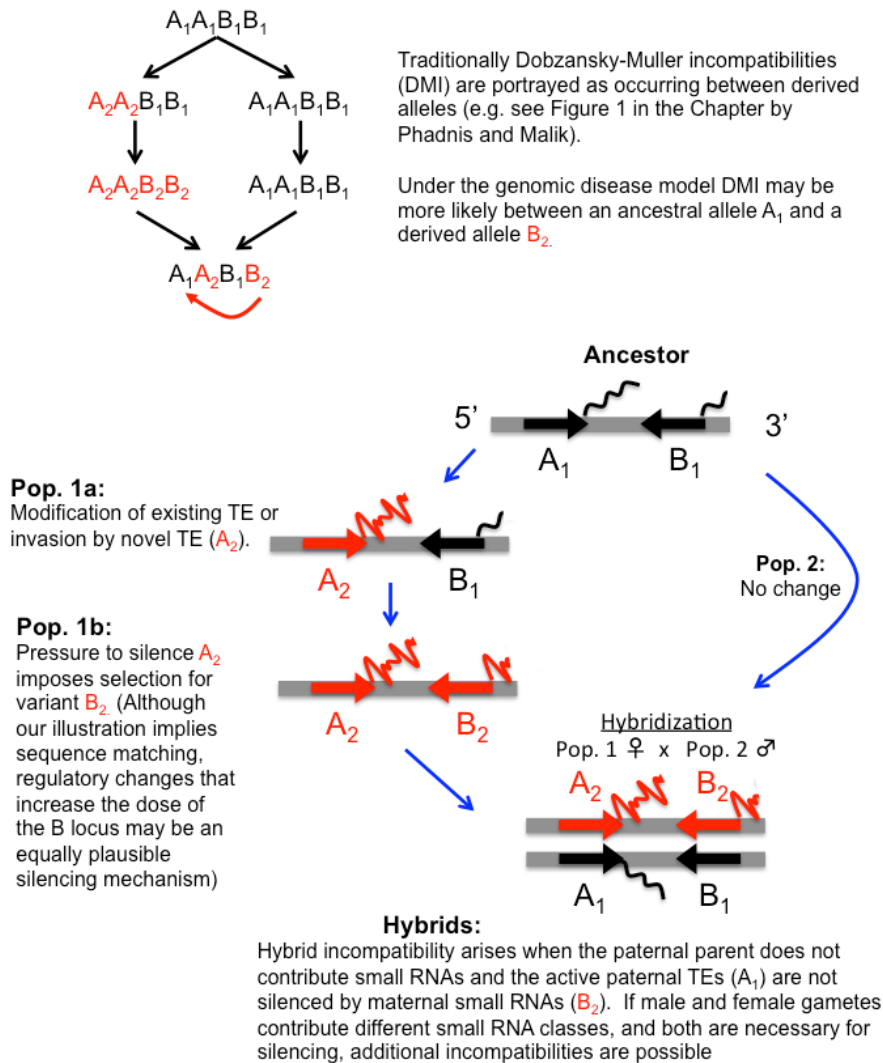


Figure 2. Sense and anti-sense orientation of matching TEs are indicated by black and red arrows. Sense (A locus) and antisense (B locus) transcription are indicated by squiggle lines. Here the B locus produces antisense transcripts which initiate small RNA production and target silencing of the A locus. For simplicity only the haploid genome is illustrated except in hybrids.

If we further abstract the genomic disease model to a two locus genetic interaction model wherein A_1 and A_2 represent TE variants, and B_1 and B_2 represent matching variants at a host defense locus, the history of work surrounding the Dobzhansky-Muller (DM) model of post-zygotic isolation become immediately relevant (Figure 1). While a complete reconciliation of the genomic disease model with DM is outside our scope, it is a potentially interesting avenue for future work. The mathematical framework surrounding DM is well developed (e.g., Turelli and Orr 2000; Orr and Turelli 2001; Demuth and Wade 2005) as is the population genetic

theory for TE proliferation (Charlesworth and Charlesworth 1983; Charlesworth and Langley 1989; Ribeiro and Kidwell 1994; Brookfield and Badge 1997). Reconciling the two may facilitate more specific predictions about the strength of asymmetry and consequences of lineage specific divergence.

TE Suppression by Small RNAs: A Molecular Mechanism for Speciation

While it is not the only mechanism by which TE activity can be controlled, the small RNA “immune response” to TEs represents an ancient, pan-eukaryotic, genomic defense against TE mobilization, and provides the kind of host-pathogen specificity necessary for antagonistic coevolution (reviewed in Aravin et al. 2007; Girard and Hannon 2008; Malone and Hannon 2009; Michalak 2009; Bourc’his and Voinnet 2010; Castillo and Moyle 2012). Small RNA-mediated TE control happens in three basic phases: detection, amplification, and repression (Girard and Hannon 2008). First, the detection phase relies on TE’s propensity to produce anti-sense, double-stranded, or aberrant RNAs, which are recognized by core RNAi machinery (e.g., Dicer RNase III family proteins) and are cleaved into small RNAs such as endogenous siRNAs. During the amplification phase, the primary pool of siRNAs are copied by an RNA dependent RNA polymerase and associate with Argonaute proteins (Ghildiyal et al. 2008). These siRNA - Argonaute complexes then recognize complimentary RNA sequences, guiding cleavage of additional transcripts. Depending on the taxon and tissue, TE repression is ultimately achieved by a combination of transcript degradation (post-transcriptional silencing) and/or siRNA target directed methylation and/or histone modification (transcriptional silencing).

In animals, there is an additional detection mechanism that provides protection to the germline and takes advantage of TE’s unique ability to move within the genome, the piwiRNA pathway. Rather than relying on Dicer dependent dsRNA recognition as above, detection begins with antisense transcription of TEs that have transposed into special RNA gene clusters (e.g., the flamenco locus in *Drosophila* contains sequences that repress Gypsy, Idefix and ZAM family TEs; Pelisson et al. 1994; Prud’homme et al. 1995; Sarot et al. 2004). Processed transcripts from these clusters, called piRNAs, associate with members of the Piwi subfamily of Argonaute proteins (e.g., *Drosophila*: Piwi, Aubergine, AGO3; Mouse: Miwi, Mili, Miwi2). Amplification occurs by a “ping pong” cycle wherein the antisense piRNAs direct cleavage of sense strand TE RNAs which also associate with Piwi proteins that then target cleavage of additional antisense piRNAs...and so on. Post-transcriptional and transcriptional repression is ultimately achieved similarly to that of the siRNA pathway above.

While we have provided only the coarsest overview, it should be clear that the genomic disease - immune response analogy is an apt one despite predating discovery of small RNA mediated TE control by more than a decade (Bingham et al. 1982; Rose and Doolittle 1983; Ginzburg et al. 1984). Breakdown at any of the three stages of small RNA mediated TE control could result in hybrid problems. At detection, uniparental inheritance of small RNA is predicted to be particularly problematic, and haploinsufficiency could follow from problems in the amplification phase. Inadequacy of either of these phases would be expected to result in breakdown of repression.

EVIDENCE FOR MECHANISM I

Drosophila Hybrid Dysgenesis

Early evidence (in fact motivation) for the genomic disease model came from two independent systems in *Drosophila* (P-M and I-R) where crossing different strains of *D. melanogaster* resulted in hybrids with a suite of maladaptive traits such as sterility, gonad hypertrophy, extensive chromosomal aberrations, male recombination, and elevated germline mutation (Picard and L'Héritier 1971; Kidwell and Kidwell 1975; Picard 1976; Engels and Preston 1979; Schaefer et al. 1979; Hiraizumi et al. 1973; Yamaguchi, 1976; Kidwell et al. 1977; Woodruff and Thompson, 1977; Thompson et al. 1978) These studies also revealed that age and rearing temperature impact the degree of hybrid phenotypes (Picard 1976; Bucheton et al. 1976) and collectively the consequences of these crosses was termed “hybrid dysgenesis” (Kidwell et al. 1977).

The underlying cause of hybrid dysgenesis was eventually traced to TE activity in offspring of P x M crosses and I x R crosses (Pelisson 1981; Rubin et al. 1982; Bingham et al. 1982; Kidwell 1983; Bucheton et al. 1984). In both cases strains more recently collected from the wild had active TEs that were not present in long-time lab strains. Hybrid dysgenesis only occurred in these crosses when the paternal parent carried the autonomous TE and the maternal parent did not. In the I-R system dysgenesis only arises when inducer (I) strain males are crossed to reactive (R) strain females. In the P-M system dysgenesis only occurs when paternal (P) strain males are crossed with maternal (M) strain females (Kidwell et al. 1977; Bingham et al. 1982; Kidwell 1983). The asymmetry in consequences of these crosses, particularly I-R sterile hybrid females, suggested that a maternal factor is responsible for offspring's ability to control inherited TEs (Bregliano et al. 1980).

Discovery that members of the Piwi subfamily of Argonaute proteins are maternally loaded into the pole plasm that will give rise to the future germline, and are necessary for TE silencing (Reiss et al. 2004; Sarot et al. 2004), added to early evidence for maternal inheritance of TE repression (Jensen et al. 1999). Later, small RNAs were also shown to be inherited maternally (Blumenstiel and Hartl 2005) and the protein-RNA complexes are responsible for TE silencing through females. Since the cytoplasm is mostly discarded from sperm there is little opportunity for male transmission of small RNAs. Thus, the hybrid dysgenesis observed in P-M and I-R hybrids appears to result from uniparental (maternal) inheritance of piRNA based TE control and the consequent inability to silence paternally derived TEs (Brennecke et al. 2008). Furthermore, maternal piRNA mediated suppression of TEs in the germline appears to be widely conserved among animals (reviewed in O'Donnell and Boeke 2007).

TE Derepression in Interspecific Animal Hybrids

Even before the mechanisms of small RNA based TE silencing were well worked out, researchers sought to test predictions of TE-mediated speciation in a comparative framework by looking for similar phenotypes in additional *Drosophila* hybrids (Coyne 1986; Hey 1988; Coyne 1989). These experiments assayed hybrids from isofemale lines derived from natural populations, as well as hybrids between natural and laboratory mutant lines. Enthusiasm over

the role of TEs in speciation subsided, when it appeared that hybrid dysgenesis was not a ubiquitous feature of interspecific hybrids from the *D. melanogaster* or *D. affinis* subgroups (Coyne 1985; 1986; 1989; Eanes et al. 1988).

Additionally, even in *Drosophila* systems where hybrid dysgenesis occurs, sterility often varies with temperature and age. This provides an opportunity for previously naïve populations to adapt to new elements; perhaps mirroring the invasion of novel TEs within populations. For example, by repeatedly crossing males from an I strain to older dysgenic females that had regained some fertility, Péliçon and Bregliano (1987) were able to introduce repression of I element proliferation in a previously reactive genome within 15 generations. Based primarily on the *Drosophila* findings, Coyne (1986) suggested that to demonstrate a convincing case for TEs causing speciation requires:

- (1) a difference in the distribution of element families among species,
- (2) that these differences cause reproductive isolation (temperature-sensitive gonadal dysgenesis or elevated mutation and recombination rates do not themselves result in isolation)
- (3) the existence of reciprocal sterility effects of the elements in cases where sterility occurs in reciprocal crosses between species.

Despite early pessimism, new findings from *D. melanogaster* x *D. simulans* hybrids have helped fuel renewed interest in TEs role in speciation. For example, Kelleher et al. (2012) recently showed that in contrast to intraspecific *D. melanogaster* crosses where maternal transmission of piRNAs fails to silence active paternal TEs, the interspecific hybrids have widespread activation of both maternally and paternally derived TEs. The study also showed that hybrid offspring are phenotypically most similar to flies with mutations in piRNA pathway genes (e.g., *Piwi*, *Aubergine* and *Argonaut3*). Ten proteins in the piRNA pathway showed excess amino acid changes between *D. melanogaster* and *D. simulans* suggesting a model wherein divergence in piRNA effector proteins between species is responsible for TE derepression in hybrids (Kelleher et al. 2012).

Several other studies also now provide evidence for both lineage-specific evolution of TE families (Table 1) and elevated activity of TEs in interspecific animal hybrids. For example, hybrids between *Drosophila koepferae* and *D. buzzati* experience increased *Osvaldo* transposition, a TE present in a repressed state in both parent genomes (Labrador et al. 1999). In mammals, hybrids between the kangaroo species *Macropus rufogriseus* and *M. agilis* exhibit unstable centromeres due to the expansion of KERV-1 TEs (Metcalf et al. 2007). Hybrids between the Wallaby species *M. eugenii* and *Wallabia bicolor* also experience elevated TE activity, leading to the expansion of centromeric heterochromatin (O'Neill et al. 1998). In lake whitefish (*Coregonus spp.*), evidence for an increase in TE activity was revealed by sequencing the transcriptomes of hybrids between normal and dwarf species (Renaut et al. 2010).

TE Derepression in Plant Hybrids

There are also several examples of TE proliferation in hybrid plants that are consistent with the genomic disease model. Indeed, McClintock's original observations of mosaic maize kernels that lead to the discovery of TEs and motivated her genomic shock hypothesis (McClintock

1950; 1984), are broadly compatible with the genomic disease model. Furthermore, the first mechanistic example of genomic defense induced by antisense TE transcription was also found in maize; silencing of the *MuDR* element involves transcription of *Mu killer*, which is the inverted duplicate of a partially deleted *MuDR* and induces heritable silencing through the small RNA pathway (Slotkin et al. 2003; 2005).

The best-studied and most suggestive link between TE derepression and speciation involves crosses between *Arabidopsis thaliana* and *A. arenosa*, which show postzygotic isolation mediated at least in part by proliferation of the pericentromeric *ATHILA* retrotransposon. Typically, *A. thaliana* ovules fertilized by *A. arenosa* pollen result in 95-100% seed abortion, and the reciprocal cross is impossible because *A. thaliana* pollen do not germinate on the *A. arenosa* stigma (Comai et al. 2000). Josefsson et al. (2006) demonstrated that seed abortion is significantly reduced when the *A. thaliana* female parent has higher ploidy than the *A. arenosa* pollen parent. In this cross, hybrid viability is correlated with the expression of only paternally derived *ATHILA* elements, which are more abundant in the *A. arenosa* (the paternal parent) genome than in *A. thaliana* (Josefsson et al. 2006). To explain this Josefsson et al. (2006) proposed the dosage dependent induction (DDI) model suggesting that increased relative maternal ploidy (i.e., dose) increases protection of the embryo and endosperm by balancing maternal suppressive factors with the paternal *ATHILA* copy number.

Since not all TE families are mobilized in the *A. thaliana* x *A. arenosa* cross, it remains unclear to what extent quantity and or specificity of repressors are important to hybrid inviability (Josefsson et al. 2006; Michalak 2009; Calarco and Martienssen 2011; Castillo and Moyle 2012). Furthermore, Martienssen and colleagues also suggest that both paternal and maternal factors may be involved (Slotkin et al. 2009; Martienssen 2010; Calarco and Martienssen 2011). Unlike animals, which only contribute small RNAs maternally, in *Arabidopsis* (and other angiosperms) both sexes contribute small RNAs that impact TE silencing. However, the two sexes produce distinct classes of small RNAs. Females produce 24-bp RNAs that interact with ARGONAUT9 in a manner reminiscent of the Piwi family - piRNA pathway in animals (Olmedo-Monfil et al. 2010). In males, pollen consists of three haploid cells - two identical sperm cells within an encompassing vegetative cell (McCormick 1993). TEs are derepressed via a regulated loss of DNA methylation in the vegetative nucleus and give rise to a distinct class of 21-bp small RNAs that are then passed to the sperm cells where they contribute to silencing their cognate elements (Slotkin et al. 2009). If it is necessary to have appropriate constituents from both the 21 and 24-bp small RNA pathways to silence TEs in hybrid offspring, the inviability of *A. thaliana* x *A. arenosa* hybrids may result from mismatches in suppression from either or both parental genomes.

Other examples that circumstantially tie TE derepression in plant hybrids to speciation include sunflower and rice. In sunflowers, three diploid species have arisen by ancient hybridization between *Helianthus annuus* and *H. petiolaris*. The hybrid species *H. anomalus*, *H. deserticola*, and *H. paradoxus* independently experienced large increases in retrotransposon content that are broadly consistent with the time of the species' origins (Welch and Rieseberg 2002; Schwarzbach and Rieseberg 2002; Gross et al. 2003; Ungerer et al. 2006; 2009). Interestingly, in five current natural hybrid zones between *H. annuus* and *H. petiolaris*, TE copy number in hybrid plants does not exceed the parental species values despite active transcription of the same TE families that are expanded in the three species derived from ancient hybridization (Kawakami et al. 2011). This finding suggests that post-transcriptional regulation currently limits TE proliferation in *Helianthus* hybrids. The fact that the three hybrid species

are each adapted to extreme environmental conditions is consistent with the ‘genomic shock’ and ‘epi-transposon’ ideas where TE proliferation in hybrids is promoted by a combination of biotic and abiotic stressors (in this case hybridization and harsh environmental conditions) upsetting the epigenetic defenses that normally keep TE activity suppressed (McClintock 1984; Wessler 1996; Lisch 2009; Zeh et al. 2009).

In rice, the genomic distribution and abundance of the miniature inverted-repeat transposable element (MITE) *mPing* differs within and between *indica* and *japonica* subspecies of *Oryza sativa* (Jiang et al. 2003), reflecting multiple rounds of differential amplification (Lu et al. 2012). Additionally, it has been shown in laboratory crosses that a hybridization signal provided by pollination with wild rice, *Zizania latifolia* remobilizes *mPing*, its supposed autonomous partner *Pong*, and at least two other TE families (*Tos17* and *Dart*; Shan et al. 2005; Wang et al. 2010). This suggests that bursts of TE proliferation in rice may be a consequence of hybridization; however, it remains far from illustrating a direct role in speciation.

MECHANISM II: MECHANICAL INCOMPATIBILITY

Rose and Doolittle’s (1983) focus on Mechanical Incompatibility dealt with two potential roles that TEs might play in hybrid infertility and inviability. First, rapid sequence turnover, particularly in heterochromatic repeats, could result in meiotic nondisjunction due to failure of homologous chromosomes to recognize each other during meiosis. Second, differential proliferation among chromosomes could disrupt necessary spatial relationships among nonhomologous chromosomes that are based on relative chromosome arm length (see Rose and Doolittle 1983). They concluded that there was little evidence for these types of mechanical isolation despite ample evidence for sequence turnover and proliferation. In fact they noted how little sequence homology was actually necessary for chromosomes to segregate properly. Somewhat ironically, the first “speciation gene” identified in animals, *Odysseus*, is now known to be a coevolutionary result of genomic conflict driven by the need to bind with rapidly changing heterochromatic repeats (see chapter by Phadnis and Malik in this volume); and this interplay between heterochromatin associated factors is emerging as a common cause of *Drosophila* incompatibilities (e.g., Brideau et al. 2006, Bayes and Malik 2009, Ferree and Barbash 2009, Cattani and Presgraves 2012).

Another way that TE activity might result in mechanical incompatibility is by generating structural variation (e.g., inversions, translocations, duplications, or deletions). Isolation due to chromosomal rearrangements could arise as a direct consequence of underdominance (i.e., lower heterozygote fitness than either homozygote) in heterokaryotypic hybrids. For instance, crossovers within inversion heterozygotes may result in gametes that do not contain a complete gene complement. If such crossovers were common, hybrids between populations or species with different inversions would be expected to show underdominance for fertility. A major difficulty with the potential for rearrangements to cause isolation directly is that fixation of an underdominant rearrangement is unlikely except in situations where genetic drift is strong (Walsh 1982; Lande 1985). If, on the other hand, there is no underdominance so that rearrangements are more likely to fix, then they are expected to have little effect on isolation.

Although direct responsibility for isolation seems unlikely, structural variants (particularly inversions) may indirectly facilitate isolation by reducing or eliminating regional

recombination near breakpoints in heterokaryotypic individuals (Rieseberg 2001; Noor et al. 2001; Hoffmann and Rieseberg 2008; Brown and O'Neill 2010; McGaugh and Noor 2012). Recombination plays a critical role in the shaping of integrated systems within species, forming adaptive peaks in the potential field of gene combinations and holding the species as a cohesive unit (Wright 1931; 1932; Gavrillets 2004). Reduced recombination facilitates the maintenance of linkage among genes involved in adaptive divergence and reproductive isolation (Rieseberg 2001; Noor et al. 2001; Navarro and Barton 2003). Under this scenario, mechanical isolation is only indirectly responsible for what is more accurately a genic model of speciation. The so-called “genomic islands of divergence” that arise in regions of low recombination only contribute to speciation when they contain factors contributing to reproductive isolation (Feder and Nosil 2009). For TEs to be implicated in speciation via this mechanism requires: 1) identifying TEs as the cause of inversion, and 2) showing that loci within the inversion cause isolation.

EVIDENCE FOR MECHANISM II

Structural Variation Caused by TE Activity

Besides normal transposition events, which depending on the type of TE, may or may not involve duplication, TE mediated structural variants arise when: 1) homologous TEs at different genome locations recombine (ectopic recombination), 2) when TE excision results in ectopic sequences being incorporated during double strand break repair (non-homologous end joining), or 3) when the ends from two TEs synapse and engage in complete or partial simultaneous transposition (alternative transposition; reviewed in Gray 2000; Feschotte and Pritham 2007b). There is abundant evidence that each of these are common causes of structural variation (Collins and Rubin 1984; Engels and Preston 1984; Lister et al. 1993; Lim and Simmons 1994; Walker et al. 1995; Hua-Van et al. 2002; Zhang et al. 2009; Xing et al. 2009; Quinlan et al. 2010; Guillén and Ruiz 2012).

In some cases TEs may occur near breakpoints as a consequence of inversion rather than a cause because they tend to accumulate in regions of low recombination. However, TEs are clearly causative in other cases; perhaps the best examples coming from *Drosophila buzzatii*, where Ruiz and colleagues have mapped members of the class II TE families, *FoldBack* (*FB*) and *hAT*, to breakpoints of several inversions that occur in natural populations (Caceres et al. 1999; Casals et al. 2003; Delprat et al. 2009; Guillén and Ruiz 2012).

Structural Variations Associated with Reproductive Isolation

Studies in *Drosophila* spp. (Noor et al. 2001; Machado et al. 2007a), *Sorex* shrews (Basset et al. 2008), *Anopheles* mosquitos (Michel 2006), *Rhagoletis* flies (Feder et al. 2003), and *Mimulus* monkeyflower ecotypes (Lowry and Willis 2010) all demonstrate that rearranged regions diverge more quickly than collinear ones, or maintain greater divergence in the presence of gene flow between closely related races or species. However, the association between islands of divergence and speciation is not necessarily straightforward, as patterns of

nucleotide differentiation are not sufficient evidence in and of themselves to infer a causal link with speciation (Noor and Bennett 2009; Turner and Hahn 2010). Additionally, regional differentiation may be maintained by selection even without chromosomal rearrangements (e.g., Turner et al. 2005; Harr 2006; Feder and Nosil 2009; Nadeau et al. 2012).

There are at least two examples where divergence in inversions has been tied to reproductive isolation. First, Noor and colleagues have mapped isolating barriers to inversion differences between *D. pseudoobscura* and *D. persimilis* (Ortiz-Barrientos et al. 2004; Ortiz-Barrientos and Noor 2005). These two species diverged within the last 500,000 years and are fixed differently for two inversions, but have experienced extensive introgression in other parts of the genome. Differences in cuticular hydrocarbons, mating preference, hybrid male sterility and inviability, and hybrid courtship dysfunction all map wholly or in part to the two fixed inversions (reviewed in Noor et al. 2001; Machado et al. 2007b; McGaugh and Noor 2012).

Second, Lowry and Willis (2010) discovered a geographically widespread inversion polymorphism responsible for local adaptation and prezygotic isolation in *Mimulus guttatus*. In a reciprocal transplant experiment wherein they isolated the effect of the inversion in nature by reciprocally introgressing into the genetic background of alternate ecotypes (perennial vs. annual) they showed that the inversion affects morphology, flowering time, survivorship, and prezygotic isolation.

A TE Induced Structural Variant That Causes Isolation

Clearly there is independent evidence for TE's role in both generating structural variation, and structural variation being tied to isolation. There is also at least one case where a TE induced variant is partially responsible for originating postzygotic isolation; the maternal effect dominant embryonic arrest (Medea) system in the red flour beetle *Tribolium castaneum*. Four Medea factors have been isolated from nature, but only 2 are stable in lab culture (M1 and M4; Beeman et al. 1992; Beeman and Friesen 1999). M1 is found in most regions of the world and has a selfish advantage when invading naïve populations because offspring of heterozygous mothers who do not receive a copy of M1 all die early in development (Wade and Beeman 1994). M4 has an even broader geographic range, and produces a very similar phenotype, but M1 and M4 do not cross rescue. The causative locus for M1 has been mapped to a 21.5 kb composite Tc-1 like transposon insertion containing several defective *Tribolium* gene duplicates (Lorenzen et al. 2008). The insertion occurs in the ~700bp intergenic region between the 3' UTR of two functional genes and while the mechanistic cause of the maternal phenotype remains unknown preliminary data show that the insertion may disrupt cis-natural antisense transcription that occurs in wildtype beetles (Demuth et al. unpublished data).

What makes Medea relevant to the present discussion of speciation is that offspring from crosses between M1 (or M4) and a second, hybrid incompatibility factor (H-factor), do not fully develop. H-factor is found widely distributed in India, and represents a strong postzygotic isolating barrier despite Medea and H-factor each producing viable, fertile offspring in combination with wildtype populations (Thomson et al. 1995; Thomson and Beeman 1999). Recently, H-factor was mapped to variation within the introns of an ecdysone receptor homolog, but the functional mechanisms underlying interaction between Medea and H-factor remain a mystery (Drury et al. 2011).

MECHANISM III: GENOMIC RESETTING

The last potential mechanism by which Rose and Doolittle suggested TEs might play a role in speciation involves their capacity to contribute regulatory sequences. Autonomous TEs encode the proteins and regulatory information necessary to catalyze their own transcription. There is now abundant evidence that proteins and regulatory sequences derived from TEs, have been exapted to perform functions for their host (Britten 1997; Feschotte 2008). Transposases in particular have been a recurrent source of “domesticated” genes. Because they are DNA binding proteins, typically with self-specificity, as a TE disperses throughout a genome it has the potential to “rewire” regulatory circuits if the transposase becomes domesticated (Cordeaux et al. 2006; Feschotte 2008). In addition to abundant evidence for isolated recruitment of functional coding and non-coding sequences originating from TEs, there is growing evidence for their role in large scale rewiring of transcriptional modules (Bourque et al. 2008; Kunarso et al. 2010; Lynch et al. 2011; Schmidt et al. 2012), but no evidence that we are aware of for this mechanism in speciation.

Despite a lack of evidence for this mechanism in its original formulation, the mechanistic details of genome defense against TE activity discussed above have several facets that could easily fall under the heading “genomic resetting” though it begins to blur the line with the genomic disease model. For instance, growing evidence shows that in the germline of plants and animals, safeguarding gametes and embryonic cells involves wholesale changes in the heterochromatin of “companion” cells such that TEs are derepressed in order to provide templates for small RNA that will go on to re-establish TE silencing in cells that form the offspring (Chambeyron et al., 2008; Slotkin et al. 2009; Calarco and Martienssen, 2011). This is genome resetting, or reprogramming, in a literal sense. Indeed the newfound interplay between small RNAs, TEs, methylation and histone modification has prompted a call for a chromatin model of speciation (Michalak 2009).

CONCLUSION

As we stated in introducing this chapter, we have learned much in the 30 years since Rose and Doolittle’s “Molecular Biological Mechanisms of Speciation”. However, in many respects TEs belong to the “dark matter” of genomes and their role in speciation remain obscure. Because of their repetitive nature they are typically viewed as a nuisance to genome sequencing projects, remaining in bins of unassembled “junk” reads. Surveying them individually poses its own challenges, so getting good estimates of variety and copy number among taxa requires significant effort and purposeful investigation. Given that the decades old search for “speciation genes” has provided few candidates and even fewer where early causation can be inferred, it would be premature to draw strong positive or negative conclusions about TEs role as causative agents in speciation. Additionally, since our ability to assay epigenetic mechanisms of genome defense (e.g., small RNA, methylation, histone modification) on a genome-wide scale in non-model organisms are still nascent technological advances, perhaps it is not surprising that we lack many strong examples. As our mechanistic understanding of TEs and ability to survey them improves, we fully expect to find additional evidence for the role of TEs in speciation.

ACKNOWLEDGMENTS

We would like to thank E. Betrán, and C. Feschotte for helpful comments and discussions about the manuscript. This work was supported by funding from National Institutes of Health grant #2R01GM065414-05A1.

REFERENCES

- Abe, H., F. Ohbayashi, T. Shimada, T. Sugasaki, S. Kawai, K. Mita and T. Oshiki. 2000. Molecular structure of a novel gypsy -Ty3-like retrotransposon (Kabuki) and nested retrotransposable elements on the W chromosome of the silkworm *Bombyx mori*. *Mol Gen Genet* 263:916–924.
- Aravin, A. A., Hannon G. J. and Brennecke J. 2007. The Piwi-piRNA Pathway provides an adaptive defense in the transposon arms race. *Science* 318: 761–764.
- Bachtrog, D. 2005. Sex chromosome evolution: Molecular aspects of Y-chromosome degeneration in *Drosophila*. *Genome Research* 15:1393–1401.
- Basset, P., Yannic G. and Hausser J. 2008. Chromosomal rearrangements and genetic structure at different evolutionary levels of the *Sorex araneus* group. *J Evol Biol* 21:842–852.
- Bayes, J. J., and Malik H. S., 2009. Altered heterochromatin binding by a hybrid sterility protein in *Drosophila* Sibling Species. *Science* 326:1538–1541.
- Beeman, R. W. and Friesen K. S. 1999. Properties and natural occurrence of maternal-effect selfish genes (“Medea” factors) in the red flour beetle, *Tribolium castaneum*. *Heredity* 82:529–534.
- Beeman, R. W., Friesen K. S. and Denell R. E. 1992. Maternal-effect selfish genes in flour beetles. *Science* 256:89–92.
- Belancio, V. P., Hedges D. J. and Deininger P. 2008. Mammalian non-LTR retrotransposons: For better or worse, in sickness and in health. *Genome Research* 18:343–358.
- Biemont, C. and Vieira C. 2005. What transposable elements tell us about genome organization and evolution: the case of *Drosophila*. *Cytogenet Genome Res* 110:25–34.
- Biémont, C. and Vieira C. 2006. Genetics: Junk DNA as an evolutionary force. *Nature* 443:521–524.
- Bingham, P. M., Kidwell M. G. and Rubin G. M. 1982. The molecular basis of *P-M* hybrid dysgenesis: the role of the *P* element, a *P*-strain-specific transposon family. *Cell* 29:995–1004.
- Blumenstiel, J. P. and Hartl D. L. 2005. Evidence for maternally transmitted small interfering RNA in the repression of transposition in *Drosophila virilis*. *Proc Natl Acad Sci USA* 102:15965–15970.
- Bourc’his, D. and Voinnet O. 2010. A small-RNA perspective on gametogenesis, fertilization, and early zygotic development. *Science* 330:617–622.
- Bourque, G., Leong B., Vega V. B., Chen X., Lee Y. L., Srinivasan K. G., Chew J. L., Ruan Y., Wei C.-L., Ng H. H., and Liu E. T. 2008. Evolution of the mammalian transcription factor binding repertoire via transposable elements. *Genome Research* 18:1752–1762.
- Brandt, J., Schrauth S., Veith A. M., Froschauer A., Haneke T., Schultheis C., Gessler M., Leimeister C. and Volff J. N. 2005. Transposable elements as a source of genetic

- innovation: expression and evolution of a family of retrotransposon-derived neogenes in mammals. *Gene* 345:101–111.
- Bregliano, J., Picard G., Bucheton A., Pelisson A., Lavige J. and L'Heritier P. 1980. Hybrid dysgenesis in *Drosophila melanogaster*. *Science* 207:606–611.
- Brennecke, J., Malone C. D., Aravin A. A., Sachidanandam R., Stark A. and Hannon G. J. 2008. An epigenetic role for maternally inherited piRNAs in transposon silencing. *Science* 322:1387–1392.
- Brideau, N. J., Flores H. A., Wang J., Maheshwari S., Wang X. and Barbash D. A. 2006. Two Dobzhansky-Muller genes interact to cause hybrid lethality in *Drosophila*. *Science* 314:1292–1295.
- Britten, R. J. 1997. Mobile elements inserted in the distant past have taken on important functions. *Gene* 205:177–182.
- Britten, R. J. and Davidson E. H. 1971. Repetitive and non-repetitive DNA sequences and a speculation on the origins of evolutionary novelty. *Q Rev Biol* 46:111–138.
- Brookfield, J. F. Y. and Badge R. M. 1997. Population genetics models of transposable elements. *Genetica* 100:281–294.
- Brown, J. D. and O'Neill R. J. 2010. Chromosomes, conflict, and epigenetics: chromosomal speciation revisited. *Annu. Rev. Genom. Human Genet.* 11: 291–316.
- Bucheton, A., Lavige J. M., Picard G. and L'Heritier P. 1976. Non-Mendelian female sterility in *Drosophila melanogaster*: Quantitative variations in the efficiency of inducer and reactive strains. *Heredity* 36:305–314.
- Bucheton, A., Paro R., Sang H. M., Pelisson A. and Finnegan D. J. 1984. The molecular basis of I-R hybrid Dysgenesis in *Drosophila melanogaster*: Identification, cloning, and properties of the I factor. *Cell* 38:153–163.
- Butter, F., Kappei D., Buchholz F., Vermeulen M. and Mann M. 2010. A domesticated transposon mediates the effects of a single-nucleotide polymorphism responsible for enhanced muscle growth. *Nat. Rev. Genet.* 11:305–311.
- Cáceres, M., Ranz, J. M., Barbadilla, A., Long, M., & Ruiz, A. 1999. Generation of a widespread *Drosophila* inversion by a transposable element. *Science*, 285:415–418.
- Calarco, J. P. and Martienssen R. A. 2011. Genome reprogramming and small interfering RNA in the *Arabidopsis* germline. *Curr. Opin. Genet. Dev.* 21: 134–139.
- Callinan, P. A. and Batzer M. A. 2006. Retrotransposable elements and human disease. *Genome dynamics* 1:104–115.
- Casals, F., M. Cáceres and A. Ruiz. 2003. The foldback-like transposon Galileo is involved in the generation of two different natural chromosomal inversions of *Drosophila buzzatii*. *Mol Biol Evol* 20:674–685.
- Castillo, D. M. and Moyle L. C. 2012. Evolutionary Implications of Mechanistic Models of TE-Mediated Hybrid Incompatibility. *Evolution* 2012:1–12.
- Cattani, M. V., and Presgraves D. C. 2012. Incompatibility between X chromosome factor and pericentric heterochromatic region causes lethality in hybrids between *Drosophila melanogaster* and its sibling species. *Genetics* 191:549–559.
- Chambeyron, S., Popkova A., Payen-Groschene G., Brun C., Laouini D., Pelisson A., and Bucheton A. 2008. piRNA-mediated nuclear accumulation of retrotransposon transcripts in the *Drosophila* female germline. *Proc Natl Acad Sci USA* 105:14964–14969.
- Charlesworth, B. and Langley C. H. 1989. The population genetics of *Drosophila* transposable elements. *Annu Rev Genet* 23:251–287.

- Charlesworth, B. and Charlesworth D. 1983. The population dynamics of transposable elements. *Genet Res* 42:1–27.
- Charlesworth, B., Sniegowski P. and Stephan W. 1994. The evolutionary dynamics of repetitive DNA in eukaryotes. *Nat Rev Genet* 371:215–220.
- Charlesworth, D., Charlesworth B. and Marais G. 2005. Steps in the evolution of heteromorphic sex chromosomes. *Nat Rev Genet* 95:118–128.
- Chénais, B., Caruso A., Hiard S. and Casse N. 2012. The impact of transposable elements on eukaryotic genomes: From genome size increase to genetic adaptation to stressful environments. *Cell* 509:7–15.
- Collins, M. and Rubin G. M. 1984. Structure of chromosomal rearrangements induced by the FB transposable element in *Drosophila*. *Nature* 308:323–327.
- Comai, L., Tyagi A. P., Winter K., Holmes-Davis R., Reynolds S. H., Stevens Y. and Byers B. 2000. Phenotypic instability and rapid gene silencing in newly formed *Arabidopsis* allotetraploids. *Plant Cell* 12:1551–1568.
- International Chicken Genome Sequencing Consortium. 2004. Sequence and comparative analysis of the chicken genome provide unique perspectives on vertebrate evolution. *Nature* 432:695–716.
- International Human Genome Sequencing Consortium. 2001. Initial sequencing and analysis of the human genome. *Nature* 409:860–921.
- Cordaux, R., Udit S., Batzer M. A., and Feschotte C. 2006. Birth of a chimeric primate gene by capture of the transposase gene from a mobile element. *Proc Natl Acad Sci USA* 103:8101–8106.
- Coyne, J. A. 1985. Genetic studies of three sibling species of *Drosophila* with relationship to theories of speciation. *Genet Res* 46:169–192.
- Coyne, J. A. 1986. Meiotic segregation and male recombination in interspecific hybrids of *Drosophila*. *Genetics* 114:485–494.
- Coyne, J. A. 1989. Mutation rates in hybrids between sibling species of *Drosophila*. *Heredity* 63 (Pt 2):155–162.
- Coyne, J. A. and Orr H. A. 2004. *Speciation*. Sinauer Associates Inc.
- de Boer, J. G., Yazawa R., Davidson W. S. and Koop B. F. 2007. Bursts and horizontal evolution of DNA transposons in the speciation of pseudotetraploid salmonids. *BMC Genomics* 8:422.
- de Koning, A. P. J., Gu, W., Castoe, T. A., Batzer, M. A., & Pollock, D. D. 2011. Repetitive Elements May Comprise Over Two-Thirds of the Human Genome. *PLoS Genetics* 7: e1002384.
- Delprat, A., Negre B., Puig M. and Ruiz A. 2009. The transposon galileo generates natural chromosomal inversions in *Drosophila* by ectopic recombination. *PLoS ONE* 4:e7883.
- Demuth, J. P. and Wade M. J. 2005. On the theoretical and empirical framework for studying genetic interactions within and among species. *Am. Nat.* 165:524–536.
- Dobigny, G., Ozouf-Costaz C., Waters P. D., Bonillo C., Coutanceau J.-P. and Volobouev V. 2004. LINE-1 amplification accompanies explosive genome repatterning in rodents. *Chromosome Res* 12:787–793.
- Dobzhansky, T. 1937. *Genetics and the Origin of the Species*. Columbia University Press, New York, NY.

- Drury, D. W., Jideonwo V. N., Ehmke R. C. and Wade M. J. 2011. An unusual barrier to gene flow: perpetually immature larvae from inter-population crosses in the flour beetle, *Tribolium castaneum*. *J Evol Biol* 24:2678–2686.
- Walker, E. L., Robbins T. P., Bureau T. E., Kermicle J. and Dellaporta S. L. 1995. Transposon-mediated chromosomal rearrangements and gene duplications in the formation of the maize *R-r* complex. *EMBO J* 14:2350–2363.
- Eanes, W. F., Wesley C., Hey J., Houle D. and Ajioka J. W. 1988. The Fitness Consequences of *P*-Element Insertion in *Drosophila melanogaster*. *Genet Res* 52:17–26.
- Eickbush, T. H. and Furano A. V. 2002. Fruit flies and humans respond differently to retrotransposons. *Curr. Opin. Genet. Dev.* 12:669–674.
- Engels, W. R. and Preston C. R. 1979. Hybrid dysgenesis in *Drosophila melanogaster*: the biology of female and male sterility. *Genetics* 92:161–174.
- Engels, W. R. and C. R. Preston. 1984. Formation of chromosome rearrangements by *P* factors in *Drosophila*. *Genetics* 107:657–678.
- Feder, J. L. and Nosil P. 2009. Chromosomal inversions and species differences: when are genes affecting adaptive divergence and reproductive isolation expected to reside within inversions? *Evolution* 63: 3061–3075.
- Feder, J. L., Roethele J. B., Filchak K., Niedbalski J. and Romero-Severson J. 2003. Evidence for inversion polymorphism related to sympatric host race formation in the apple maggot fly, *Rhagoletis pomonella*. *Genetics* 163: 939–953.
- Fedoroff, N. V. 1999. Transposable Elements as a Molecular Evolutionary Force. *Annals NY Acad Sci* 870:251–264.
- Fedoroff, N. V. 2012. Transposable Elements, Epigenetics, and Genome Evolution. *Science*. 338:758-767.
- Ferree, P. M. and Barbash D. A. 2009. Species-specific heterochromatin prevents mitotic chromosome segregation to cause hybrid lethality in *Drosophila*. *PLoS Biology* 7:e1000234.
- Feschotte, C. 2008. Transposable elements and the evolution of regulatory networks. *Nat. Rev. Genet.* 9:397–405.
- Feschotte, C. and Pritham E. J. 2007a. Computational Analysis and Paleogenomics of Interspersed Repeats in Eukaryotes. Pp. 31–54 in N. Stojanovic, ed. *Computational Genomics: Current Methods*. Horizon Bioscience, Norfolk, UK.
- Feschotte, C. and Pritham E. J. 2007b. DNA transposons and the evolution of eukaryotic genomes. *Annu Rev Genet* 41:331–368.
- Finnegan, D. J. 1992. Transposable elements. Pp. 1096–1107 in D. L. Lindsley and G. G. Zimm, eds. *The genome of Drosophila melanogaster*. San Diego Academic Press, San Diego, CA.
- Gavrilets, S. 2004. *Fitness landscapes and the origin of species*. Princeton Univ Pr.
- Ghildiyal, M., Seitz H., Horwich M. D., Li C., Du T., Lee S., Xu J., W. Kittler E. L., Zapp M. L., Weng Z. and Zamore P. D. 2008. Endogenous siRNAs derived from Transposons and mRNAs in *Drosophila* Somatic Cells. *Science* 320:1077–1081.
- Ginzburg, L. R., Bingham P. M. and Yoo S. 1984. On the theory of speciation induced by transposable elements. *Genetics* 107:331–341.
- Girard, A. and Hannon G. J. 2008. Conserved themes in small-RNA-mediated transposon control. *Trends Cell Biol.* 18:136–148.

- Goodier, J. L. and Kazazian H. H. Jr. 2008. Retrotransposons Revisited: The Restraint and Rehabilitation of Parasites. *Cell* 135:23–35.
- Gray, Y. H. M. 2000. It takes two transposons to tango: transposable-element-mediated chromosomal rearrangements. *Cell* 16:461–468.
- Gross, B. L., Schwarzbach A. E. and Rieseberg L. H. 2003. Origin (s) of the diploid hybrid species *Helianthus deserticola* (Asteraceae). *Am. J. Bot.* 90: 1708–1719.
- Guillén, Y. and Ruiz A. 2012. Gene alterations at *Drosophila* inversion breakpoints provide *prima facie* evidence for natural selection as an explanation for rapid chromosomal evolution. *BMC Genomics* 13:53.
- Haas, B. J., Wortman J. R., Ronning C. M., Hannick L. I., Smith R. K. J., Maiti R., Chan A. P., Yu C., Farzad M., Wu D., White O. and Town C. D. 2005. Complete reannotation of the *Arabidopsis* genome: methods, tools, protocols and the final release. *BMC Biol* 3:7.
- Haldane, J. 1922. Sex ratio and unisexual sterility in hybrid animals. *J Genet* 12:101–109.
- Harr, B. 2006. Genomic islands of differentiation between house mouse subspecies. *Genome Res* 16:730–737.
- Hey, J. 1988. Speciation via hybrid dysgenesis: negative evidence from the *Drosophila affinis* subgroup. *Genetica* 78:97–103.
- Hiraizumi, Y., Slatko B., Langley C. and Nill A. 1973. Recombination in *Drosophila melanogaster* male. *Genetics* 73:439–444.
- Hoffmann, A. A. and Rieseberg L. H. 2008. Revisiting the Impact of Inversions in Evolution: From Population Genetic Markers to Drivers of Adaptive Shifts and Speciation? *Annu. Rev. Ecol. Evol. Syst.* 39:21–42.
- Hua-Van, A., Langin T. and Daboussi M. J. 2002. Aberrant transposition of a Tc1-mariner element, impala, in the fungus *Fusarium oxysporum*. *Mol Gen and Genom* 267:79–87.
- Jensen, S., Gassama M.-P. and Heidmann T. 1999. Taming of transposable elements by homology-dependent gene silencing. *Nat. Rev. Genet.* 21: 209–212.
- Jiang, N., Bao Z., Zhang X., Hirochika H., Eddy S. R., McCouch S. R. and Wessler S. R. 2003. An active DNA transposon family in rice. *Nat. Rev. Genet.* 421:163–167.
- Josefsson, C., Dilkes B. and Comai L. 2006. Parent-dependent loss of gene silencing during interspecies hybridization. *Curr. Biol.* 16:1322–1328.
- Jurka, J., Bao W. and Kojima K. K. 2011. Families of transposable elements, population structure and the origin of species. *Biol. Direct* 6:44.
- Kapitonov, V. V., and Jurka J. 2003. Molecular paleontology of transposable elements in the *Drosophila melanogaster* genome. *Proc Natl Acad Sci USA* 100: 6569–6574.
- Kapitonov, V. V., and Jurka J. 2005. RAG1 core and V (D) J recombination signal sequences were derived from Transib transposons. *PLoS Biology* 3: e181.
- Kawakami, T., Dhakal P., Katterhenry A. N., Heatherington C. A. and Ungerer M. C. 2011. Transposable Element Proliferation and Genome Expansion Are Rare in Contemporary Sunflower Hybrid Populations Despite Widespread Transcriptional Activity of LTR Retrotransposons. *Genome Biol Evol* 3:156–167.
- Kazazian, H. H. 1998. Mobile elements and disease. *Curr. Opin. Genet. Dev.* 8:343–350.
- Kazazian, H. H. 2004. Mobile Elements: Drivers of Genome Evolution. *Science* 303:1626–1632.
- Kidwell, M. G. 1983. Evolution of hybrid dysgenesis determinants in *Drosophila melanogaster*. *Proc Natl Acad Sci USA* 80:1655–1659.

- Kidwell, M. G. and Lisch D. R. 2001. Perspective: Transposable elements, parasitic DNA, and genome evolution. *Evolution* 55:1–24.
- Kidwell, M. G. and Kidwell J. F. 1975. Cytoplasm–chromosome interactions in *Drosophila melanogaster*. *Nature* 253:755–756.
- Kidwell, M. G., Kidwell J. F. and Sved J. A. 1977. Hybrid dysgenesis in *Drosophila melanogaster*: A syndrome of aberrant traits including mutation, sterility and male recombination. *Genetics* 86:813–833.
- Kim, T.-M., Hong S.-J. and Rhyu M.-G. 2004. Periodic explosive expansion of human retroelements associated with the evolution of the hominoid primate. *J. Korean Med. Sci.* 19:177–185.
- Kelleher, E. S., Edelman N. B. and Barbash D. A. 2012. *Drosophila* interspecific hybrids phenocopy piRNA-pathway mutants. *PLoS Biology* 10(11): e1001428.
- Kunarso, G., Chia N. Y., Jeyakani J., Hwang C. and Lu X. 2010. Transposable elements have rewired the core regulatory network of human embryonic stem cells. *Nature* 42:631–634.
- Labrador, M., Farre M., Utzet F. and Fontdevila A. 1999. Interspecific hybridization increases transposition rates of *Osvaldo*. *Mol Biol Evol* 16: 931–937.
- Lande, R. 1985. The fixation of chromosomal rearrangements in a subdivided population with local extinction and colonization. *Heredity* 54:323–332.
- Li, W. L., Zhang P., Fellers J. P., Friebe B. and Gill B. S. 2004. Sequence composition, organization, and evolution of the core *Triticeae* genome. *Plant J.* 40:500–511.
- Lim, J. K. and Simmons M. J. 1994. *Gross chromosome rearrangements mediated by transposable elements in Drosophila melanogaster*. 16:269–275.
- Lisch, D. 2009. Epigenetic regulation of transposable elements in plants. *Annu Rev Plant Biol* 60:43–66.
- Lister, C., Jackson D. and Martin C. 1993. Transposon-induced inversion in *Antirrhinum* modifies *nivea* gene expression to give a novel flower color pattern under the control of *cycloidearadialis*. *Plant Cell* 5:1541–1553.
- Litman, G. W., Rast J. P., and Fugmann S. D. 2010. The origins of vertebrate adaptive immunity. *Nat Rev Immunology*, 10:543–553.
- Lorenzen, M. D., Gnirke A., Margolis J., Garnes J., Campbell M., Stuart J. J., Aggarwal R., Richards S., Park Y. and Beeman R. W. 2008. The maternal-effect, selfish genetic element *Medea* is associated with a composite Tc1 transposon. *Proc Natl Acad Sci USA* 105:10085–10089.
- Lorenzi, H., Thiagarajan M., Haas B., Wortman J., Hall N. and Caler E. 2008. Genome wide survey, discovery and evolution of repetitive elements in three *Entamoeba* species. *BMC Genomics* 9:595.
- Lowry, D. B. and Willis J. H. 2010. A Widespread Chromosomal Inversion Polymorphism Contributes to a Major Life-History Transition, Local Adaptation, and Reproductive Isolation. *PLoS Biol* 8:e1000500.
- Lu, C., Chen J., Zhang Y., Hu Q., Su W. and Kuang H. 2012. Miniature Inverted-Repeat Transposable Elements (MITEs) Have Been Accumulated through Amplification Bursts and Play Important Roles in Gene Expression and Species Diversity in *Oryza sativa*. *Mol Biol Evol* 29: 1005–1017.
- Lynch, V. J., Leclerc R. D., May G. and Wagner G. P. 2011. Transposon-mediated rewiring of gene regulatory networks contributed to the evolution of pregnancy in mammals. *Nature Reviews Genetics* 43:1154–1159.

- Machado, C. A., Matzkin L. M., Reed L. K. and Markow T. A. 2007a. Multilocus nuclear sequences reveal intra- and interspecific relationships among chromosomally polymorphic species of cactophilic *Drosophila*. *Mol Ecol* 16:3009–3024.
- Machado, C. A., Haselkorn T. S. and Noor M. A. F. 2007b. Evaluation of the genomic extent of effects of fixed inversion differences on intraspecific variation and interspecific gene flow in *Drosophila pseudoobscura* and *D. persimilis*. *Genetics* 175:1289–1306.
- Maheshwari, S. and Barbash D. A. 2011. The Genetics of Hybrid Incompatibilities. *Annu Rev Genet* 45:331–355.
- Maksakova, I. A., Romanish M. T., Gagnier L., Dunn C. A., van de Lagemaat L. N. and Mager D. L. 2006. Retroviral Elements and Their Hosts: Insertional Mutagenesis in the Mouse Germ Line. *PLoS Genet* 2(1):e2.
- Malone, C. D. and Hannon G. J. 2009. Small RNAs as guardians of the genome. *Cell* 136:656–668.
- Mandal, P. K. and Kazazian H. H. J. 2008. SnapShot: Vertebrate transposons. *Cell* 135:192–192.e1.
- Market, E. and Papavasiliou F. N. 2003. V (D) J Recombination and the Evolution of the Adaptive Immune System. *PLoS Biol* 1:e16.
- Martienssen, R. A. 2010. Heterochromatin, small RNA and post-fertilization dysgenesis in allopolyploid and interploid hybrids of *Arabidopsis*. *New Phytol* 186:46–53.
- Mayr, E. 1942. *Systematics and the origin of species, from the viewpoint of a zoologist*. Columbia Univ. Press, New York.
- McClintock, B. 1956. Controlling Elements and the Gene. *Cold Spring Harb Symp Quant Biol* 21:197–216.
- McClintock, B. 1950. The Origin and Behavior of Mutable Loci in Maize. *Proc Natl Acad Sci USA* 36(6):344.
- McClintock, B. 1984. The Significance of Responses of the Genome to Challenge. *Science* 226:792–801.
- McCormick, S. 1993. Male gametophyte development. *Plant Cell* 5:1265–1275.
- McGaugh, S. E. and Noor M. A. F. 2012. Genomic impacts of chromosomal inversions in parapatric *Drosophila* species. *Philos. Trans. R. Soc. Lond., B, Biol. Sci.* 367:422–429.
- Metcalfe, C. J., Bulazel K. V., Ferreri G. C., Schroeder-Reiter E., Wanner G., Rem W., Obergfell C., Eldridge M. D. B. and O'Neill R. J. 2007. Genomic instability within centromeres of interspecific marsupial hybrids. *Genetics* 177:2507–2517.
- Michalak, P. 2009. Epigenetic, transposon and small RNA determinants of hybrid dysfunctions. *Heredity* 102:45–50.
- Michel, A. P. 2006. Divergence With Gene Flow in *Anopheles funestus* From the Sudan Savanna of Burkina Faso, West Africa. *Genetics* 173:1389–1395.
- Nadeau, N. J., Whibley A., Jones R. T., Davey J. W., Dasmahapatra K. K., Baxter S. W., Quail M. A., Joron M., French-Constant R. H., Blaxter M. L., Mallet J. and Jiggins C. D. 2012. Genomic islands of divergence in hybridizing *Heliconius* butterflies identified by large-scale targeted sequencing. *Philos. Trans. R. Soc. Lond., B, Biol. Sci.* 367:343–353.
- Naito, K., Zhang F., Tsukiyama T., Saito H., Hancock C. N., Richardson A. O., Okumoto Y., Richardson A. O., Okumoto Y., Tanisaka T. and Wessler S. R. 2009. Unexpected consequences of a sudden and massive transposon amplification on rice gene expression. *Nature* 461:1130–1134.

- Navarro, A. and Barton N. H. 2003. Accumulating postzygotic isolation genes in parapatry: a new twist on chromosomal speciation. *Evolution* 57:447–459.
- Noor, M. A. F. and Bennett S. M. 2009. Islands of speciation or mirages in the desert? Examining the role of restricted recombination in maintaining species. *Heredity* 103:439–444.
- Noor, M. A. F., Grams K., Bertucci L. and Reiland J. 2001. Chromosomal inversions and the reproductive isolation of species. *Proc Natl Acad Sci USA* 98:12084–12088.
- Nosil, P. and Schluter D. 2011. The genes underlying the process of speciation. *Trends Ecol. Evol.* 26:160–167.
- O'Donnell, K. A. and Boeke J. D. 2007. Mighty Piwis Defend the Germline against Genome Intruders. *Cell* 129:37–44.
- O'Neill, R. J., O'Neill M. J. and Graves J. A. M. 1998. Undermethylation associated with retroelement activation and chromosome remodelling in an interspecific mammalian hybrid. *Nature* 393:68–72.
- Oliver, K. R. and Greene W. K. 2011. Mobile DNA and the TE-Thrust hypothesis: supporting evidence from the primates. *Mob DNA* 2:8.
- Oliver, K. R. and Greene W. K. 2009. Transposable elements: powerful facilitators of evolution. *Bioessays* 31:703–714.
- Olmedo-Monfil, V., Durán-Figueroa N., Arteaga-Vázquez M., Demesa-Arévalo E., Autran D., Grimanelli D., Slotkin R. K., Martienssen R. A. and Vielle-Calzada J. P. 2010. Control of female gamete formation by a small RNA pathway in *Arabidopsis*. *Nature* 464:628–632.
- Orr, H. A. 1996. Dobzhansky, Bateson, and the genetics of speciation. *Genetics* 144:1331–1335.
- Orr, H. A. and Turelli M. 2001. The evolution of postzygotic isolation: Accumulating Dobzhansky-Muller incompatibilities. *Evolution* 55:1085–1094.
- Ortiz-Barrientos, D., Counterman B. and Noor M. A. F. 2004. The genetics of speciation by reinforcement. *PLoS Biol* 2:2256–2263.
- Ortiz-Barrientos, D. and Noor M. A. F. 2005. Evidence for a one-allele assortative mating locus. *Science* 310:1467–1467.
- Pascale, E., Valle E. and Furano A. V. 1990. Amplification of an ancestral mammalian L1 family of long interspersed repeated DNA occurred just before the murine radiation. *Proc Natl Acad Sci USA* 87:9481–9485.
- Pelisson, A. 1981. The *I-R* system of hybrid dysgenesis in *Drosophila melanogaster*: Are *I* factor insertions responsible for the mutator effect of the *I-R* interaction? *Mol Gen Genet* 183:123–129.
- Pelisson, A. and Bregliano J. C. 1987. Evidence for rapid limitation of the *I* element copy number in a genome submitted to several generations of *I-R* hybrid dysgenesis in *Drosophila melanogaster*. *Mol Gen Genet* 207:306–313.
- Pelisson, A., Song S. U., Prud'homme N., Smith P. A., Bucheton A. and Corces V. G. 1994. *Gypsy transposition correlates with the production of a retroviral envelope-like protein under the tissue-specific control of the Drosophila flamenco gene*. 13:4401.
- Picard, G. and L'Heritier P. 1971. A maternally inherited factor inducing sterility in *D. melanogaster*. *Drosophila Information Service* 46:54.
- Picard, G. 1976. Non-Mendelian female sterility in *Drosophila melanogaster*: hereditary transmission of *I* factor. *Genetics* 83:107–123.

- Prud'homme, N., Gans M., Masson M., Terzian C. and Bucheton A. 1995. Flamenco, a gene controlling the gypsy retrovirus of *Drosophila melanogaster*. *Genetics* 139:697–711.
- Quinlan, A. R., Clark R. A., Sokolova S., Leibowitz M. L., Zhang Y., Hurles M. E., Mell J. C. and Hall I. M. 2010. Genome-wide mapping and assembly of structural variant breakpoints in the mouse genome. *Genome Research* 20:623–635.
- Ray, D. A., Feschotte C., Pagan H. J. T., Smith J. D., Pritham E. J., Arensburger P., Atkinson P. W. and Craig N. L. 2008. Multiple waves of recent DNA transposon activity in the bat, *Myotis lucifugus*. *Genome Res* 18:717–728.
- Reiss, D., Josse T., Anxolabehere D. and Ronsseray S. 2004. *aubergine* mutations in *Drosophila melanogaster* impair *P* cytotype determination by telomeric *P* elements inserted in heterochromatin. *Molecular Genetics and Genomics* 272:336–343.
- Renaut, S., Nolte A. W. and Bernatchez L. 2010. Mining transcriptome sequences towards identifying adaptive single nucleotide polymorphisms in lake whitefish species pairs (*Coregonus* spp. Salmonidae). *Mol Ecol* 19: 115–131.
- Ribeiro, J. M. and Kidwell M. G. 1994. Transposable elements as population drive mechanisms: specification of critical parameter values. *J Med Entomol* 31:10–16.
- Rieseberg, L. H. 2001. Chromosomal rearrangements and speciation. *Trends Ecol. Evol.* 16:351–358.
- Rose, M. R. and Doolittle W. F. 1983. Molecular biological mechanisms of speciation. *Science* 220:157–162.
- Rubin, G. M., Kidwell M. G. and Bingham P. M. 1982. The molecular basis of *P-M* hybrid dysgenesis: The nature of induced mutations. *Cell* 29:987–994.
- Sarot, E., Payen-Groschene G., Bucheton A. and Pelisson A. 2004. Evidence for a piwi-dependent RNA silencing of the gypsy endogenous retrovirus by the *Drosophila melanogaster* flamenco gene. *Genetics* 166:1313–1321.
- Schaefer, R. E., Kidwell M. G. and Fausto-Sterling A. 1979. Hybrid Dysgenesis in *Drosophila melanogaster*: Morphological and Cytological Studies of Ovarian Dysgenesis. *Genetics* 92:1141–1152.
- Schmidt, D., Schwalie P. C., Wilson M. D., Ballester B., Gonçalves Â., Kutter C., Brown G. D., Marshall A., Flicek P. and Odom D. T. 2012. Waves of Retrotransposon Expansion Remodel Genome Organization and CTCF Binding in Multiple Mammalian Lineages. *Cell* 148:335–348.
- Schwarzbach, A. E. and Rieseberg L. H. 2002. Likely multiple origins of a diploid hybrid sunflower species. *Mol Ecol* 11:1703–1715.
- Shan, X., Liu Z., Dong Z., Wang Y., Chen Y., Lin X., Long L., Han F., Dong Y. and Liu B. 2005. Mobilization of the active MITE transposons mPing and Pong in rice by introgression from wild rice (*Zizania latifolia* Griseb.). *Mol Biol Evol* 22:976–990.
- Shedlock, A. Takahashi M., K. and Okada N. 2004. SINEs of speciation: tracking lineages with retroposons. *Trends Ecol. Evol.* 19:545–553.
- Slotkin, R. K., Freeling M. and Lisch D. 2005. Heritable transposon silencing initiated by a naturally occurring transposon inverted duplication. *Nat. Rev. Genet.* 37:641–644.
- Slotkin, R. K., Freeling M. and Lisch D. 2003. *Mu* killer causes the heritable inactivation of the Mutator family of transposable elements in *Zea mays*. *Genetics* 165:781–797.
- Slotkin, R. K., Vaughn M., Borges F., Tanurdžić M., Becker J. D., Feijó J. A. and Martienssen R. A. 2009. Epigenetic Reprogramming and Small RNA Silencing of Transposable Elements in Pollen. *Cell* 136:461–472.

- Thompson, J. N., Woodruff R. C. and Schaefer G. B. 1978. An assay of somatic recombination in male recombination lines of *Drosophila melanogaster*. *Genetica* 49:77–80.
- Thomson, M. and Beeman R. W. 1999. Assisted suicide of a selfish gene. *J Hered* 90:191–194.
- Thomson, M. S., Friesen K. S., Denell R. E. and Beeman R. W. 1995. A hybrid incompatibility factor in *Tribolium castaneum*. *J Hered* 86:6.
- Turelli, M. and Orr H. A. 2000. Dominance, epistasis and the genetics of postzygotic isolation. *Genetics* 154:1663–1679.
- Turelli, M. and Moyle L. C. 2007. Asymmetric postmating isolation: Darwin's corollary to Haldane's rule. *Genetics* 176:1059–1088.
- Turner, T. L. and Hahn M. W. 2010. Genomic islands of speciation or genomic islands and speciation? *Mol Ecol* 19:848–850.
- Turner, T. L., Hahn M. W. and Nuzhdin S. V. 2005. Genomic Islands of Speciation in *Anopheles gambiae*. *PLoS Biol* 3:e285.
- Ungerer, M. C., Strakosh S. C. and Stimpson K. M. 2009. Proliferation of Ty3/gypsy-like retrotransposons in hybrid sunflower taxa inferred from phylogenetic data. *BMC Biol* 7:40.
- Ungerer, M. C., Strakosh S. C. and Zhen Y. 2006. Genome expansion in three hybrid sunflower species is associated with retrotransposon proliferation. *Curr. Biol.* 16:R872–R873.
- Venner, S., Feschotte C. and Biémont C. 2009. Dynamics of transposable elements: towards a community ecology of the genome. *Trends Genet* 25: 317–323.
- Verneau, O., Catzeflis F. and Furano A. V. 1998. Determining and dating recent rodent speciation events by using L1 (LINE-1) retrotransposons. *Proc Natl Acad Sci USA* 95:11284–11289.
- Wade, M. J. and Beeman R. W. 1994. The population dynamics of maternal-effect selfish genes. *Genetics* 138:1309–1314.
- Walsh, J. B. 1982. Rate of Accumulation of Reproductive Isolation by Chromosome Rearrangements. *Am. Nat.* 120:510–532.
- Wang, N., Wang H., Zhang D., Wu Y., Ou X., Liu S., Dong Z. and Liu B. 2010. Transpositional reactivation of the Dart transposon family in rice lines derived from introgressive hybridization with *Zizania latifolia*. *BMC Plant Biol* 10:190.
- Wang, S., Lorenzen M. D., Beeman R. W. and Brown S. J. 2008. Analysis of repetitive DNA distribution patterns in the *Tribolium castaneum* genome. *Genome Biol* 9:R61.
- Welch, M. E. and Rieseberg L. H. 2002. Patterns of Genetic Variation Suggest a Single, Ancient Origin for the Diploid Hybrid Species *Helianthus paradoxus*. *Evolution* 56:2126–2137.
- Wessler, S. R. 1996. Plant retrotransposons: Turned on by stress. *Cell* 6:959–961.
- Wicker, T., Sabot F., Hua-Van A., Bennetzen J. L., Capy P., Chalhoub B., Flavell A., Leroy P., Morgante M., Panaud O., Paux E., San Miguel P. and Schulman A. H. 2007. A unified classification system for eukaryotic transposable elements. *Nat. Rev. Genet.* 8:973–982.
- Woodruff, R. C. and Thompson J. N. 1977. An analysis of spontaneous recombination in I males. *Heredity* 38:291–307.
- Wright, S. 1931. Evolution in Mendelian Populations. *Genetics* 16:97–159.
- Wright, S. 1932. The Roles of Mutation, Inbreeding, Crossbreeding, and Selection in Evolution. in D. F. Jones, ed. *Brooklyn Botanic Garden*, Ithaca, New York.
- Xing, J., Zhang Y., Han K., Salem A. H., Sen S. K., Huff C. D., Zhou Q., Kirkness E. F., Levy S. and Batzer M. A. 2009. Mobile elements create structural variation: analysis of a complete human genome. *Genome Research* 19:1516–1526.

- Yamaguchi, O., Cardellino R. A. and Mukai T. 1976. High Rates of Occurrence of Spontaneous Chromosome Aberrations in *Drosophila melanogaster*. *Genetics* 83:409.
- Zeh, D. W., Zeh J. A. and Ishida Y. 2009. Transposable elements and an epigenetic basis for punctuated equilibria. *Bioessays* 31:715–726.
- Zhang, J., Yu C., Pulletikurti V., Lamb J., Danilova T., Weber D. F., Birchler J. and Peterson T. 2009. Alternative *Ac/Ds* transposition induces major chromosomal rearrangements in maize. *Gene Dev* 23:755–765.
- Zhou, L., Mitra R., Atkinson P. W., Burgess Hickman A., Dyda F. and Craig N. L. 2004. Transposition of hAT elements links transposable elements and V (D) *J recombination*. 432:995–1001.

Chapter 11

HYBRID INCOMPATIBILITY: ARE CELLULAR PROCESSES THE BATTLEFIELDS OF GENOMIC CONFLICT?

Norman A. Johnson

Department of Biology, Department of Environmental Conservation,
Graduate Program in Organismal and Evolutionary Biology,
University of Massachusetts, Amherst, MA, US

ABSTRACT

Recent studies suggest that much hybrid incompatibility results as a consequence of conflict between different components of the genome (genomic conflict). In this chapter, I argue that the battlegrounds for much of the conflict-driven hybrid incompatibility consist of various cellular structures and processes. These include the recombination machinery, centromeres and kinetochores, and heterochromatin and its associated proteins. I conclude with a call to integration between cell biology and evolutionary biology in the study of the evolutionary genetics of hybrid incompatibility.

1. HYBRID INCOMPATIBILITY

"The spread within a population of genes that may eventually induce isolation between populations is probably due to their properties other than those concerned with isolation." (Dobzhansky 1937, p. 258).

Hybrids between closely related species often have reduced fertility or viability, sometimes to the extent of being completely sterile or inviable (reviewed in Coyne and Orr 2004, chapters 7 and 8). These hybrids can also exhibit abnormalities in morphology (e. g., Wade and Johnson 1994), behavior (e.g., Noor 1997), and/or physiology (e.g., Willet 2011). Collectively known as hybrid incompatibility (HI), these fitness reductions and associated abnormalities, continue to draw the attention of evolutionary biologists. This attention arises in large part because hybrid incompatibility is a form of reproductive isolation, and thus central to speciation.

HI is also of interest because its evolution appears paradoxical: why would not being able to produce offspring be favored by natural selection? Both Darwin and Wallace weighed in on this puzzle (reviewed in Cronin 1991; Johnson 2008). Wallace thought there were circumstances where HI (notably sterility) could be favored by natural selection, but Darwin maintained that hybrid incompatibility was only a consequence of other adaptive or neutral changes (see Cronin 1991, for more information). Subsequent studies of HI have largely vindicated Darwin's position (Dobzhansky 1937; Muller 1942; Coyne and Orr 2004). With only rare exceptions (discussed in Johnson 2008; Turner et al. 2011), HI evolves as a by-product of evolutionary divergence, and not because of direct selection.

What, then, is the nature of this evolutionary divergence that underlies HI? At the start of the 21st century, that answer appeared to be that differences in ecological factors led to selectively driven changes in nascent species (Schluter 2000, 2009; Coyne and Orr 2004). This supposition was supported in part by ecological studies showing that divergent natural selection could lead to reproductive isolation (Nosil 2012). More support comes from a comparative study by Funk et al. (2006) showed that the amount of reproductive isolation between populations or species correlated with the extent of ecological divergence, once genetic distance was factored out. Finally, several genes connected to HI in *Drosophila* showed the signature of positive selection as revealed by the McDonald-Kreitman test (e.g., Ting et al. 1998; Barbash et al. 2003; Presgraves et al. 2003).

More recent studies challenge the generality of this ecological explanation for the evolution of HI. The discovery of new genes that contribute to HI and further analysis of previously studied ones suggest that few of these HI genes diverged due to standard ecological differences (Presgraves 2007; 2010; Johnson 2010). Instead, many of these genes appear to be related to conflict between different parts of the genome (Johnson 2010; Presgraves 2010; Mahshwari and Barbash 2011; Crespi and Nosil 2013).

2. CONFLICT-DRIVEN HI

"The unity of the organism is an approximation, undermined by these continuously emerging selfish elements within their alternative, narrowly self-benefitting means for boosting transmission to the next generation. The result: a parallel universe of (often intense) sociogenetic interactions *within* the individual organism--a world that evolves according to its own rules, as modulated by the sexual and social lives of the hosts and Mendelian systems that act in part to suppress them." (Burt and Trivers 2006, p. 475).

The evolutionary interests of different parts of the genome vary, and these differing interests lead to what is known as genetic or genomic conflict (Burt and Trivers 2006). Although genomic conflict has been recognized since the modern evolutionary synthesis of the middle-20th century (e.g., Ostergren 1945), its importance as an evolutionary force has been recognized only fairly recently (Burt and Trivers 2006). More specifically, various authors have proposed genomic conflict as an explanation for HI over the last thirty years (e.g., Rose and Doolittle 1983; Frank 1991; Hurst and Pomiankowski 1991), but these claims were generally dismissed by those studying HI (e.g., Johnson and Wu 1992; Coyne and Orr 1993). Additional data in recent years has led to a re-examination of the role of genomic conflict to the evolution of HI, and thus, speciation (McDermott and Noor 2010).

In a recent review paper, Crespi and Nosil (2013) coined the term conflictual speciation, and defined it as "the evolution of reproductive isolation as a byproduct of antagonistic selection among genomic elements with divergent fitness interests". Crespi and Nosil present conflictual speciation as one of the alternatives to ecological speciation, the process by which reproductive isolation evolves as a consequence of divergent natural selection (Schluter 2000, 2009; Nosil 2012). Both ecological speciation (Schluter 2000, 2009; Via and West 2009; Nosil 2012) and conflictual speciation (Johnson 2010; Presgraves 2010; Crespi and Nosil 2013; see below) are supported by several lines of evidence.

One line of support for the importance of ecological speciation comes from a meta-analysis by Funk et al. (2006). Prior studies had examined patterns of the accumulation of reproductive isolation between pairs of species in a variety of taxa including *Drosophila* frogs, butterflies, fish, birds, and plants (reviewed in Coyne and Orr 2004; Johnson 2006). These studies generally found an increase in indices of reproductive isolation with increasing genetic distance (a proxy for time). Funk et al. (2006) found that when genetic distance was factored out of the analysis, that species with greater ecological differences evolved more reproductive isolation between them than species with lesser ecological divergence. They then claimed that this result supports ecological speciation. Although this logic is sound, a closer examination of the data (Funk et al. 2006) shows that the correlation between reproductive isolation and ecological divergence is considerably stronger for premating reproductive isolation than it is for postmating reproductive isolation (i.e., HI). Thus, ecological speciation appears to be more prevalent for premating than for postmating isolation.

Many HI genes do show a signature of repeated selection based on McDonald-Kreitman (1991) and similar tests. Such a signature could arise from the selection associated with genomic conflict instead of ecological adaptation (Johnson 2010; Presgraves 2010). In fact, we would expect such signatures to occur more often when the selection involves conflict than when it involves ecological adaptation. The rationale is that conflict-associated selection is more likely to occur in repeated bouts than selection associated with ecological changes because conflict entails co-evolving units of the genome (Presgraves 2010; Crespi and Nosil 2013).

I argue below that conflict-driven HI involves cellular functions and processes. Thus, to understand the mechanics of HI, we need more studies of the evolutionary biology of the cell.

A surprisingly large number of genes or genetic regions involved in hybrid incompatibility are either heterochromatin or bind heterochromatin (reviewed in Johnson 2010; Johnson and Lachance 2012; Maheshari and Barbash 2011; Sawamura 2012). Heterochromatin is not an inert part of the genome; in fact, it plays several important roles including regulating gene transcription and chromosome segregation. Notably, chromosome segregation abnormalities can result from improper heterochromatin formation (Yasuhara and Wakimoto 2006).

One HI gene that is linked to heterochromatin is the *Odysseus* gene, a gene that is involved in the sterility of male hybrids of *Drosophila simulans* and *D. mauritiana*. It was one of the first hybrid incompatibility genes to be molecularly characterized, and was shown to contain a homeobox (Perez et al. 1993; Ting et al. 1998). Thus, it was reasonable to assume that *Odysseus* encodes a transcription factor, and the associated hybrid sterility arose from dysfunctions in transcriptional regulation (reviewed in Oritz-Barrientos et al. 2007). Recently, Bayes and Malik (2009) discovered that *Odysseus* has another role: its protein product binds heterochromatin, and inappropriate binding of heterochromatin appears to be associated with the hybrid sterility.

In this volume, Phandis and Malik discuss studies of *Odysseus* in further detail. Phandis and Malik (this volume) also discuss the HI gene *Overdrive*, a gene involved in both hybrid sterility and meiotic drive in hybrids between *Drosophila* species. While genomic conflict certainly has driven the evolution of *Overdrive*, the role (if any) of cellular processes for this gene is unknown.

Below I focus on genomic conflict and its possible relationships with HI genes in two arenas: recombination-associated genes (section III) and centromere/ kinetochore-associated genes (section IV).

3. PRDM9 AND OTHER RECOMBINATION-ASSOCIATION GENES

"The 'hotspot paradox' states that due to biased gene conversion, a hotspot tends to kill itself, nevertheless, there are many hotspots in extant genomes." (Wu et al., 2012, p. 2)

Some F1 male hybrid genotypes between *Mus musculus musculus* females and *M. m. domesticus* males result in complete sterility. Other F1 male hybrid genotypes are fertile, to varying degrees; a substantial part of that variation is due to what was called the *Hst-1* (*Hybrid sterility 1*) locus (Forejt and Ivanyi 1974; Mihola et al. 2009). Recent studies have shown that *Hst-1* is *Paired domain number 9* (*Prdm9*), which is also known as *Meisitz* (Meiotic gene with SET/PR domain and Zinc fingers) (Mihola et al. 2009; Flachs et al. 2012). Thus far, *Prdm9* is the only mammalian gene known to contribute to hybrid incompatibility that has been characterized at the molecular level (Mihola et al. 2009; Flachs et al. 2012); its sterility phenotype is manifest as meiotic arrest in the pachytene stage of meiosis I (Flachs et al. 2012, and references within). F1 males from the reciprocal cross (*M. m. d.* females x *M. m. m.* males) are fertile, but suffer from reduced fecundity, as compared to pure species males (Mihola et al. 2009; Flachs et al. 2012).

A recent transgenic study showed *Prdm9* can rescue fertility in sterile hybrid males and the rescue effect correlates with dosage (Flachs et al. 2012). The sensitivity to *Prdm9* varied across genetic backgrounds. Moreover, reciprocal hybrids (which are mostly fertile) are also sensitive to *Prdm9*. Interestingly, the hybrid incompatibility effect of *Prdm9* does not appear to be due to its own transcript (Flachs et al. 2012).

One of the roles of *Prdm9* concerns the regulation of recombination hotspots. In mammals, recombination events at the local scale are concentrated in regions referred to as "hotspots" (Paigen and Petkov 2010). Such hotspots often have associated short motifs (10-15 nucleotides), and recombination appears to involve the binding of proteins to the sequences associated with the hotspots. In addition to a SET methyltransferase domain (Ponting 2011), *Prdm9* also encodes a variable number of zinc finger domains (Segurel et al. 2011). These zinc finger domains bind preferentially to recombination hotspots (Paigen and Petkov 2010; Wu et al. 2012).

The exact mechanism by which *Prdm9* affects hybrid fertility is not yet completely worked out, but the number of zinc finger repeats in the protein is associated with the sterility phenotype (Mihola et al. 2009). It should also be noted that the SET methyl transferase domain methylates histones, and thus can alter chromatin composition (Ponting 2011).

Why would we expect genomic conflict to be involved in the evolution of *Prdm9* and other recombination-associated genes? The answer is related to what has been called the "hotspot

paradox" (Friberg and Rice 2008; Paigen and Petkov 2010; Ubeda and Wilkins 2011). Recombination hotspots induce double stranded breaks, and can cause biased gene conversion, with the result that the hotspots are converted into coldspots (Paigen and Petkov 2010). The paradoxical result is that hotspots cause their own destruction. Yet, hotspots exist. The apparent resolution of the hotspot paradox comes from recombination-modifiers like *Prdm9* that cause specific DNA sequences to be hotspots. Evolutionary changes in recombination-modifiers create new hotspots (Friberg and Rice 2008; Ubeda and Wilkins 2011). In this model, the conflict arises over the level of genetic recombination, with coldspots driving against hotspots and *Prdm9* and other recombination modifiers acting as modulators of the conflict (Ubeda and Wilkins 2011). Jeffreys and Neumann (2005) provide evidence showing that cold alleles are preferentially transmitted over hot alleles. This process is one akin to the Red Queen "arms races", where there is relative stability in the overall recombination rate, but a dynamic turnover of hotspots, coldspots, and sites within the recombination-modifiers (Ubeda and Wilkins 2011).

Consistent with this Red Queen model, the zinc finger domains of *Prdm9* are rapidly evolving in several lineages of mammals (Oliver et al. 2009; Ponting 2011). Additional evidence comes from dogs and their close relatives, which lack functionality in the zinc finger domain of *Prdm9* (Munoz-Fuentes et al. 2011). Associated with this loss of *Prdm9* function, the recombination hotspots are more stable in canids than they are in primates (Axelsson et al. 2012). Incidentally, there is no *Prdm9*-like mechanism for recombination hotspots in *Drosophila* (Heil and Noor 2012). Variation at *Prdm9* also affects genome instability (rearrangements) with some variants within humans being associated with varying risk for genetic abnormalities and non-disjunction (Berg et al. 2010; Borel et al. 2012).

Prdm9's binding sites explain only about 18% of the variation in mouse recombination (Wu et al. 2012, and references within). Thus, we should expect other proteins to also contribute to recombination hotspots. Some of these proteins may be candidates for HI genes. In a bioinformatic analysis, Wu et al. (2012) search for other modifiers of recombination. They looked for transcription factors that bind preferentially to hotspots. A Gene Ontology (GO) analysis reveals that these hotspot-binders were more likely to play a role in recombination, as well as in histone modification and other epigenetic functions (Wu et al. 2012). This study also pointed to a few candidate genes, one of which is *ZFX*, an X-linked gene that encodes a zinc finger domain. Interestingly *ZFX* appears to have spermatogenic function, and mutations in it in mice can lead to reduced germ cell number in both males and females (Wu et al. 2012). There are no known cases of *ZFX* contributing to hybrid sterility, but based on its features as a binder of recombination hotspots and spermatogenesis function, a logical place to look for genes involved in hybrid sterility is *ZFX*.

4. THE CENTROMERE AND ASSOCIATED STRUCTURES

"A central paradox in biology is that organisms must remain stable to maintain the integration of complex developmental and functional systems, but they must also adapt to accommodate changing environments" (Schwenk et al., 2009, p. 11)

As evidenced by the quote above, whole organism evolutionary physiologists recognize a tension between the competing needs of evolutionary flexibility (or evolvability) and stability (homeostasis). The same is also true with cellular structures and functions. One place where this

homeostasis/ evolvability tension is particularly acute occurs in the machinery used in cell division: the centromere and kinetochore structures (Burrack and Berman 2012). During mitosis and meiosis, sister chromatids are attached at the centromere. The kinetochore, which is assembled on the centromere, is a complex of proteins on the chromatids where the spindle fibers attach, allowing for the chromatids to segregate to different poles. Thus, these structures play crucial roles in chromosome segregation, a complicated and intricate process (Burrack and Berman 2012). If this process does not occur properly, a likely result is aneuploidy, a deviation from the standard component of chromosomes in a cell. Like mutations, aneuploidy can be advantageous in rare occurrences, it is much more often deleterious (ref).

The kinetochore can involve at least a hundred different proteins, some of which must interact with others in order to properly assemble the complex. Moreover, physiological changes can affect kinetochore assembly; thus, the process must be buffered against perturbations (Burrack and Berman 2012). Both centromeres and kinetochores can evolve rather rapidly, as seen by variation in features of these structures even among closely related species, and sometimes even within species. This diversity is reviewed in Burrack and Berman (2012).

Centromere DNA evolves at rates much higher than rates at other non-coding regions (Burrack and Berman 2012). At least some of the increased rate of evolution of centromeric DNA arises from genomic conflict, specifically the conflict that arises over asymmetries in female meiosis (Burt and Trivers 2006; Malik 2009). During meiosis in females, one of the four cells will become the functional gamete (the egg), while the other three are "dead end" polar bodies. There is, thus, intense competition for chromatids to get in the functional egg as opposed to the polar bodies, a process that is called centromere drive (Henikoff et al. 2001; Malik 2009). One mechanism by which drive could arise is by centromeres varying in how well they attract microtubules (Henikoff et al. 2001; Malik 2009).

The rest of the genome would have an evolutionary advantage to suppress such centromere drive (Burt and Trivers 2006). Heterochromatic proteins may thus evolve as drive suppressors, which is consistent with the rapid evolution of some proteins associated with heterochromatin (e.g., Vernack and Malik 2009). Interestingly, a large number of genes involved in HI either are centromeric heterochromatin or bind to it (Johnson 2010; Presgraves 2010). Thomas et al. (2009) also suggested that the recombination-associated HI gene, *Prdm9* (see section III), may modulate centromeric drive.

The kinetochores may also be involved in HI. Two nucleoporins (Nup) genes, *Nup 96* and *Nup 160* were found in screens for hybrid inviability between *Drosophila melanogaster* and *D. simulans* (Presgraves et al. 2003; Tang and Presgraves 2009). Recently, Sawamura et al. (2010) showed that *Nup160* also contributes to hybrid female sterility in this species pair. Nup genes are involved in the selective transport between the nucleus and the cytoplasm via the nuclear pore complex (NPC) (deSouza and Osmani 2009). *Nup160* and other Nups are also involved in kinetochore assembly (Zuccollo et al. 2007) An additional line of evidence linking Nup genes to genomic conflict is the similarity between Nups and the *SD* segregation disorder system in *D. melanogaster* (Presgraves 2007).

5. A CALL FOR THE EVOLUTIONARY GENETICS OF THE CELL

"Must we geneticists become bacteriologists, physiological chemists and physicists, simultaneously with being zoologists and botanists? Let us hope so." (Muller 1922, p. 50)

Ninety years after Muller's remarks, let us now hope that evolutionary geneticists interested in hybrid incompatibility become cell biologists, along with all of the other hats that they wear. Let us also hope that cellular and molecular genetics interested in sterility also become evolutionary biologists. Hybrid incompatibility arises when the cellular processes break down when diverged genomes come together in the same cell. An understanding of how the cellular processes work, and how they can break down, appears ever more important to understanding hybrid incompatibility.

As shown above, many of the HI processes appear to be associated with genomic conflict. However, genomic conflict is not always involved. For instance, the processes of dosage compensation and Meiotic Sex Chromosome Inactivation (MSCI), which are related to the hemizyosity of sex chromosomes in the heterogametic sex, appear to play a large role in HI (Johnson and Lachance 2012). Yet, these processes are probably unrelated to genomic conflict. Ecological adaptation and compensatory processes also contribute to HI. Still, genomic conflict looms large. A detailed understanding of the cell will likely shed light on how these evolutionary processes lead to HI.

REFERENCES

- Axelsson E, Webster MT, Ratnakumar A, Ponting CP, Lindbald-Toh K. 2012. Death of PRDM9 coincides with stabilization of the recombination landscape in dogs. *Gen. Res.* 22: 51-63.
- Barbash DA, Siino DF, Tarone AM, Roote J. 2003. A rapidly-evolving MYB-related protein causes species isolation in *Drosophila*. *Proc. Nat. Acad. Sci. USA.* 100: 5302-5307.
- Bayes JJ, Malik HS. 2009. Altered heterochromatin binding in a hybrid sterility protein in *Drosophila* sibling species. *Science* 326: 1538-1541.
- Berg II, Neumann R, Lam KWG, Sarbajna S, Odenthal-Hesse L, May CA, Jeffreys AJ. 2010. PRDM9 variation strongly influences recombination hot-spot activity and meiotic instability in humans. *Nat. Genet.* 42: 859-864.
- Borel C, Cheung F, Stewart H, Koolen DA, Phillips C, Thomas NS, Jacobs PA, Eliez S, Sharp AJ. 2012. Evaluation of PRDM9 as a risk factor for recurrent genomic disorders and chromosomal non-disjunction. *Hum. Genet.* 131: 1519-1524.
- Burrack LS, Berman J. 2012. Flexibility of centromere and kinetochore structures. *Trends Genet.* 28: 204-212.
- Burt A, Trivers R. 2006. *Genes in Conflict: The Biology of Selfish Genetic Elements*. The Belknap Press of Harvard University Press, Cambridge, MA.
- Coyne J A, Orr HA. 1993. Further evidence against meiotic-drive models of hybrid sterility. *Evolution* 47: 685-687.
- Coyne J A, Orr HA. 2004. *Speciation*. Sinauer Associates, Sunderland, MA.
- Cronin H. 1991. *The Ant and the Peacock: Altruism and Sexual Selection from Darwin to Today*. Cambridge University Press.

- Crespi B, Nosil P. 2013. Conflictual speciation: species formation via genomic conflict. *Trends Ecol. Evol.* (in press).
- De Souza MA, Osmani SA. 2009. Double duty for nuclear proteins-- the price of more open forms of mitosis. *Trends Genet.* 25: 545-554.
- Dobzhansky T. 1937. *Genetics and the Origin of Species*. Columbia University Press, New York.
- Flachs P, Mihola O, Simecek P, et al. (11 co-authors). 2012. Interallelic and intergenic incompatibilities of the *Prdm9 (Hst1)* gene in mouse hybrid sterility. *PLoS Genet.* 8: e1003044.
- Forejt J. and Ivanyi P. 1974. Genetic studies on male sterility of hybrids between laboratory and wild mice (*Mus musculus L.*) *Genet. Res.* 24: 189-206.
- Friberg U. and Rice WR. 2008. Cut thy neighbor: cyclic birth and death recombination hotspots via genetic conflict. *Genetics* 179: 2229-2238.
- Frank SA. 1991. Divergence in meiotic drive-suppressors as an explanation for sex-biased hybrid sterility and inviability. *Evolution* 45: 262-267.
- Funk DJ P Nosil, Etges WJ. 2006. Ecological divergence is consistently positively associated with reproductive isolation across disparate taxa. *Proc. Nat. Acad. Sci. USA.* 103: 3209-3213.
- Heil CSS, Noor MAF. 2012. Zinc finger binding motifs do not explain recombination rate variation within or between species of *Drosophila*. *PLoS ONE* 7: e45055.
- Henikoff S, Ahmad K, Malik HS. 2001. The centromere paradox: stable inheritance with rapidly evolving DNA. *Science* 293: 1098-1102.
- Hurst LD, Pomiankowski A. 1991. Causes of sex-ratio bias may account for the unisexual sterility in hybrids: a new explanation of Haldane's rule and related phenomena. *Genetics* 128: 841-858.
- Jeffreys AJ, Neumann R. 2005. Factors influencing recombination frequency and distribution in a human meiotic crossover hotspot. *Hum. Mol. Genet.* 14: 2277-2287.
- Johnson NA. 2006. The evolution of reproductive isolating barriers. Pp. 374-386 in *Evolutionary Genetics: Concepts and Case Studies*. (Fox CW, Wolf JB. eds.) Oxford University Press, New York.
- Johnson NA. 2008. Direct selection for reproductive isolation: the Wallace effect and reinforcement. Pp. 114-124 in *Alfred Russel Wallace* (Smith CH, Beccaloni GW, eds.) Oxford University Press, New York.
- Johnson NA. 2010. Hybrid incompatibility genes: remnants of a genomic battlefield? *Trends Genet.* 26: 319-325.
- Johnson, NA, Lachance J 2012. The genetics of sex chromosomes: evolution and implications for hybrid incompatibility. *Ann. NY. Acad. Sci.* 1256: e1-e22.
- Johnson NA, Wu CI. 1992. An empirical test of the meiotic drive models of hybrid sterility: Sex-ratio data from hybrids between *Drosophila simulans* and *Drosophila sechellia*. *Genetics* 130: 507-511.
- Maheshwari S. and Barbash DA. 2011. The genetics of hybrid incompatibilities. *Ann. Rev. Genet.* 45: 331-355.
- Malik HS. 2009. The centromere-drive hypothesis: a simple basis for centromere complexity. *Prog. Mol. Subcellular Biol.* 48: 33-52.
- McDermott SR, Noor MAF. 2010. The role of meiotic drive in hybrid male sterility. *Phil. Trans. Roy. Soc. Lond. B.* 365:1265-1272.

- McDonald JH, Kreitman M. 1991. Adaptive protein evolution at the *Adh* locus in *Drosophila*. *Nature* 351: 652-654.
- Mihola O, Trachtulec Z, Vlcek C, Schimentii JC, Forejt J. 2009. A mouse speciation gene encodes a meiotic histone H3 methyltransferase. *Science* 323: 376-379.
- Muller HJ. 1922. Variation due to change in the individual genes. *Amer.Natur.* 56: 32-50.
- Muller HJ. 1942. Isolating mechanisms, evolution and temperature. *Biol. Symp.* 6: 71-125.
- Munoz-Fuentes V, Di Rienzo A, Vila C. 2011. *Prdm9*, a major determinant of meiotic recombination hotspots, is not functional in dogs and their wild relatives, wolves and coyotes. *PLoS ONE* 6: e25498.
- Noor MAF. 1997. Genetics of sexual isolation and courtship dysfunction in male hybrids of *Drosophila pseudoobscura* and *D. persimilis*. *Evolution* 51: 809-815.
- Nosil P. 2012. *Ecological Speciation*. Roberts and Co.
- Oliver PL, Goodstadt L, Bayes JJ et al. (10 co-authors) 2009. Accelerated evolution of the *Prdm9* speciation gene across diverse metazoan taxa. *PLoS Genetics* 5:
- Oritz-Barrientos D, Counterman B, Noor MAF. 2007. Gene expression divergence and the origin of hybrid dysfunctions. *Genetica* 175: 1289-1306.
- Ostergren G. 1945. Parasitic nature of extra fragment chromosomes. *Botaniska Notiser* 2: 157-163.
- Paigen K, Petkov P. 2010. Mammalian recombination hot spots: properties, control, and evolution. *Nat. Rev. Genet.* 11: 221-233.
- Perez DE, Wu CI, Johnson NA, Wu ML. 1993. Genetics of reproductive isolation in the *Drosophila simulans* clade: DNA marker-assisted mapping of a hybrid-male sterility gene, *Ods*. *Genetics* 134: 261-275.
- Phandis and Malik, this volume.
- Ponting C. 2011. What are the genomic drivers of the rapid evolution of PRDM9? *Trends Genet.* 27: 165-171.
- Presgraves DC. 2007. Does genetic conflict drive rapid molecular evolution of nuclear transport genes in *Drosophila*? *Bioessays* 29: 386-391.
- Presgraves DC. 2010. The molecular genetic basis of species formation. *Nature Reviews Genetics* 5: 288-298.
- Presgraves DC, Balagopalan L, Abmayr S, Orr HA. 2003. Adaptive evolution drives divergence of a hybrid inviability gene between two species of *Drosophila*. *Nature* 423: 715-719.
- Rose RR, Doolittle, WF. 1983. Molecular mechanisms of speciation. *Science* 220: 157-162.
- Sawamura K. 2012. Chromatin evolution and molecular drive in speciation. *International J. Evol. Biol.* 2012: 301894.
- Schluter D. 2000. *The Ecology of Adaptive Radiations*. Oxford University Press.
- Schluter D. 2009. Evidence for ecological speciation and its alternative. *Science* 323: 737-741.
- Schwenk K, Padilla DK, Bakker GS, Full RJ. 2009. Grand challenges in organismal biology. *Integ. Comp. Biol.* 49: 7-14.
- Segurel L, Leffler EM, Przeworski M. 2011. The case of the fickle fingers: How the PRDM9 zinc finger protein specifies meiotic recombination hotspots in humans. *PLoS Biol.* 9:e1001211.
- Tang S, Presgraves DC. 2009. Evolution of the *Drosophila* nuclear pore complex results in multiple hybrid incompatibilities. *Science* 323: 779-782.
- Ting CT, Tsaur SC, Wu ML, Wu CI. 1998. A rapidly evolving homeobox at the site of a hybrid sterility gene. *Science* 282: 1501-1504.

- Turner E, Jacobson DJ, Taylor JW. 2011. Genetic architecture of a reinforced, postmating, reproductive isolation barrier between *Neurospora* species indicates evolution via natural selection. *PLoS Genetics* 7: e1002204.
- Ubeda F, Wilkins, JF. 2011. The Red Queen theory of recombination. *J. Evol. Biol.* 24: 541-553.
- Vermaak D, Malik HS. 2009. Multiple roles for Heterochromatin Protein 1 genes in *Drosophila*. *Ann. Rev. Genet.* 43: 467-492.
- Via S, West J. 2009. The genetic mosaic suggests a new role for hitchhiking in ecological speciation. *Mol. Ecol.* 17: 4334-4345.
- Wade MJ, Johnson NA. 1994. Reproductive isolation between two species of flour beetles, *Tribolium castaneum* and *T. freemani*: Variation among and within geographical populations of *T. castaneum*. *Heredity* 72: 155-162.
- Willett CS. 2011. Complex deleterious interactions associated with malic enzyme may contribute to reproductive isolation in the copepod, *Tigriopus californicus*. *PLoS ONE* 6: e21177.
- Yasuhara JC, Wakimoto BT. 2006. Oxymoron no more: the expanding world of heterochromatic genes. *Trends Genet.* 22: 330-338.
- Wu M, Kwoh CC, Przytycka TM, Li J, Zheng, J. 2012. Epigenetic functions enriched in transcription factors binding to mouse recombination hotspots. *Prot. Sci.* 10 (Suppl. 1): 511.

Chapter 12

PROCESSES IN ORGANIC AND CULTURAL EVOLUTION

Börje Ekstig*

Department of Education, Uppsala University, Uppsala, Sweden

ABSTRACT

In the present work, I give an all-embracing macroevolutionary perspective on processes of the evolution of life and culture on earth. First, I investigate a complementary form of natural selection that diverges from the traditional form in that it is acting independently of the external environment. This form of natural selection is found as a result of a mathematical analysis of the conditions for population growth. I extend my investigation as well to other evolutionary processes than the organic, such as the evolution of human language and the evolution of science, thereby suggesting other possible forms of underlying explanatory processes.

I examine the concept of complexity and show that it implies new insights into the ways natural selection has been acting in forming the evolutionary and developmental processes. Especially, complexity is found to be growing in an accelerating mode, a process that is explained as the combined result of natural selection and a self-reinforcing feedback process.

The use of the concept of complexity opens the possibility of construction of a new form of a Tree of Life, which, in contrast to traditional forms, combines complexity and time. Such an illustration of the evolutionary process explains the observation that most species live without great changes over vast periods of time. For species at the highest level of complexity there is no competition from species at still higher levels and these species can therefore, if conditions are beneficent, form new species at still higher levels. The process explains the emergence of new species and the general trend of evolution towards cumulatively higher complexity levels.

The cumulative addition of species with successively higher complexity implies that the latest appearing species is the one of the highest level of complexity. At present, this species is the human species.

* Corresponding Author's Email: borje.ekstig@gmail.com.

Keywords: organic evolution, cultural evolution, natural selection, complexity, self-reinforcing feedback, the tree of life

INTRODUCTION

Individual Development and Evolution

Life is fantastic. To me it has always been irresistible to ponder upon the seemingly highly improbable course of events that has resulted in the appearance of mankind on earth and the likewise unfathomably long time during which this very process has incessantly been advancing. Equally enigmatic to me seems the marvellous developmental process of an animal or a human being, merely arising from the molecular genetic code, inherited from the long and unbroken chain of individual developmental courses ever since the dawn of life. In addition, it is to me the highest intellectual challenge to try to understand the intricate coupling between these two processes. Indeed, this is part what this work is about.

All the way during the course of evolution, natural selection is by the majority of researchers considered as the key mechanism. In this selection process, competition between organisms has been a decisive factor. Highly chaotic external factors have formed the underlying conditions for this contest in which natural selection has appointed the winner. The competition may have been about mates and food, as well as about capability to handle social life and the impact of diseases, predators and climate. Thus natural selection gives rise to a population's adaptation to its environment, thereby producing the enduring though irregular evolutionary changes. Due to the irregular nature of these external conditions it has been difficult, as testified by the history of the science of evolution, to find an all-embracing pattern and direction of evolution at large.

It is important to notice that the genetic material, the DNA, doesn't determine the evolutionary process. Instead, the DNA-molecule gives the plan for the individual's growth and its realization in the building up of an embryonic and juvenile individual creature. In the present work, I'm not entering the field of molecular biology – rather I discuss the macroscopic aspects of development and its evolutionary outcome.

A common notion of evolution seems to be to regard it as an extended sequence of adult organisms sustaining successive transformations due to natural selection. However, such a notion encompasses limited truth only. A central tenet forming the basis of the present thesis is that *natural selection primarily is acting on the developmental process* of individual growth and that *the process of evolution above all proceeds as a result of insensibly small and continuous modifications in the development programs* of living organisms, modifications having during the long extent of time given rise to the vast diversity of life on our planet.

This view is by no means new. It is expressed, for instance by Richard Dawkins, in saying that “evolutionary change is change in genetically controlled processes of embryonic development, not literal change from adult form to adult form (Dawkins 2004 B p. 200), and furthermore “that we have to understand development before we can speculate constructively about evolution” (ibid. p. 201). These aspects of development indicate the existence of some kind of relationship between development and evolution; a notion that has caused a long and engaged discussion ever since Darwin's discoveries.

In the current work, I present my own discovery of a concrete manifestation of this relationship and my interpretation of it as well as some far-reaching outcomes of the discovery.

The Problem of Evolutionary Change

In a tradition that harks back to antiquity, living organisms have been seen as arranged in a hierarchical order, thus placing the most simple at the lowest level and the more advanced higher up. This principle was developed without any time dimension, because the world with all its organisms in their present shape was thought to have been created at one and the same occasion. During the end of the eighteenth century, observations of successive temporal changes of organisms became impossible to deny, a notion that got its fulfillment by Charles Darwin in his explication of this successive change, an explication that, after unprecedented turmoil, finally became accepted by the scientific community. As we know, Darwin called his principle natural selection and the process of temporal change became called evolution.

Then however, the idea of hierarchical order came to be considered as obsolete and became abandoned. This is by and large the situation of today's scientific conception of the evolutionary process, even if there are attempts to revitalize the notion of hierarchical order. Thus Daniel McShea (2001) has argued that an increase of the complexity of organisms in evolution can be measured on a hierarchy scale. In the present work, I intend to show the high explanatory possibilities of arranging organisms in a progressively rising complexity scale.

But the situation isn't that easy because most organisms show only marginal changes over long periods of time. How then, as one may ask, to explain this observation by means of natural selection, supposedly being acting on behalf of the great variations in the external conditions that doubtlessly have occurred? And what's more, how to explain the fact that the oldest organisms of today, that is to say those having been exposed to natural selection for the longest time, still retain their archaic characteristics?

Actually, in the beginning of the nineteenth century, Lamarck pondered over this enigmatic observation already when the very idea of evolution was quite new and even before Darwin discovered the principle of natural selection. He asked how we still can see the complete hierarchy of living organisms today, if the active power of nature compels life to mount steadily up the chain of being. Why haven't all living things raised themselves to the same level as man? In his broad exposition of the history of evolution, Peter Bowler (1989 p. 85) has called attention to Lamarck's thought-provoking conundrum. Although much of Lamarck's views now are obsolete, I find his questions still challenging and my intention with the present thesis is to suggest an answer.

The reason for the scientific denial of a hierarchical order of organisms is obviously a lack of means by which to identify and measure such evolutionary levels. Such identification requires a measure of evolutionary progress, a measure that has turned out to be difficult to find.

However, there is a vast literature aiming at a solution of this deficiency. The most frequently suggested candidate for a concept to be used for the assessment of evolutionary progress is complexity. Before a discussion of this concept, I suggest that we prepare ourselves by making acquaintance with a couple of research fields associated to the evolutionary process.

Aim

The aim of the present work is to examine natural selection and its impact on the macroevolutionary process of life and human culture. Special attention is given to my discovery that natural selection in certain situations is acting independently of the external environment. The aim is furthermore to examine the concept of complexity, a concept, as is my intention to show, having great explicatory power for our understanding of the evolutionary process at large. Especially, I analyse the impact of a feedback process in complexity, a mechanism that I suggest as a complement to natural selection. I finally examine to what extension the discussed processes of biological evolution are applicable to human cultural evolution.

Evolutionary Developmental Biology

As I already have emphasized, my main view is to see evolution as the result of changes in the developmental course of individual creatures – by no means a new view. It is given attention in a special theory called Evolutionary developmental biology (*Evo-devo*). In his introduction to *Evo-devo*, Wallace Arthur (2011 p. 28) states that development is both a source for, and a result of, evolutionary change.

Likewise, in his emphasis on the genetic background, Sean Carroll accentuates that the evolution of form occurs through changes in development (Carroll 2005 p. 4) and that molecular changes contribute substantially to organismal adaptation (Carroll 2008). In his summary of *Evo-devo*, Gerd Müller (2007) states that this field explores the relationships between the processes of individual development and phenotypic change during evolution. Müller discusses how developmental systems have evolved and probes the consequences of these changes for organic evolution. He contends that *Evo-devo* aims at explaining how development itself evolves and how the developmental processes are formed by the interplay between genetic, epigenetic and environmental factors.

Unfortunately, the embryonic development of ancestral animals is not available for observations – they are not fossilized – and therefore conclusions regarding evolutionary trends of developmental changes are difficult. It seems that this difficulty to some degree is surmounted by the study of invertebrates. Thus Tills et al. (2011) have performed studies of snails taking advantage of the fact that snail embryos have benefits over those of vertebrates as they allow the study of the continuous developmental process.

Some authors point out that modifications in early embryos are very uncommon because of their supposedly profound effects on subsequent developmental events (Raff and Wray 1989). This view is even more clearly expressed by Brian Hall (2002 p. 13), suggesting that early development is immune to change or, if altered, would so drastically deflect subsequent development that early changes would be lethal or actively selected against. Yet, there is a type of changes that are not impeded by this obstacle.

Thus, Kalinka and Tomancak (2012) contend that many changes in early development result in increase in the tempo of embryogenesis and a shift from late to early-patterning processes; a view studied in the field of heterochrony. These remarks are of high significance for the continued analysis in this work.

Heterochrony

As emphasized in the field of Evo-devo, modifications of the developmental programs of growing creatures provide an important entrance to the understanding of evolution. A special kind of such modifications is temporal displacements of the formation of developmental traits, a phenomenon called *heterochrony*. Central contributions to this field of research are given by Stephen Jay Gould's (1977) classical *Ontogeny and Phylogeny*, by McKinney and McNamara's (1991) *Heterochrony* and by Rudolf Raff's (1996) *The Shape of Life*.

In his summary of heterochrony, Kenneth McNamara (2002, 2012) defines this process as a change of timing or rate of development relative to the ancestor. He states that heterochrony takes the form of both increased and decreased degrees of development. He also recognizes that genes regulating embryonic or larval development play a major role in evolution. These genes are the targets of mutations causing heterochronic change in phylogeny. A link between evolution and development is formed by the inherent directionality in evolutionary trends also occurring in organisms' development. McNamara concludes that also human features are moulded by the all-pervasive influence of heterochrony. Our overall large body size, relatively large legs, structure of our pelvis, enlarged cranium and brain, and even our big feet can be seen as resulting from heterochronic processes. As Raff (1996 p. 267) points out, studies of heterochronies have predominance for late development, but he stresses that heterochronies in early development have been found to be as prevalent as in late development. I would here like to remark that all heterochronic changes are more or less tacitly assumed to be formed by natural selection.

In the present work, I restrict the discussion to two simple forms of heterochronies – condensation and terminal addition. *Condensation* means the successive shortening of the interval of time needed for the shaping of each specific trait in the developmental process of individual creatures, a process causing displacement to earlier emergence of subsequent traits, a process called *acceleration*. Thus acceleration is caused by condensation and I therefore focus the discussion on condensation.

It should be emphasized that condensation, like other heterochronic processes, is a process expressed both in ontogeny and phylogeny, thus forming an issue for Evo-devo. One representative of this field of research maintains that when comparisons are made between very different levels of complexity, the emerging pattern is broadly recapitulatory, although only in a very imprecise way, in the sense of recapitulating levels of complexity rather than precise morphological details (Arthur 2002). It is interesting to note that Arthur includes complexity in his discussion, a concept that is central in the present thesis as well and will be analysed in forthcoming sections.

Terminal addition means the emergence of new traits with new functions at the end of the maturation period. One may think that terminal addition cannot be included in a strict definition of heterochrony since it means no change of timing or rate of development of existing early organs. Nonetheless, for instance McKinney and McNamara (1991) include it in their abundant terminology. They argue that much of the overall pattern of evolution can be explained by terminal addition and that one particularly important arrow of evolution, that of increasing complexity, seems to owe much of its existence to this process (ibid. p. 381).

Terminal addition occurs near the stage of maturity and is explained by the traditional interpretation of natural selection. An addition of a new trait is of course possible only on behalf of its reproductive advantage in the prevalent environment. Otherwise it wouldn't be inserted.

Terminal addition causes prolonged generation time owing to the insertion of novel traits at the end of the juvenile period. Such novel traits can lead to the emergence of new species and in this way terminal addition explains the diversity of life (*ibid.* p. 381). It must in this context be mentioned that after the stage of maturation great changes are rare because after reproduction, the fate of the adult animal has no bearing on the genes of its offspring. Of course, this rule is not absolute in the case of the adult's caring of its offspring. Furthermore, it isn't abrupt for iteroparous species.

The two processes of condensation and terminal addition result in opposite evolutionary trends; condensation results in shortening of the time from conception to adulthood, i.e., a shortening of generation time; terminal addition on the other hand prolongs maturity time, i.e., it increases generation time. A common notion, as Stephen Jay Gould (2002 p. 367) mentions, seems to have been that new stages cannot be added indefinitely to the end of previous ontogenies, lest growth of adulthood take untold years to reach completion. Some process must therefore, as it is thought, produce general speeding-up of development, leaving time at the end for novel features. Such a view, I think, implies evolution to embrace an element of intentionality, a view that must be resolutely rejected. An important goal of the present work is to explain the speeding-up of development in a scientific way.

As to heterochrony in general, it maybe isn't that all-pervasive as McKinney and McNamara contend. Thus Rice (1997) proposes that although heterochrony has become a central organizing concept relating development and evolution to each other, it can only accurately describe a small subset of the possible ways through which ontogeny can change. The somewhat bewildering nomenclature of heterochrony is meaningful only when there is a uniform change in the rate or timing of the ontogenetic process, with no change in the internal structure. In the present thesis, such a uniform temporal change without change in the internal structure forms a basic principle.

As mentioned above, there is a generally expressed contention that changes of traits in early development would so drastically deflect subsequent development that such changes might be lethal. But a change restricted merely to an increase in the tempo of the shaping of a trait, without change of internal structure, wouldn't change its function and therefore wouldn't be prohibited.

This is what characterizes the process of condensation. But there is another important consequence of this conclusion. If there is no change of function, natural selection has no alternatives among which to choose. How then to explain condensation? It seems that we have to examine the very process of natural selection somewhat closer.

ENVIRONMENT-INDEPENDENT NATURAL SELECTION

Population Biology

In his study of population biology, Stephen Stearns (1992) points out that a shortening of generation time is beneficial for the population growth because it increases its chances of surviving the often-risky juvenile period. A higher number of individual creatures will reach maturation and get the possibility to contribute to reproduction. In addition, a shorter generation time implies more frequent occasions of reproduction over time.

Therefore, these double advantages generate a selection pressure on all developmental traits to develop as rapidly as possible with retained functionality. Such a selection may give rise to a shortening of developmental traits once they have been established. We recognize this shortening as condensation.

There is simultaneously also a benefit of the opposite process – the prolongation of generation time. A delayed maturity, as Stearns points out, may imply greater size and a higher quality of the offspring, which leads to higher initial fecundity. This process, I maintain, is only occurring at the end of maturity time as terminal addition, since great changes of early developmental traits often are deleterious. Such terminal additions are explained by the traditional interpretation of natural selection.

Thus, there are two counteracting processes influencing generation length; one causing its shortening, associated to condensation of developmental traits, the other causing its prolongation, associated to terminal addition. An important contention of the present analysis is that these two processes are acting independently of each other. Hence, even if a species prolongs its generation time by means of terminal addition, its developmental traits may nonetheless be subject to condensation, implying that the actual change of generation time is the total result of both processes. In the majority of cases, the prolongation dominates but an assessment of the contribution of each of the processes seems difficult.

As we may conclude so far, there is a selection pressure on all developmental traits to develop as rapid as possible. An important question is now if the traditional form of natural selection is sufficient for explaining such a shortening because, as I already have pointed out, if there is no change of function, natural selection gets no alternatives among which to choose. One may thus conclude that condensation is invisible to natural selection, at least in its traditional interpretation. There is need for a more detailed analysis.

A Mathematical Analysis

The question is if the traditional form of natural selection is sufficient to explain condensation of developmental traits. I suggest an investigation of this question by means of a simple mathematical analysis. In this investigation, I examine the combined impact of reproduction, survival, and generation length for the growth of a population. Reproduction is associated to the ability to attract mates, to the size of the litter, and to the feeding and defending of the offspring, and is primarily of importance at the transition from one generation to the next. Survival is associated to the ability to catch prey and/or to avoid predators, to have good digestion, to have sensitive senses, to behave conformingly in the current social framework, and is primarily of importance during the lifetime of individual creatures. I examine by means of a mathematical analysis the impact of reproduction and survival on the growth of population size for different generation lengths.

Suppose that we have a population of which the fraction s survives every year. Then, if the generation time is g , the fraction of the initial number of the population at the end of a generation is s^g . We next suppose that the level of reproduction at the end of each generation will increase or decrease the size of the population by the factor r .

At the onset of a new generation the population size has thereby, due to the two factors r and s , changed to rs^g .

The number of such changes after time t is the number of generations during this time, which is t/g . Then the total number of individuals N in the population after time t will be

$$N/N_0 = (rs^g)^{t/g} \quad (1)$$

Where N_0 denotes the size of the population at the beginning of the studied period. We examine the impact on population size for different values of the three variables r , s , and g . Calculations are performed by means of numerical examples for four different cases.

Case 1. $rs^g > 1$; The population increases.

Case 2. $rs^g = 1$; The population is stable.

Case 3. $1 < r < s^{-g}$; The population decreases.

Case 4. $r < 1$; The population decreases.

I calculate the growth of the population N/N_0 by means of formula (1). The values of the parameters r and s are chosen for clarifying illustrations of three of the four cases (Case 2 is trivial since the size of the population is constant). In each case the growth of the population over generations is calculated for two different generation lengths and the result is illustrated in Figure 1.

This result is understood in the following way. The annual survival factor s , since always less than one, reduces the population, a reduction that is compensated for at the occasions of reproduction with the factor r at the end of generation time g .

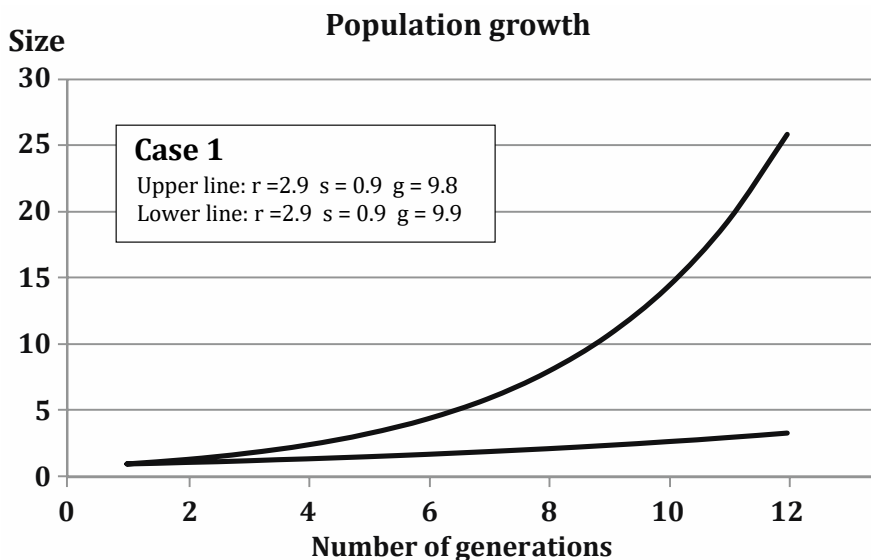


Figure 1. (Continued).

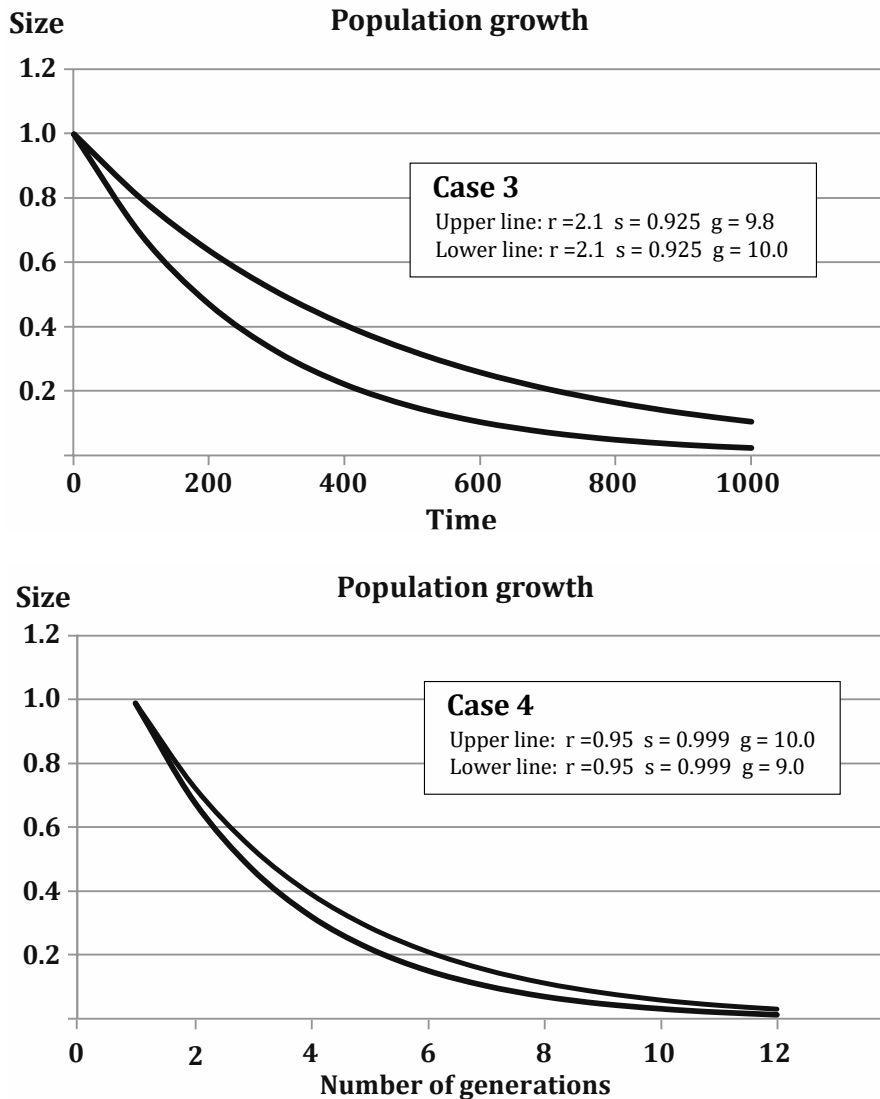


Figure 1. The graphs of Equation 1 shown for some illustrative values of the variables r , s , and g .

In Case 1, the graphs show that the size of the population will grow faster for a shorter generation time. In this case the environment is beneficial for population growth implying that the total reduction at the end of each generation time is exceeded by the reproduction and if generation time is shortened, the number of annual reductions falls and the number of occasion of reproduction increases, both factors of which contribute to more rapid growth of the population size over time. This means that a shortening of generation time is beneficial for the growth of the population and hence there is a selection for shortening of generation time.

In Case 2, the reproduction constant at the end of generation time balances the annual reductions and the population is stable. In this case shortening of generation time is neither beneficial nor deleterious for the population. Yet, in this case the three parameters must have very special values for satisfying these conditions, which seems quite improbable, especially

as the features are stretching over time. Therefore, due to the natural variation of r and s , the conditions will very soon change to either Case 1 or Case 3.

In Case 3, the graphs of the formula show that the size of the population will decrease slower for a shorter generation time. In this case the reproduction at the end of generation time cannot fully compensate for the sum of annual reductions and the population size will decrease over time. But if generation time is shorter, the number of the annual reductions falls and the occasions of reproduction will be more numerous. Since the reproduction factor is bigger than 1, both these circumstances result in slower decrease of population size as testified by the graphs. Therefore, even under these harsh circumstances, a shortening of generation time is beneficial. This is an especially important implication of the mathematical model that, I think, hardly could have been intuitively anticipated. It clarifies the general selection pressure for shortening of generation time even in a deleterious environment. Together with Case 1 one can conclude that there is *a selection pressure for shortening of generation time, and therefore also for condensation, whether the environment is beneficial or deleterious*. Therefore the traditional interpretation of natural selection as being coupled to the external environment seems insufficient in explaining condensation and that there is an environment-independent form of natural selection responsible for the condensation of developmental traits.

If extended over a long time, this situation of course leads to extinction, a disaster however being postponed by the slower decrease of population size. In this way the population raises its chances of survival until less perilous external conditions appear or until it by means of ordinary natural selection has got time to adapt to the hostile environment. Because extinction is a far from infrequent occurrence in the history of evolution, the proposed process of shortening of generation time, I think, isn't without importance since one may think that a species despite hard external conditions in such critical periods may have avoided extinction by means of this process.

One should here remember that in the above discussion of condensation we concluded that condensation is invisible to natural selection, at least in its traditional interpretation because, since condensation implies no change of function of a trait, natural selection has no alternatives to choose amongst. This conclusion is, as we can see, concurrent with the result of the calculation, which shows that condensation actually can be independent of the environment. Still, I think one shouldn't abandon the process of natural selection when it is about condensation, just that it is of a form that is environment-independent. Because of this independence of the haphazardly varying external conditions it seems possible, maybe even plausible, that condensation could exhibit some degree of regularity.

In Case 4, the graphs of the formula show that the size of the population will decrease faster for shorter generation time.

In this case the low value of the reproduction constant adds to the decrease coming about by the small survival factor, implying that the population size will suffer a successive decrease. But in this case, as the analysis shows, the population decreases faster in case of shorter generation time. Therefore there is a disadvantage of a shortened generation time. The situation leads to extinction if not interrupted by a change to a higher value of the reproduction constant, thus bringing the population to the conditions of some of the former cases. Therefore species having in fact survived over time cannot have been applying the conditions of the fourth case except for short periods of time and therefore this case doesn't contradict the general selection advantage of shortened generation time for normal, long living species.

The main result of the mathematical analysis shows that, when a population is increasing, there is a benefit of shortened generation time. However, even if the population is decreasing as in situations when the external environment is moderately deleterious, there is a reproductive advantage of shortened generation time as well.

Hence, as the result of the theoretical analysis shows, the selection for shortened generation time is independent of the external conditions. This means, I claim, *that there is a form of natural selection that is independent of the external environment.*

One may perhaps dare to say that the tendency for shortened generation time is an inevitable characteristic of living creatures. I will in a coming section demonstrate an empirical support of this conclusion. As we see, this result is at odds with the traditional interpretation of natural selection according to which evolutionary changes are seen as adaptations to the external environment.

However, in the case of the selection pressure for shortened generation time, I cannot identify any possible feature in the environment to which condensation could be an adaptation. Therefore, natural selection in the adaptationist interpretation isn't applicable. Instead, as I have demonstrated, condensation is explained as a result of a form of natural selection that is independent of the external environment. This conclusion is concurrent with the conclusion as posed above, implying that condensation of developmental traits occurs without change of their function, and that therefore natural selection has no alternatives among which to choose.

Richard Dawkins (2004 B, p. 88) mentions the notion of evolutionary change *per se*, and I suggest one may use this idea and state that environment-independent natural selection is a process that, as opposed to adaptive change, might be characterized as evolutionary change *per se*. I think that this general notion can be meaningfully used in a comprehensive description of evolution.

An Application to Bacteria

The result of the analysis in this section is primarily of theoretical interest though I can think of an application regarding the growth of bacterial colonies. In a high-nutrient environment a population of bacteria increases and there is a selection for shorter generation time according to Case 1. In a situation of depleted nutrients or exposition to antibiotics, the population may decrease. At the application of antibiotics to pathogenic bacteria, the result of Case 3 is especially interesting, implying as it does a selection pressure for shorter generation time and a postponement of eradication under the conditions of moderate population decrease. The bacteria then have a somewhat prolonged time for adaptation to the antibiotics thus getting an opportunity of developing resistance. It is therefore important that the antibiotic treatment is given as a high dose therapy so that the decrease of the bacterial colony is fast enough to enter Case 4. The vital importance of a complete extinction of pathogenic bacterial colonies is already well understood though I think the present result gives additional emphasis to the importance of high antibiotic therapy.

As I already have pointed out, the tenet of the present work is that evolution proceeds as a result of the continuous changes in developmental programs of living organisms. After the above analysis of the developmental process, I continue by examining the coupling of the developmental process to that of evolution.

DEVELOPMENT AND EVOLUTION

The Relationship between Development and Evolution

There has since long been noticed that vestiges of old evolutionary stages can be observed in the embryogenesis of present-day individual organisms, observations that have been subject to enduring discussions related to the idea of recapitulation (for a comprehensive survey, see Richards 1992). These discussions were intense in the second part of the nineteenth century, but since they seem not to have given any satisfactory conclusion they have faded out. In the present work I revitalize the principle of the concept but avoid the term recapitulation because of its bad reputation.

I have been criticised for dealing with this connection, which by most biologists is considered an obsolete and abandoned idea. I must therefore try to make myself very clear, especially, as it's my conviction, that the connection gives a possible key to a deeper understanding of the evolutionary process. In addition, the connection constitutes the main empirical support of present work.

The interpretations of the observed vestiges have mostly been restricted to morphological and anatomical features within the field of biology. When I also included cultural features of the human species in the analysis, I discovered a conspicuous pattern, indicating a relationship between the developmental process of a modern individual human being and its evolutionary history. This pattern is depicted in Figure 2.

The pattern as revealed in the diagram of Figure 2 makes the notion of a relationship between development and evolution more concrete. As we can see, the pattern to a large extent comprises human cultural traits as well. Otherwise, the regularity would never have been disclosed and this, I think, is why biologists didn't discover it.

As we can see, the diagram reveals a concrete connection between development and evolution or, as used to be said, between ontogeny and phylogeny. Darwin himself observed and discussed this relationship as we can see in the following quotation:

I shall attempt to show that the adult differs from its embryo, owing to variations supervening as a not early age, and being inherited at a corresponding age. The process, whilst it leaves the embryo almost unaltered, continually adds, in the course of successive generations, more and more difference to the adult. Thus the embryo comes to be left as a sort of picture, preserved by nature, of the former and less modified condition of the species. This view may be true, and yet may never be capable of proof (Darwin 1859 p. 338).

The observation of such a relationship has been subject of a vigorous though bewildering discussion during the later half of the nineteenth century. Ernst Haeckel, known as a forceful supporter of Darwin's theory, stated the phrase "ontogeny recapitulates phylogeny," also known as the biogenetic law.

One of Darwin's great contributions, besides that of natural selection, is the concept of common ancestry. In one of his notebooks from 1837, i.e., one year before his discovery of natural selection, one can find a rough drawing showing this principle (Darwin 1837). It's interesting to note that it was probably not a direct observation of the phenomenon of evolution that gave him this brilliant idea but rather the study of embryos.

As Carl Zimmer (2002 pp. 316-317) points out, Darwin acquainted himself with studies of embryos of gorillas and chimpanzees and saw conspicuous similarities. Bone for bone, they are almost identical. As human embryos develop, they pass through virtually identical stages as gorillas or chimps. Only relatively late in their development do they start to diverge, taking on different proportions. These similarities, Darwin argued, were signs that apes and humans descended from some ancient common ancestor.

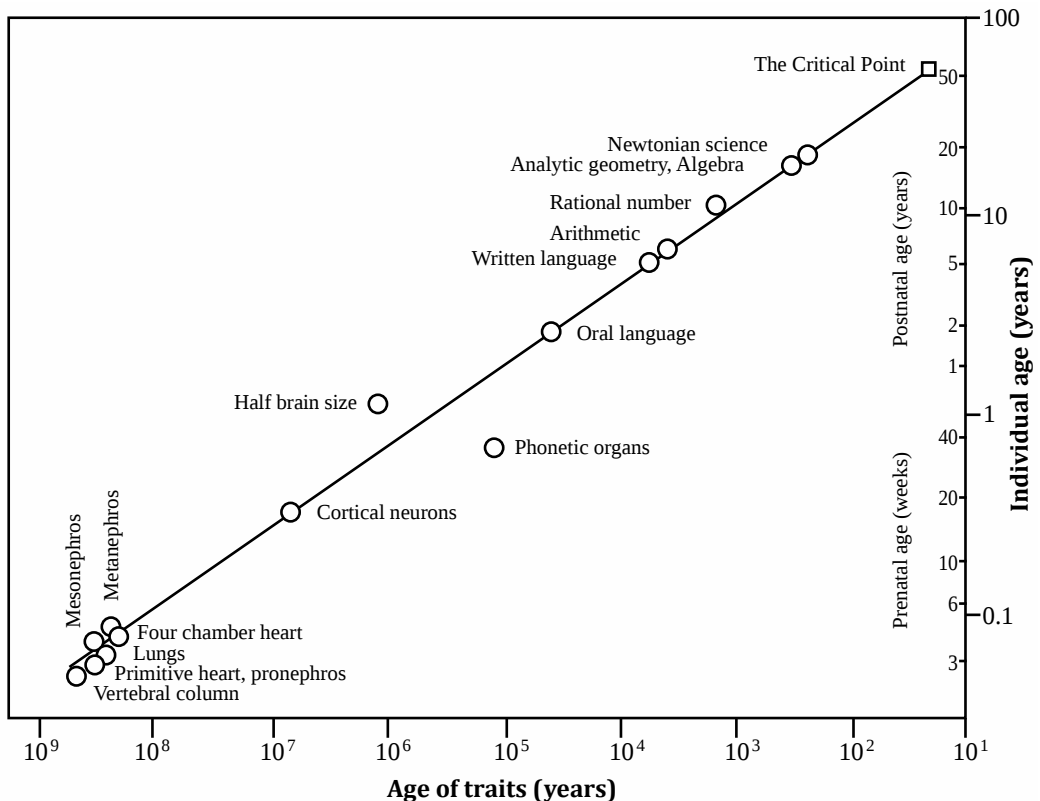


Figure 2. Individual development versus evolution of the lineage of man. The horizontal axis gives evolutionary time measured backwards from the present point of time. The vertical axis shows the individual age. Both scales are logarithmic. References to and discussion of the traits that are included in the diagram are given in my previous publication of the pattern (Ekstig 1994). This version of the diagram was published and further discussed in Ekstig (2007, 2010 A, 2015).

In looking at Darwin's principle of common ancestry from the perspective of developmental programmes, one may conclude that species having deviated from a common lineage share the same developmental course up to the stage corresponding to their divergence, a principle called *Spencer's theorem*. McKinney and McNamara (1991 p. 390) express this observation in saying that the more closely related two groups are, the more similar their ontogenies, and the later their ontogenies diverge.

Let me also refer to a remark on this issue made by Dawkins who reminds us that organisms, including human beings, form overlapping sequences of generations in which each generation exhibits a close resemblance to its progenitors. He then suggests an expressive metaphor: "Everything about a modern animal, especially its DNA, but its limbs and its heart,

its brain and its breeding cycle too, can be regarded as an archive, a chronicle of its past, even if that chronicle is a palimpsest, many times overwritten.” (Dawkins 2004 A p. 23) (A palimpsest is a manuscript page from a scroll or an old book that has been overwritten and used again.) In this metaphor, the DNA has a special weight because, as we know, the DNA is copied over and over again, with only small modifications, over many generations.

What makes the evolutionary process so complicated is a double causality. Evolution proceeds as a result of modifications in the development programs of living organisms and the development course recapitulates traits that are consecutively added during the evolutionary history. This mutual causality has come into being at the construction of the straight line in Figure 2.

When considering the diagram in Figure 2, I would like to emphasize that the traits forming the pattern are the same, whether appearing in an ancient ancestor or in a modern organism. Thus for instance, the three forms of kidneys, pronephros, mesonephros, and metanephros, are identical, at least as seen to their function, in the ancient reptile and in a modern human being. Of course, we cannot observe the ancient reptile but, as is generally observed, the embryonic organs don’t change much over evolutionary time, and we may therefore conclude that in today’s reptiles, the kidneys are the same as they were in the ancient reptiles. This is supposed to be valid for all traits in the diagram. By means of this reasoning, I conclude that the similarities giving rise to the biogenetic law can scarcely be said to be similarities at all because instead, with respect to their function, they are identical. Consequently, there is one and only one point for each trait in the diagram, though given two separate dates for its appearance.

What really is enigmatic, though, is the existence of such a regular pattern over such a great part of the evolutionary process, actually over all the time of animal evolution. Of course, this regularity insists upon an explication, especially as the process of evolution is supposed to be driven by natural selection depending on the highly irregular external contingencies that certainly have been prevailing all over this long time. Step by step, I will develop an explication of this conundrum.

The straight line invites to two ways of interpretation. The first concerns condensation and the second concerns complexity.

We first discuss condensation, the assessment of which is obtained by means of a mathematical analysis of the line.

The Regularity of Condensation

The straight line of course invites to a mathematical analysis. The equation of line is

$$\ln t_o = C_1 - C_2 \ln t_p \quad (2)$$

where t_o and t_p denote the developmental (ontogenetic) and evolutionary (phylogenetic) age of each trait, respectively. The values of dimensionless constants are determined by the scales on the axes yielding $C_1 = 5.12$ and $C_2 = 0.39$.

Due to the fact that the evolutionary timescale in the diagram of Figure 2 starts at the present point of time and displays time in the negative direction, the position of each particular event will be displaced to the left as time proceeds and, because of the fact that the time scale

is logarithmic, this displacement per unite time as seen in the diagram will appear smaller the higher the age of the trait.

Then the line wouldn't be linear except for the present point of time, which is a highly unsatisfactory feature. Instead it is reasonable to assume that the line reflects a general characteristic of life and not just an accidental coincidence of the present point of time, because, of course, our understanding of the evolutionary process cannot rely on the presumption that the present point of time has a special status in the evolutionary process.

However, there is a possibility that the linearity of the pattern is preserved over time if one assumes that the points are displaced downwards towards earlier developmental appearance simultaneously as they are moving to the left due to the increasing age of each trait with time. In accepting this, as I claim, highly reasonable assumption, one gets not only a means to explain the straight line, but a possibility of a quantitative measure of acceleration and condensation as well. If the displacement downwards, recognized as acceleration, occurs at a pace that is determined by the value of the derivative $a = dt_o/dt_p$ of Equation (2), the linearity of the straight line will be preserved over time. Differentiation of Equation (2) yields

$$a = -C_2 t_o / t_p \quad (3)$$

The quantity a thus denotes the acceleration of each trait appearing at the ontogenetic age t_o .

Such a regular displacement of developmental traits is the result of an appropriate shortening of all preceding stages, a shortening recognized as condensation. The quantity of condensation, denoted q , is the relative difference of two values of acceleration at adjacent points on the lines. This difference is the derivative of acceleration with respect to t_o . Hence $q = da/dt_o$, and differentiation of Equation (3) yields

$$q = -(C_2 + 1)/t_p \quad (4)$$

This analysis shows that the straight line in the diagram is coupled to a continuous and regular shortening of developmental stages, i.e., condensation. The formula implies a systematic decrease of condensation over time, which is intuitively sensible since the more a stage is shortened, the more difficult it must be to shorten it still more. However, I find it remarkable that condensation obeys such a simple formula, depending as it does on one variable exclusively – the age of the trait.

Maybe, though, that the simple regularity of condensation, and hence of the presence of the straight line, isn't so remarkable after all, because, as we found in a previous section, condensation is a consequence of a selection process that is independent of the irregular external contingencies.

As we can see, the formula implies the possibility of a quantitative assessment of condensation and a couple of examples may clarify the significance of the formula. Since the traits of the human embryo emerged very early on the evolutionary time scale, in fact several hundred million years ago, the condensation of these traits is quite small, actually of the order of 3% over 10 million years. As a second example, we may choose the acquisition of oral language of a modern child, which, according to the model, is condensed 3% per millennium. Such small values are of no practical significance. The values of condensation of traits

appearing at higher age are bigger. Thus, for a modern child, the acquisition of Newtonian science is found to be 5% per decade. In the present work, though, there is no need to use these specific values of condensation, because the suggested model of evolution is merely outlined in a most general form.

It may seem, though, that condensation has little influence on evolution at large. I can imagine a process of evolution without condensation, proceeding in much the same way that it actually has gone through. Generation times would just be longer. However, one must assume a random variation of generation times. Then, as discussed in the above section of population biology, a population with somewhat shorter generation time would reproduce more effectively and win the competition with populations with longer generation times. This is how natural selection works, in this case actually in a form that is independent of external conditions. I thus conclude that condensation is an inseparable part of the evolutionary process.

Condensation and Terminal Addition

It is important to clear out the how of the counteracting processes of condensation and terminal addition are working. I therefore try to make it easy to grasp with a highly idealized illustration, intended to facilitate the understanding of the concepts. At the top these columns, sexual maturation sets in, marking the end of ontogeny. The figure illustrates condensation of early appearing organs during the course of evolution. Due to condensation of preceding organs, subsequent organs are displaced towards earlier appearance, a process recognized as acceleration.

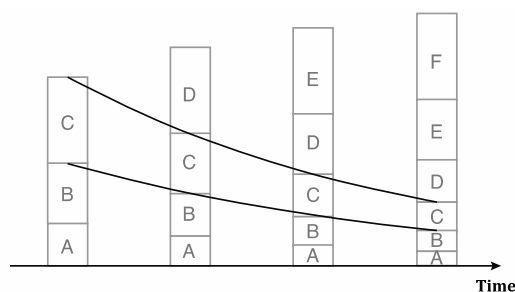


Figure 3. The vertical columns illustrate four life histories at different points of time in the evolutionary history of a species at occasions when terminal additions have occurred. The blocks in the columns represent developmental traits. We may think block C as representing a primitive back-bone, block D a primitive heart, E the lungs, and F the modern heart. It may also be suggested that the blocks represent stages in the human cultural evolution. Thus block C may represent oral language, D written language, E arithmetic and F algebra. Between these columns one may think of a lot of intermediate life histories in which there has been a continuous condensation and shortening of generation length. The curved lines illustrate acceleration and the distance between them at each point of time corresponds to condensation.

The figure also illustrates how new organs are added at the end of ontogeny, a process identified as terminal addition. As we can see in this picture, the lengthening of generation time and the often-accompanying growth of body size, being normal trends in the great perspective of evolution, don't contradict the conclusion of a simultaneous presence of condensation. Such a lengthening might however be somewhat reduced by condensation. In this way, maturation

time escapes from becoming awkwardly long, but this conclusion is a consequence, not a cause, of condensation. The fact that condensation and terminal addition give a superimposed result is probably responsible for the fact that the occurrence of condensation has not been given appropriate attention in prevailing discussions of evolutionary processes.

Terminal additions (including increasing body sizes) must in most cases have been more advantageous than the disadvantages due to the consequential prolongation of generation time, because otherwise the new traits could never have come about. This means in these situations that ordinary environment-dependent natural selection in the formation of new traits is stronger than the accompanying selection for shortening of generation time. Nonetheless, it seems reasonable to assume that, during the process of evolution, the counteracting processes of condensation and terminal additions have been in action simultaneously and independently of one other.

The sequence of stages at the top of the columns illustrates what is normally considered as the evolutionary history of the species. Of course, it is unavoidable to observe that this sequence of traits is observed in the sequence of traits building each individual's life history as well. This is recognized as the contentious concept of recapitulation that I have mentioned above.

An Explication of the Regular Temporal Structure of Evolution

The straight line in Figure 2 indicates a large structure of evolution that insists on explanation. In Figure 3, the four columns illustrate in an idealized way four individual developmental courses at the occasions of terminal additions. Such major transitions are certainly occurring at highly different time intervals because they are molded by natural selection and thus dependent of the haphazardly varying external conditions. Between these occurrences one may envisage a great number of individual developmental courses (not illustrated in the diagram), incessantly exposed to selection pressure for condensation. In my analysis of population biology, I found condensation to be independent of external contingencies. Therefore it seems possible, maybe even plausible, that it could exhibit some degree of regularity. If so, the lines of acceleration in Figure 3, i.e., the lines connecting the onsets of developmental traits would be even as indicated in the figure. As a consequence, the irregular time intervals at which terminal additions have occurred in the course of evolution would be preserved as corresponding time intervals at which traits are added in subsequent individual developmental courses. This structural similarity could then be illustrated as a straight line as that in Figure 2. Since this line, formed on the basis of empirical data, in fact is found to be straight, the assumption of the regularity of condensation is confirmed. One may thus conclude that in spite of highly irregular conditions, the evolutionary process has nevertheless, under the influence of the environment-independent functioning of condensation, accomplished a large-scale regularity in its outcome as manifested as a straight line in Figure 2.

A different way to arrive at this conclusion is to assume that if condensation hadn't been independent of the environment, it would probably not have been regular and if so, the temporal structure of the evolutionary process wouldn't have been preserved in the developmental process. Then the line wouldn't have been straight.

Let us, in order to make this reasoning easier to follow, consider an example related to the features of the diagram of Figure 2. In this diagram we can observe that the temporal gap

between the evolutionary occurrences of the primitive heart and the four-chamber heart is much smaller than that between, say, the appearance of cortical neurons and oral language. The same difference is observed in the timing of these traits in the individual growth process as seen on the vertical axis. This demonstrates that the temporal structure of the evolutionary process is preserved in the developmental process, thus giving additional support to the above reasoning. Of course, one should not forget that the mathematical analysis of the straight line has shown that the regularity of condensation is a consequence of the linearity of the line, but the task of the reasoning has actually been to explain the feature of the line of being straight.

The main conclusion is that, *in spite of the irregular time intervals at which changes have occurred in the course of evolution, there is a large-scale regular structure of these changes, revealed when also the developmental process is taken into account.*

This regular structure is a result of a process being coupled to a form of natural selection that is independent of the external environment.

Empirical Support

An important question is of course if the discussed selection pressure for condensation in fact has been exerted in evolutionary lineages of living creatures and if so, if such a process can empirically be studied. A fundamental difficulty hampering this question is that the embryonic development of ancestral organisms is not available for studies and therefore that longitudinal study of condensation in a lineage cannot be performed.

This difficulty restrains all attempts of examinations of temporal displacements as studied in the field of heterochrony. For the present analysis, studies of the age of an embryo are the most urgent. However, as Balinsky (1975 p. 302) points out, the age of an embryo is often not known, and in animals other than mammals, the rate of development is dependent on the temperature of the environment making an assessment of the age of the embryo especially uncertain. Moreover, the size of the embryo is no true indicator of its degree of development, as the dimensions of the embryo vary to a great extent. As to the morphological peculiarities of the embryo, there are certain limitations even to this approach. Finally, as he indicates, the development of different parts or organs of the embryo is not always strictly coordinated in time. In this situation, I think, we are restricted to look at the largest patterns of the evolutionary process. Thus, from textbooks on embryology like the one by Balinsky, we may contend that in present-day vertebrates, the most vital somatic organs (the heart, the kidneys, and the like) are moulded very early in the gestation period.

One must remember in this context that new traits in their first emergence appear as terminal additions because, as is generally assumed, novel traits appearing early in the embryogenesis are most probably affecting subsequent forms in a deleterious way.

Then, after this occurrence, condensation of this and preceding traits are causing the observed early appearance of basic embryonic and juvenile traits.

Since, as we have seen, developmental traits are found to have a regular temporal dependence of the intervals at which evolutionary events have occurred, it would be interesting to find out if the very process of evolution would display some degree of regularity as well. In order to find such regularity, one needs some sort of measure of evolution itself. This leads us to the second way of interpretation of the straight line in the diagram of Figure 2 – an interpretation in terms of complexity.

COMPLEXITY

Common Views

In the extensive literature on evolution, many authors seem to be hesitating about how to determine the direction of the evolutionary process; about how to characterize, as they say, the arrow of time. A common though intuitive notion is that the direction can be characterized as a steady increase of complexity.

In the present work I maintain that the concept of complexity provides a promising way to get a deeper understanding of the very process of evolution. A recent contribution to the literature on complexity in connection to evolution is given by Lineweaver et al. (2013), having collected contributions from several researchers in the field.

A common problem, expressed by many of these authors, is the lack of a definition of complexity. But, as the editors ask, even without a definition or a way to measure it, isn't it qualitatively obvious that biological complexity has increased? Do we really need to wait for a precise definition to think about complexity and its limits? (*ibid.* p. 5). I find this remark highly relevant for my present analysis.

Melanie Mitchel (2009 pp. 96-111) describes many forms of complexity in specific fields such as complexity of size, entropy, information, thermodynamic depth, fractals, and degree of hierarchy. She contends that complexity is discussed in several fields of research such as in genetics, in Evo-Devo, and in studies of chaos.

As the history of science shows, the comprehension of a new concept often starts with its measurement. This practice can be seen as exemplified by Newton's measurement of force and Joule's measurement of energy. Actually, there are some fundamental concepts that cannot be defined by reference to still more fundamental concepts. I think that complexity may be such a concept. If so, it is futile to wait for a definition and instead it seems better to develop methods of its measurement and explore the explanatory consequences of such a measurement. I have in a previous publication (Ekstig 2010 A) suggested such a method that shortly will be presented in the next section.

Even if one cannot formulate a stringent definition, there is need of a description in ordinary terms of what complexity means. As an attempt at such a description, I would here like to use a simple formulation by McShea and Brandon (2010), implying that complexity means the number of parts or the amount of differentiation among parts within an individual.

Most authors express the intuitive notion that complexity has been increasing during the course of evolution. Since natural selection is generally seen as furnishing the prime mechanism of evolution, it is more or less tacitly assumed that increasing complexity has come about by the action of natural selection.

But there are also attempts to explain increasing complexity even without the action of natural selection (McShea and Brandon 2010, Kauffman 1993, 2013). An objection against these attempts is raised by Lineweaver (2013 p. 7), maintaining that a definition of complexity without the involvement of natural selection merely describes an increase in entropy. However, many authors resolutely deny the possibility of a meaningful use of the concept of complexity in the context of evolution altogether, referring as they do, to the lack of a stringent definition. It seems in this context that Robert Trivers's opinion has had a pervasive influence through his statement: "There exists no objective basis on which to elevate one species above another.

Chimp and human, lizard and fungus, we have all evolved over some three billion years by the process known as natural selection” (Trivers 1976). In a certain meaning, all species can of course be said to be emanating from the very first forms of living organisms, since all have parts of their DNA conserved from these old organisms. I think that that is what Richard Dawkins has in mind when stating that all lineages have had exactly equal time to evolve since the dawn of life and he ardently discards the idea that animals could be ranked on a “higher or lower” scale (Dawkins 1992 p. 263). He is somewhat more open, though, to the concept of progress, especially as a result of arms race (Dawkins 2004 A pp. 496-497).

However, Dawkins doesn’t think that anybody would deny that there has been a broad overall trend towards increased information content during the course of human evolution from our remote bacterial ancestors (Dawkins 2004 B p. 101) and then, in the same context, he refers to another scientist in suggesting that information content of a biological system is another name for its complexity (ibid p. 102), a contention that, at least, he doesn’t deny. It seems to me that Dawkins here comes quite near the notion of complexity as a quantity that is increasing in evolution.

I think that Trivers’s and Dawkins’s opinions in this matter can be understood insofar as when expressing a notion of a change of an entity, one has to rely on some kind of quantitative assessment, a measure that, as commonly observed, is lacking in the case of complexity. Yet despite this problem, there is a widespread opinion, even as we have seen amongst scientists, to hold on to the view of a generally increasing complexity in the evolutionary process, a view that is maintained with or without the involvement of natural selection.

In the present work, I strongly adhere to the view, expressed by so many scientists, that complexity has successively been increasing during the evolutionary process and that this process mainly is a result of natural selection. However, the view that increasing complexity is driven by natural selection requires a more detailed analysis that I will come back to. But, first of all, can complexity really be measured?

The Assessment of Complexity

As a point of departure, I make the assumption that there is such a concept as complexity that has been continually increasing during the course of evolution. The intuitive view of such a process relies on how evolutionary key novelties have been appearing over time. Such an indication is discussed by Carl Sagan in the form of a picture he calls *The Cosmic Calendar*. In this picture he presents evolutionary key novelties, including some cultural events as well, between which there are successively shorter time intervals (Sagan 1978 pp. 14-16).

In his comprehensive overview of organic evolution, Richard Dawkins (2004 A) demonstrates the same pattern and Yuval Noah Harari (2012 p. viii) provides in the same vein a timetable of biological and cultural history.

As we can see from these accounts, the novelties appear irregularly but at rapidly shortened time intervals thus harmonizing roughly with the logarithmic time scale of Figure 2. One may raise the objection that the choice of these key novelties is subjective, but it seems to me that the choices, though intuitive, are reasonable and probable shared by many people. It must be emphasized, however, that these accounts of evolution cannot be used for an assessment of the rate of evolution because to that end one also needs an estimation of the degree of change or quality of each novelty. I suggest that the concept of complexity can be used for that purpose.

I have in my construction of the diagram of diagram of Figure 2 suggested the line in the diagram to be a representation of complexity (Ekstig 2010 A). The justification of this conjecture is relying on the explanatory possibilities it may infer. I developed a numerical analysis making possible an estimation of the increasing values of complexity during the evolutionary process, a procedure that can be seen as an operational definition of complexity. I suggest this particular form of complexity to be called *evolutionary complexity*.

The interpretation of the straight line in Figure 2 as illustrating evolutionary complexity relies on two assumptions. The first is that it has been incessantly increasing during the course of evolution. This assumption is supported by the fact that many scientists agree with it as I have pointed out above. Therefore, I don't find this assumption controversial.

The second assumption is that evolutionary complexity increases during the developmental process as well. The problem with this assumption is that all information about the growth of an individual creature is present in its genome at the very beginning of the developmental process and doesn't thereafter change (with some sporadic exceptions) and therefore one could say that the information for the developmental process has been present all the time from its beginning. I think, though, that it is important to distinguish two features of the DNA. First, it contains information about how to build a body and, second, it performs the building work as well. To use a metaphor – the DNA is both the architect and the builder's workman.

As we know, the architect's plan is available from the beginning and is step-by-step accomplished by the builders. In a similar way, the building of an animal body is accomplished, at the supervision of the genes, by means of successive additions of new parts and more connections between the parts, which means an increase of complexity according to the definition of complexity that I suggested above. Therefore I conclude that evolutionary complexity is increasing during the developmental growth of an individual animal. The same reasoning can be applied to the development of our intellectual faculties, though in this case, the information and instructions are stored in the social environment.

To begin with at the formation of the procedure of measuring evolutionary complexity (Ekstig 2010 A). I made the assumption that its growth may be following an exponential function. Therefore, I assumed as a starting hypothesis that evolutionary complexity is following increasing values similar to the negative tail of an exponential function. Hence, its growth is assigned values between 0 and 1.

The next step is to construct a diagram of complexity on a linear time scale. To this end, one must take into account that the position of each point on the line in the diagram of Figure 2 is determined by two time coordinates, one for evolution and one for development.

These two time coordinates for each trait is now represented by a number of pairs of points with their time coordinates on the common linear time axis stretching backwards from the present point of time, thus forming two separate sequences. Next, these pairs of points should be distributed in the diagram according to their coordinates of complexity, up to now undetermined, which is a most crucial procedure.

To that end, one must keep in mind that every pair of these points originates from one and the same point on the line in the diagram of Figure 2, thus representing the same value of complexity. The coordinates of complexity of these points are determined by a tentative formation of an evolutionary curve that gives the most sensible shape of the developmental curve. Because the time axis of the diagram is linear, it is impossible, due to the great difference of temporal extension of the processes, to display the two curves in one and the same diagram. Nonetheless, the procedure can be executed in a numerical mode.

The result of this adjustment of the two curves to each other indicates that an exponential form of the evolutionary curve is unable to give a sensible form of the development curve. Rather, I found that complexity during the course of animal evolution is best described by means of a sequence of exponential functions with increasing bases.

This means that complexity has increased in a superexponential pace, thus growing in an extremely slow pace in early evolutionary times and much faster in the cultural realm (Ekstig 2012) (A superexponential function is a variant of an exponential function in which the base is continuously increasing). The resulting diagram from this procedure is reproduced in Figure 4 in which only the last 100 years of the evolutionary curve is shown.

I suggest that we consider this assessment as a hypothesis and examine its explanatory possibilities. As I already have pointed out, the understanding of evolution has been strongly hampered by the lack of a method for its measurement and I maintain that the suggested method for the measurement of evolutionary complexity gives a possibility to discuss the progress and direction of evolution.

The numerical analysis indicates that evolutionary complexity has been steadily increasing during the course of evolution, although the increase isn't following a simple mathematical formula. A table of its values as obtained from the suggested procedure is given in Table 1.

In the present work, though, I think there is no need of using the specific values of evolutionary complexity as received by the measurement. The measurement has its justification insofar as it makes statements about increasing complexity better grounded. Furthermore, it makes possible the construction a diagram of complexity versus time for the evolutionary process at large, without the requirement of specifying the precise values on the axes. As it turns out, such a diagram implies an entry for several basic insights into the evolutionary process being presented and analyzed in later sections.

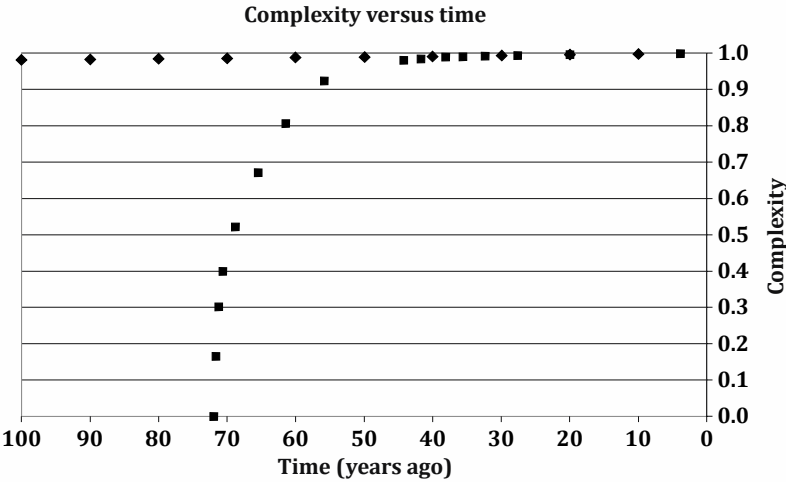


Figure 4. A combined diagram over development (squares) and evolution (diamonds) on a linear time scale. The curves are adjusted to each other to give the most sensible shape of the development curve. The evolutionary curve has its onset 600 million years ago which is a point on the linear temporal axis of this diagram, positioned 600 km to the left. At the point of time 20 years ago, the curves coincide, which corresponds to the point called The critical point in the diagram in Figure 2. The diagram is reproduced from (Ekstig 2010 A).

Table 1. Some relative values of evolutionary complexity as given by the numerical program as reproduced from Ekstig (2010 A)

Evolutionary time (years ago)	Individual age (years from conception)	Relative value of evolutionary complexity
600 million	0.06	10^{-26}
6 million	0.38	0.16
1 million	0.76 (birth)	0.30
25,000	3.2	0.52
1,200	10	0.80
400	16	0.92
100	28	0.98
20	52	0.99

It is interesting to note that the evolutionary process can best be explained by means of a time scale running backwards, starting from the present point of time. Such a feature contrasts sharply to the ordinary way of studying processes in nature by following the direction of time itself, and I think it reflects the fact – by no means unnoticed – that the features of living organisms are best understood on behalf of their bygone history. This result also concurs with the fact that future routes of evolution cannot be predicted. However, the nearest future can in a limited meaning be seen as determined by natural selection. As Dawkins (2004 B p. 103) points out, natural selection is by definition a process whereby information is fed into the gene pool of the next generation. This, as I see it, explains the continuity of the evolutionary process.

Accelerating Evolution

There is a common notion that evolution is an accelerating process. According to the suggested procedure of measuring evolutionary complexity, I determined the successively increasing bases of the consecutive exponential functions representing evolutionary complexity. From these bases one can derive the doubling times of the complete process.

This procedure is described and the resulting doubling times at different evolutionary epochs are given in Ekstig (2012). Here, I reproduce in Figure 5 the obtained graph of the doubling times which gives a conspicuous picture of a significant feature of evolution. Added to my own values are a couple of other measurements for the latest period of time.

If evolution were following an exponential course because of its cumulative character, the doubling time would be constant. However, the present result indicates the overall trend to be much more rapidly increasing with rapidly decreasing doubling times, implying a superexponential evolutionary course. The result implies that at about 100 million years ago the doubling time was about seven million years. Then during the following period of animal evolution it decreased to 600,000 years. At the onset of human evolution the doubling time was about 200,000 years to successively decrease to about 3,500 years at about 400 years ago, a value being characteristic for the pre-Galilean scientific evolution.

For the most recent 200 years, human scientific activity shows a dramatic progress with a doubling time of only 20 years. Finally, information technology over the last century follows a still more rapid rate with a doubling times quickly decreasing to the present value of only 18

months. The diagram shows that evolution at large is following an accelerating pace that in the present time has reached a tremendous rate. Ray Kurzweil (2005) has discussed this trend, suggesting that evolution soon will reach an unlimited speed, thus arriving at what he calls the singularity; a really alarming prospect. If evolution had been found to follow a hyperbolic function, one could expect it to end in such a singularity.

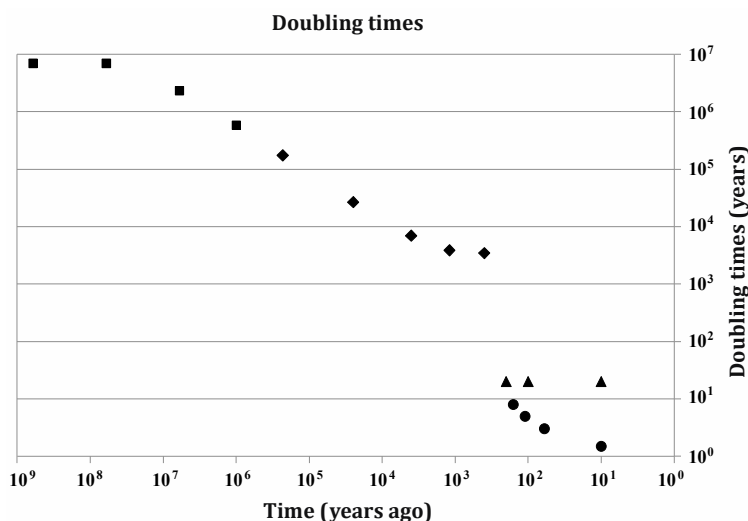


Figure 5. The doubling time of complexity versus evolutionary time. Squares represent animal evolution, diamonds hominid and human evolution including, for the last three points, pre-Galilean scientific evolution, triangles post-Galilean scientific evolution adapted from Jinha (2010), and circles the evolution of information technology adapted from Nagy (2010). The diagram is reproduced from (Ekstig 2012).

However, I have built the analysis of increasing complexity on a sequence of exponential functions that, although rendering the complexity a rapidly increasing pace, doesn't necessarily end up in such a precarious singularity.

The concept of evolutionary complexity has hitherto in this work been discussed in descriptive terms. I now suggest an explication of its growth.

A Self-Reinforcing Feedback Process

An intrinsic feature of life is that living organisms apply the remarkable principle of starting, so to say, from scratch in the formation of each new generation. Only the genetic information is preserved, thus allowing the growing individuals to take advantage of the experiences collected by parents and previous generations. This information is then sieved by the external environment before it is fed back to the next generation or, as I have discussed above, even sieved by a selection process without the involvement of the external environment. In short, this is natural selection combined with a feedback process.

The concept of positive feedback is a source of growth of a system. It originates from the technology of electronic amplifiers, in which part of the output voltage is coupled to the input side. In this way the voltage, which I here call the substrate, is amplified by a feedback loop, a mechanism that can give a rapid and nearly unlimited degree of voltage amplification. This

principle can be transferred to many kinds of processes. Of special interest in the present context is its application to the evolutionary process. Bernard Crespi (2004) points out that positive feedback can be instrumental in driving many of the most important and spectacular processes in evolutionary ecology. He emphasizes that the self-reinforcing dynamics of a positive feedback generates the conditions for changes that might otherwise be difficult or impossible for selection or other mechanisms to achieve and that it can generate large-scale changes in genetic systems.

Crespi mentions the peacock's tail, the human brain, the extinction of passenger pigeons and the infestation of our genomes by repetitive elements as examples of the many remarkable phenomena in evolution and ecology for the growth of which he suggests the mechanism of positive feedback. To his broad account of cases of positive feedback, I would like to add the implementation of complexity, which I suggest can be seen as a common concept behind Crespi's examples.

According to a common description of complexity it means the number of parts or the amount of differentiation among parts within individuals as I have discussed above. Therefore, the addition of a new trait contributes to the complexity of the species. One may assume that this increase of complexity opens still more opportunities for the species to adapt to the environmental peculiarities and thus to add still more traits. In this way, a species is increasing its complexity according to the principle: the higher its level of complexity, the higher the number of opportunities to add still more traits and thus to increase complexity still more.

Such a procedure, even if the implication of its repetition in the developmental course is taken into account, implies a cumulative process that normally leads to an exponential increase.

Let us consider a typical example of a situation in which natural selection gives rise to a cumulative process. An animal's fur is a shelter against cold climate – the colder the climate, the thicker the fur. Thus, the climate stimulates the fur to grow cumulatively but the fur has no influence on the climate. In order that a feedback process should be applicable, there should be a coupling back from the thickened fur to the climate thus leading to a runaway process typical of a feedback in action. In this case there is no such coupling. In many cases in organic evolution natural selection works that way. Are there cases in which natural selection could give rise to a feedback process?

Let us consider such an example. *Arms race* between predators and prey implies that an improvement of the capability of one side will have influence on that of the other thus leading to runaway improvements of the capabilities of both sides. As Dawkins (2004 A p. 496) points out, arms races are deeply and inescapably progressive in a way that, for example, evolutionary accommodation to weather is not. Therefore, the mutual coupling occurring in arms race comprises, I think, the hallmark of a self-reinforcing feedback process. The improvements of the capabilities involved in this process certainly entails an increase of complexity of both sides leading to the conclusion that the self-reinforcing feedback process in arms race causes a rapid increase of complexity that may be much faster than if generated by a cumulative process only.

Sexual selection is another example of a self-reinforcing feedback process. A typical example is found in the conspicuous ornamental plumage of the peacock's tail. In this case the plumage evolution in the male is coupled to a sexual preference for such a plumage in the female. This means that an improvement of a feature on one side will have influence on that of the other. An important feature of this process, as discussed at length by Dawkins (1988 p. 203), is that the genes for male qualities, and the genes for making females prefer those qualities are evolving together leading to mutual runaway improvements, typical of a feedback process.

Another application of a self-reinforcing feedback process is found in cultural evolution in which the evolution of language is especially elucidating. This case will be discussed in a forthcoming section on cultural evolution.

Of course, such a rapidly growing process as that accomplished by the self-reinforcing feedback process cannot go on unlimitedly. The peacock's tail cannot grow infinitely. It seems that the benefit of the new trait starting the feedback process sooner or later is exhausted leading to the kind of natural selection called stabilizing selection; a situation that may go on until interrupted by the emergence of another new trait that may start a new feedback process.

It is conceded that feedback will give rise to an exponential amplification. The cases just discussed are examples of feedback amplification of complexity thus being expected to increase exponentially with time. Such an exponential increase is determined by the base of the exponential function making possible a nearly unlimited degree of amplification. However, during the extended course of organic and cultural evolution several such amplifications have occurred consecutively, each of which supposedly provided with its own characteristic base that in each case may result in a high increase of complexity over short time intervals. In this way, evolution may be characterized by a stepwise growth of complexity, a feature that will be discussed in a forthcoming section.

In addition to this process, natural selection contributes to a successive cumulative addition of new traits in the evolutionary process that may lead to a large-scale exponential increase of complexity. When also the contributions to the growth of complexity from the feedback instances are added, one may find a large-scale increase of complexity that by far exceeds a merely exponential growth.

Despite the periods of stabilizing selection, when seen in the large-scale perspective of evolution, the total pattern of increasing complexity may very well explain the observed increase of complexity as I have previously found and illustrated in Figure 5.

In this way it seems to me that feedback is an inherent principle of evolution that, together with natural selection, accomplishes the generally large-scale pattern of increasing complexity in evolution – a pattern that natural selection by itself is incapable to accomplish. It is now time to discuss complexity a bit closer.

COMPLEXITY AND THE TREE OF LIFE

Some Authors' Views of Evolution

Many renowned authors discuss the general trend of evolution. Thus John Maynard Smith, together with his co-author, Eörs Szathmáry, contend that living organisms are highly complex and that increases of complexity have depended on a small number of major transitions (Maynard Smith and Szathmáry 1995 p. 3). Such notions indicate a stepwise increase of complexity. Edward O. Wilson, though without using the concept of complexity, expresses a comprehensive view of how life has evolved:

Species emerge quickly and fully formed after a rapid burst of evolution, then persist almost unchanged for millions of years. And, conversely, rapid evolution is driven mostly or entirely during species formation. ...The models of population genetics, the foundation of quantitative theory, predict that evolution by natural selection can be so rapid as to seem nearly

instantaneous in geological time. The models also allow for stasis, or long periods with little or no evolution of a kind detectable in fossils (Wilson 1992 pp. 80, 81).

Somewhat later on he contends:

All contemporary dynastic successions taken together present a complex and strikingly beautiful pattern across the surface of the earth. Now the comparison is to a palimpsest, an ancient parchment on which the current dominant groups are boldly spread and past rulers survive as faded traces in spaces between the lines, in shrunken niches. Mammals, the dominant large vertebrates on the land today, are accompanied by turtles and crocodilians, among the last survivors of yesteryear's ruling reptiles. Forests of flowering plants shelter scattered ferns and cycads, remnants of the prevailing vegetation of the Age of Reptiles (*ibid.* p. 86).

It is interesting to note that Dawkins in his analysis of evolution uses the same metaphor of a palimpsest (Dawkins 2004 A p. 23). This metaphor is concurrent with the present view of evolution inasmuch as new species are added while the previously existing species continue their way of living unaffected.

These instances of thoughts by these three renowned scientists have important bearings on the present analysis. Thus Maynard Smith and Szathmáry's views imply evolution to have proceeded in a stepwise way.

Likewise Wilson argues that there have been stepwise bursts of evolutionary changes between which there have been long periods with little change. I interpret Wilson's view in such a way that evolution is proceeding cumulatively while most species continue to live in the mostly unchanged niches to which they are adapted. Because Maynard Smith as well as Wilson base their statements on initiated observations, I take their words as a basic empirical support for my interpretation.

I think that these rationales can be understood in a more comprehensive way by the application of the concept of complexity and the most direct way to do so is by means of a principal diagram over complexity versus time. From now on in this work, instead of using the concept of evolutionary complexity, I use the generalized concept of complexity.

Complexity and a New View of the Evolutionary Process

I suggest that we start by illustrating the evolutionary process by means of a highly idealized diagram of complexity versus time as shown in Figure 6.

I have previously developed the diagram of Figure 6 as a result of an analysis of the evolution of individual developmental courses (Ekstig 2010 B). In Ekstig (2015) I have given a complementary analysis of the diagram as seen as a result of different forms of natural selection. In the present work I extend this analysis and suggest an explanation of the stepwise increase of complexity as a result of a joint action of natural selection and a self-reinforcing feedback process as we have discussed in the previous section.

It is common knowledge (see for instance Reece et al. 2011, p. 527) that natural selection works in three modes – disruptive selection, stabilizing selection, and directional selection. I have analyzed increasing complexity in terms of these concepts (Ekstig 2015), and in the present work I deepen this analysis.

In Figure 3 the uppermost blocks in each column illustrate the addition of new traits to the individual's developmental course. In this process, the new trait must have some advantage of its establishment in comparison to the already existing traits, and I think such an advantage involves an addition of complexity.

This process is in Figure 6 depicted by the steps in the uppermost line, formed by the self-reinforcing feedback process giving rise to a rapid rise of complexity in evolution, a process that is associated to the traditional concept of disruptive selection.

In the diagram of Figure 6, the stepwise raising curve thus represents species emerging in the complexity space after great though rare bursts of evolutionary transitions. After such a transition, the new species normally stabilizes at a new level of complexity, thus forming a new horizontal line above those previously formed. In this way, the diagram illustrates Wilson's comparison with a palimpsest, a metaphor for the cumulative emergence of new species above former ones.

Furthermore, previously formed species are stagnant, illustrated as horizontal lines, persist to exist in what Wilson expresses as shrunken niches – an indication of the hard competition for room in the complexity space. In order to clarify the steps in the uppermost line in the diagram, I suggest as major transitions in the biological part of evolution the emergence of multi-cellular organisms, vertebrates, terrestrial animals, mammals, and man. Other similar or less prominent transitions may be envisaged in between. After these occurrences, human cultural evolution sets in, the steps of which is discussed later on.

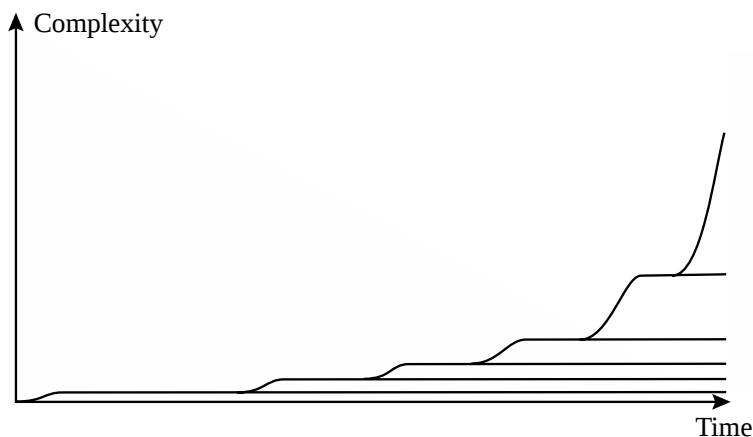


Figure 6. A view of evolution as formed by the processes suggested in the present analysis. The lines represent the complexity of animal species in a highly idealized way although the lines can be interpreted as representing the species per se as well. In this second interpretation, the diagram can be seen as a Tree of Life. The horizontal lines represent stagnant species and the stepwise curve represents the emergence of novel species at the highest degree of complexity. This stepwise curve at the same time represents the common descent of all species. The diagram was first published in (Ekstig 2010 B).

This procedure, when repeated over and over again, leads to a cumulative formation of new species at successive higher levels of complexity. I have suggested this rapid increase to be explained by natural selection and a self-reinforcing feedback process. However, the form of the stepwise curve in Figure 6 isn't just depending on the time interval between the steps but on the height of each step as well, a feature that can be seen as a measure of the quality of the elevation of complexity at that occasion. An analysis of this feature remains to be made. In this

work, I don't specify the details of the curves in Figure 6. It is in the present work sufficient to say that the stepwise curve is a principal illustration of the structure of complexity growth accomplished by the joined effects of natural selection and a self-reinforcing feedback process.

It must be emphasized that the diagram of Figure 6 is highly simplified. Thus, the vast number of species is represented by few lines only. From one such line one may imagine that a lot of related species have emerged, just with slightly diverging features on slightly different levels of complexity. In this way, the complexity space is more or less tightly filled. In the diagram these new species would form horizontal lines lying quite near those of their parent species, although such lines are not shown in the diagram. In this way one may expect that the complexity space will be tightly filled with species.

Such a situation is discussed by Conway Morris (2013 pp. 150-156), concluding that any room for exploring complexity much further is now highly restricted.

Stabilizing selection and directional selection are associated to the cases in which species are kept unchanged or with only small successive changes for long periods of time. Directional selection implies a successive change in a certain direction, for instance towards increased body size, which is a common trend in many species. Such an increased body size keeps most traits unchanged with regard to their function and implies, I think, only a slow increase of complexity. Therefore, the complexity of species applying stabilizing selection or directional selection is represented by the horizontal lines in the diagram in Figure 6.

I summarize this section by pointing out that the rare emergences of radically new species are occurring cumulatively on the uppermost level of complexity, thus rendering the direction of the evolutionary process a significant meaning.

Competition for Room in the Complexity Space

For each of the tightly placed species forming horizontal in the diagram of Figure 6, it may be hard to change to another position in the complexity space because all nearby positions are already occupied. This contention is actually clearly expressed already by Charles Darwin in saying that "competition will generally be most severe between those forms which are most nearly related to each other" (Darwin 1859 p. 121). Daniel Dennett (1995 p. 89) has expressed a similar view, in stating that the odds are heavily against any mutation being more viable than the theme on which it is a variation. What is new in the present view is that competition is seen as occurring in the complexity space.

However, competition of this kind is not present *for species at the highest level of complexity at each point of time, because for them there are no species on higher levels causing competition*. This explains an important feature of this form of the Tree of Life implying that elevations to radically higher levels of complexity occur merely for species dwelling already at the highest level of complexity. Moreover, it gives an explanation of the direction of the evolutionary process, since these elevations to higher levels of complexity occur cumulatively. In this way, evolution becomes asymmetric – it moves mainly in one direction only, the direction of increasing complexity.

This observation is mentioned by Blum (1968 p.175) in his statement that the whole picture of evolution is one in which derivation of new patterns come from modifications of those existing – never by return to a former starting place for a new "try," unless in some very minor instances which we may neglect in our overall view.

Such an instance, though, is accomplished by parasites. These creatures are considered to be evolving towards lower levels of complexity, in the diagram diverging from a horizontal line downwards. However, they are diverging from a species that previously has raised itself to a certain higher level, and therefore parasites don't form a primary type of living organism. Hence they don't, I think, disturb the general conclusions developed in the present work. But maybe parasitism doesn't necessarily lead to a decrease of complexity. As Conway Morris (2013 p. 150) suggests, the interlocking of the genomes of hosts and parasites seems to point towards an under-appreciated degree of complexity.

The reasoning by reference to competition has a most pregnant implication. It means that the evolutionary process is going on in a continuously cumulative way, implying that old species do not suddenly jump up amongst the species dwelling on a much higher level of complexity. Such a transition would need the simultaneous change of many genetic features, which is highly improbable. This reasoning explains the simplicity of the illustration of evolution in Figure 6. The lines are not randomly pointing in different directions or crossing each other. There are no such things as hopeful monsters.

This feature of evolution is clearly demonstrated by Richard Dawkins in his great survey of biological evolution, *The Ancestors' Tale* (Dawkins 2004 A). Although Dawkins doesn't apply the concept of complexity, I find his illustrations of species evolution having much in common with the present model.

The Tree of Life

In the diagram in Figure 6, the lines represent the levels of complexity of animal species in a highly principal way, although the lines may be interpreted as representing the species per se as well, implying that the diagram can be seen as a Tree of Life. This metaphor, introduced by Darwin even before he disclosed natural selection (Darwin 1837), is now widely spread in discussions of evolution. One may find many forms of the Tree of Life in the literature. It seems that Ernst Haeckel in the nineteenth century got pervasive influence with his widely spread picture of the Tree of Life in the form of an old oak, the multitude of branching of which represents the diversity of species.

The form of the Tree of Life suggested in Figure 6 differs in important ways from the traditional form in that it includes the dimension of complexity, perpendicular to the dimension of time.

Furthermore, common descent is formed by the lineage with the highest level of complexity, a lineage represented by the stepwise curve at the uppermost part of the diagram.

Despite the highly simplified form of the Tree of Life it has, I claim, great explanatory value. It can be seen as a summary of many ideas in the present work. Especially, it explains Lamarck's challenging observation mentioned in the introduction, implying that the oldest species are the most primitive in spite of their long exposition to natural selection. The traditional solution of this conundrum is to exclude the use of the terms primitive and advanced – lower or higher – in analyses of evolution, by the same token refraining from the possibility of talking about the direction of evolution.

As I found in my previous measurement of complexity, it displays an accelerating growth. This growth is what is illustrated as the uppermost line in Figure 6. Furthermore, in the reasoning above, I concluded that the accelerating growth of complexity is caused by a

combination of natural selection and repeated self-reinforcing feedback processes. However, if one refrains from the interpretation of Figure 6 as built on complexity and regards it as an illustration of the evolutionary process per se, then also the conclusions in connection to the concept of complexity may be generalized. Thus we may conclude that, although each step in the uppermost line is a result of the disruptive form of natural selection, this process is insufficient in explaining the large-scale accelerating pace of evolution. Instead, these steps are mainly caused by repeated self-reinforcing feedback processes.

The Appearance of Mankind

The cumulative addition of species with successively higher complexity implies that the latest appearing species at each point of time is the one of the highest degree of complexity. This principle has been in action all the way during organic evolution. At present, this species is the human species. As I already have commented, the vast diversity of animal species fills the complexity space very tightly. However, due to the enigmatic extinction of all hominids, there is a broad empty region in the complexity space between modern man and all other animals. This gap makes it easier to regard the human species as different because the borderline between animals and man needn't being drawn in a continuum. It is thought-provoking to fancy about the situation if the *Australopithecus* species or the Neanderthals were still living – by no means an improbable state of affairs. Then it had been a much greater ethical problem to determine which species could be enslaved or be used as food. In the actual state of things, this problem has vanished in that the human species is distinctly separated from the rest of all animals. Harari (2012) discusses this situation in detail.

The question of whether the human species actually can be seen as the one with the highest evolutionary complexity is a highly contentious issue and many authors seem to avoid to comment it. In his book, *The Ancestors' Tale*, Richard Dawkins starts his journey to our ancestors with a point of departure in the human species and he seems to be very careful of how to justify this choice. Thus his justification of his chosen perspective goes like this:

Instead of treating evolution as aimed towards us, we *choose* modern *Homo sapiens* as our arbitrary, but forgivably preferred, starting point for our reverse chronology. We choose this route, out of all possible routes to the past, because we are curious about our own great grancestors (*italics in the original*) (Dawkins 2004 A p. 13).

I think that Dawkins's choice can be understood by use of the concept of complexity. If, for instance, he had chosen to start with the hippopotamus, his diagrams wouldn't have been so simple. I think that his choice to build his diagrams on the human lineage gives the simplest diagram, but I claim that it is only by means of the inclusion of a measure of evolution, here suggested to be complexity, that his choice is justified. Dawkins's many detailed pictures of the evolution of species are, except for this important difference, quite analogous to my picture in Figure 6.

CULTURAL EVOLUTION

Universal Darwinism

As we have seen, the preceding analyses have been based on the discovery of a regular pattern over the last 600 million years of evolution on our planet. The pattern includes the most recent time as well, thus including human cultural and scientific evolution. Indeed, these parts comprise a great deal of the pattern. Therefore, it is interesting to see in what way the discussed principles of biological evolution are applicable to culture as well.

Ever since Darwin, there have been attempts of extension of his original ideas to the field of cultural evolution. Most of these notions have been built on analogies. A great step was taken by Dawkins (1976) with his introduction of the concept of memes as equivalence to genes. Evolution as based on memes is in this conception considered to work by the same principles as that based on genes. Thus, first there is variation so that not all forms of life are identical; undeniably true for cultural forms. Second, there is a selection that chooses those variants that have the highest ability to be spread in the prevailing environment. Third, there is some kind of heredity mechanism through which features of creatures and cultural forms are transmitted, a heredity that in the realm of culture is upheld by memes.

These conditions are thoroughly discussed by Dawkins (1976 and in many other of his works), by Dennett (1995), and by Susan Blackmore (1999), just to mention a few contributors to the extensive literature on cultural evolution in its connection to evolution.

In the above discussion of organic evolution, I emphasized the importance of self-reinforcing feedback in complexity as a central cause of the large-scale structure of evolution. Such a feedback is observed in the field of cultural evolution as well. Thus Richard Alexander (1989) proposes that social competition between and within human groups has implied a feedback of intelligence constantly producing successively enhanced human intelligence. Needless to say, intelligence is a feature of high complexity contents, implying that this feedback can be seen as working on the substrate of complexity.

We will now examine in what way the present model of evolution, as represented in Figure 6, is a good model for cultural evolution as well.

Steps of Cultural Evolution

Figure 6 illustrates the main features of the steady growth of complexity in organic evolution, characterized by major transitions in species at the highest level of complexity, and by the great diversity of stagnant species. The transitions are illustrated as steps in the graph. We now examine to what extent this model is applicable to cultural evolution.

Zimmer (2002 p. 317) suggests that human evolution is marked by five great transitions, the first of which is the separation from the apes appearing some 5 million years ago, pushing our ancestors out onto the African savannahs. The second was the invention of stone tools about 2,5 million years ago and the third the appearance of hand axes. Some half a million years ago, our ancestors learned to master fire and developed a better ability at making spears and other tools. Finally Zimmer mentions the signs of truly modern minds – paintings on cave walls,

jewellery, weapons, and elaborate burials. All these transitions, I think, are good examples of increasing complexity.

Harari (2012) discusses the reasons of the human success and especially emphasizes the significance of the acquisition of language as a unique transition event of the evolutionary emergence of the human species. In the present context, I think one must consider the acquisition of language as an important step in the growth of complexity, because it certainly assumes a nervous system of high complexity. It laid the ground for the uniqueness of the human species and our conspicuous cultural faculties. I suggest the inclusion of some more recent events in our account of important evolutionary transitions, events that are included in the diagram of Figure 2.

The domestication of plants and animals, first appearing around 11,500 years ago, opened the option of abandoning the hunter-gatherer way of living and instead to live in permanent villages or cities. With this way of living followed the possibility of personal possessions and of differentiated tasks in the community. Of special importance was the invention of written language, first appearing around 5000 years ago, that I claim to be an especially important step in the growth of complexity.

I suggest that the next decisive step forward in this summarized historic review was taken in the ancient Greek society. One may especially connect this society to the rise of the ability of arithmetic and geometry. However, after this period, religious forces kept human thought and culture imprisoned for some 1000 years during the medieval period. By the onset of the Copernican revolution, mankind took the next great leap leading to the scientific era starting in western countries and now thriving in a rapidly accelerating pace nearly all over the world. In this way, cultural evolution follows an analogous pattern as that of species evolution, being depicted by the suggested form of a Tree of Life characterized by a stepwise cumulative increase of complexity and an accelerating rate as seen in the decreasing time intervals between the steps. As to the horizontal lines in the diagram, they are interpreted as the level of complexity in societies that by some reason or another not have participated in the growth of science and technology.

A principal difference between organic and cultural evolution is that as soon as a species is split, no reunion is possible and no exchange of experiences is practiced whereas different cultures carry out far-reaching mutual influences. Such mutual influences certainly contribute in a decisive way to the rapid evolution of human culture.

The phenomenon of culture is of course a much more intricate concept than to be represented by a few simple lines in a diagram. And it is indeed highly controversial to suggest different societies to be related to different levels of complexity. In this context I would like to refer to Jared Diamond (1997). In his attempt to explain Eurasian hegemony throughout history he argues that the gaps in power and technology between human societies do not reflect cultural or racial differences, but rather originate in environmental differences.

I would like to highlight Diamond's rationales in that it is not fundamentally a question of ethnical features that makes a distinction between different societies – these differences are by and large due to accidental environmental conditions, especially, as Diamond emphasizes, due to the advantageous occurrence of plants and animals possible to domesticate. After this crucial step, the differences between different societies have been powerfully amplified by various self-reinforcing feedback loops.

Analogies between Biological and Cultural Evolution

My discovery of the regular relationship between development and evolution includes human cultural traits as an important part. Otherwise, the regularity would never have been disclosed and this is, I think, why biologists never discovered it. The diagram of Figure 2 demonstrates how the cultural part of evolution has been united with organic evolution in an integrated, regular, and continuous pattern. As a matter of fact, if we accept the concept of complexity to be a meaningful measure of evolution, the model indicates that human culture provides about half of the total amount of complexity built up on this planet, as demonstrated in Table 1 – by no means an unreasonable result.

Regarding the interpretation of the Tree of Life as depicted in Figure 6 in the realm of human cultural evolution, it is important to differentiate the construction of the human body from our cultural features. The human species can in this context be regarded as an animal, being depicted by a horizontal line in Figure 6, thus indicating that our somatic features have much in common with animals and are changing only slowly. Our cultural manifestations on the other hand, as seen as expressions of complexity, are growing at an extremely rapid pace as compared with the pace of organic evolution. It should be noted that this reasoning doesn't mean that biological evolution has ceased to be in action. It is just that biological evolution proceeds so slowly as compared to cultural evolution that its effects to a large extent are hidden.

As I have pointed out, the diagram in Figure 6 initially depicts the degree of complexity of species. I also suggested that the lines in the diagram could be interpreted as representing species per se, thus forming a Tree of Life. When it comes to the cultural part of the evolutionary process this interpretation must be modified. I suggest that the lines can only be interpreted as representing the degree of complexity of specific cultural manifestations such as the faculties of verbal language, written language, and scientific advancements. However, I would like to suggest that the interpretation of the diagram shouldn't just be restricted to a descriptive demonstration of analogies but being interpreted as an indication of underlying mechanisms as well, especially the question of the application of natural selection.

Natural Selection in Cultural Evolution

In the study of cultural evolution, one may ask to what extent natural selection is applicable to this process. As to this issue, Sir Karl Popper is known to have sharply pointed out the analogy between scientific progress and genetic evolution. He suggested science may be regarded as a means used by the human species to adapt itself to the environment and asserts a fundamental similarity between the three levels of adaptation: genetic adaptation, behavioural learning, and scientific discovery (Popper 1972).

Within sociobiology, as developed by Edward O. Wilson, many human behavioural and social patterns are considered to be evolved through natural selection. Especially, Wilson concludes that gene-culture coevolution is a special extension of the more general process of evolution by natural selection (Wilson 1998 p. 128). Wilson emphasizes that culture and hence the unique qualities of the human species will make complete sense only when linked in causal explanation to the natural sciences, biology in particular (*ibid.* p. 267). However, I think that the role of natural selection remains to be discussed in more detail.

The Process of Heredity in Cultural Evolution

In his thought-provoking book *The Selfish Gene*, Richard Dawkins (1976) suggests cultural evolution to be analogous to biological evolution, primarily inasmuch as there is a corresponding kind of hereditary principle, the notion of memes, in the field of cultural evolution. Thus for instance, it has been asked if religion in human groups has evolved due to an assumed advantage for survival of the group. This might be an open question, but the main reason, as Dawkins expresses it, for a cultural trait to have evolved in the way that it has, is simply because it is advantageous to itself (ibid. p. 214).

When applying the concept of competition as discussed in the explanation of the diagram of Figure 6, I think one may conclude that competition is a driving force for cultural change as well. Thus Dawkins distinctly expresses such a notion: "If a meme is to dominate the attention of a human brain, it must do so at the expense of 'rival' memes (ibid. p. 211). Daniel Dennett articulates a similar view: "Minds are in limited supply, and each mind has a limited capacity for memes, and hence there is a considerable competition among memes for entry into as many minds as possible" (Dennett 1995 p. 349). It is easy to observe that, for instance, there is an incessant competition between different religions and sects, which means a competition of space in people's brains.

Complexity in Cultural Evolution

Regarding biological evolution, I have suggested the concept of complexity to have high explanatory power. We will now discuss if the concept of complexity is applicable to culture as well and if it has increased with time. Many authors intuitively anticipate the notion of a high degree of complexity in the human cultural evolution. Thus John Tyler Bonner points out that instead of all neurons and their connections being specified in the genome, only certain general parameters are specified. In this fashion, it became possible to produce nervous-system structures far more complex than anything the genes might have specified in detail. These structures are the prerequisites for human abilities in learning, teaching, and communicating (Bonner 1988 p. 192).

As to the increase of complexity, Ray Kurzweil (2005) has emphasized the rapid growth of complexity in human evolution. He regards technological evolution as an extension of organic and cultural evolution. He points out that each paradigm develops through three faces in an S-shaped curve and that evolution, both in its biological and technological expression, evolves through a series of such S-shaped curves forming a soft stair-like curve, a view in precise concordance with the diagram in Figure 6.

The Evolution of Verbal Language

As I already have pointed out, the acquisition of language forms a decisive transition event laying the ground for the emergence of the human species. I think that this event marks the transition from natural selection of the traditional form to memetic selection of the form suggested by Dawkins – both these forms of selection being superimposed during this event to be successively dominated by memetic selection. This is so because this form of selection,

driven by the immaterial units of memes, results in a mode of evolution that is enormously much faster than natural selection working on the inheritance of the material units of genes. Such a great difference of the rate of change is perfectly depicted and understood by means of the logarithmic time scale allowing for a superposition of the two processes.

Susan Blackmore (1999) claims that it is the ability to imitate that makes the human species different. She assumes that people will both preferentially copy and preferentially mate with people with the best memes – especially in the case of the best language (*ibid.* p. 104). This is sexual selection, which, as I have emphasized above, implies a source of strong increase of complexity. Since mating is coupled to genetic evolution whereas imitation is related to a memetic evolution, the first steps of verbal evolution initiated, one could say, the transition from genetic to memetic evolution. It must be pointed out, though, that not all biologists, not to say all linguists, accept the notion of human culture as a product of Darwinian evolution. Especially regarding language, the renowned linguist Noam Chomsky is known to fervently disagree with such an idea. However, Daniel Dennett rejects Chomsky's approach (Dennett 1995 p. 390).

The question in what degree the evolution of language is driven by natural selection is a controversial question, discussed at length for instance by Dennett (1995). It seems plausible that the ability to acquire language is promoted by high intelligence and cleverness, features that certainly had high survival value in the days when all kinds of environmental hazards constantly threatened the survival of the small bands of humans. In such situations, language has certainly been beneficial for survival. But is this process really Darwinian? I think that the survival isn't coupled to a particular physical environment because the causal relation between environment and the acquisition of language isn't as apparent as that for instance between a cold climate and a thick fur. One may say that the verbal ability is equally beneficial in any physical environment. But if the social environment is taken into account, the situation may be quite different, because what is important in the social environment in this context is actually the verbal ability of the population.

Therefore, there is a kind of mutual causal coupling between the individual verbal ability and that of the population. When a child grows up it will contribute to the verbal level of its social environment that enhances the learning of language in the next generation of children. This situation means of a type of self-reinforcing feedback mechanism that I think has had a strong effect on the evolution of language. We may recall that the feedback principle is discussed above in connection with the organic part of evolution.

Thus, in the case of the evolution of language, there is a kind of mutual coupling inasmuch as the child, when grown up, has impact on its environment that has consequence for the next generation of children. This process really involves the hallmark of self-reinforcing feedback: the same causal loop repeating generation after generation. This feedback mechanism, I claim, has had a strong effect on the evolution of language and has caused its rapid changes in comparison to organic changes.

There may be a more general application of the action of feedback in the evolution of the human brain. Thus Daniel Dennett (1991) concedes that "the haven all memes depend on reaching is the human mind, but the human mind is itself an artifact created when memes restructure a human brain in order to make it a better habitat for memes." Then he continues: "The memes enhance each others' opportunities: the meme for education, for instance, is a meme that reinforces the very process of meme-implantation" (*ibid.* p.207). Such mutual

couplings, typical for feedback, have, I think, had a decisive importance for the extremely rapid growth of the human brain as seen in comparison with other somatic traits.

Condensation and Terminal Addition

The increase of verbal ability in a society is coupled to an enhancement of children's verbal skillfulness. Such an enhancement is associated to an earlier acquisition during the child's growth. In the terminology of the present work, we may conclude that there is a condensation of verbal ability. Likewise, as all of us have observed, young people continuously invent new words and a new vocabulary that, in the terminology of the present work, may be seen as terminal additions. Thus I conclude that the concepts of condensation and terminal addition, as developed in the field of organic evolution, are acting in the field of language evolution as well.

Processes in the Evolution of Written Language

The next point after that of verbal language on the line in Figure 2 is written language. I think it is plausible that for instance in the ancient Greek society, writing was an activity exclusively performed by adult persons. I imagine the situation to be the same for the runic writing amongst the Nordic Vikings. But nowadays, most children exercise writing. This has of course to do with intentional education. I think that progress in the western society and the Eurasian hegemony to a large extent must be seen as a result of the organization of universities and, somewhat later, of school systems. During the last 200 years, there has been an awakening political awareness of the importance of education for the wellbeing of the population and that this education must be extended over merely religious indoctrination as essentially was its task in medieval times.

Moreover, education has substantially improved its pedagogical methods resulting in a rapid enhancement of learning. This also means that the children learn the specific technics at successively earlier age; in other words, there is condensation in children's learning.

It may be interesting to see in this context that even the formation of the letters has gone through a process of simplification. The gothic letters from the sixteenth century have successively been developed into more simple forms that certainly have made reading and writing more easily acquired, in this way making earlier learning possible.

In the field of evolution of written language, the changes are intentionally brought into use. This leads as to the application of intentional selection; the primary type of selection in the fields of science and technology.

Processes in Scientific Evolution

As we can see in the diagram of Figure 2, the scientific evolution forms a considerable part of the pattern. It is therefore of interest to see if one can talk about analogies between the biological, cultural and the scientific evolutionary processes as well as similarities in mechanisms. There is a remarkably early formulation of such a notion, expressed by the 1905

Swedish Nobel laureate in chemistry, Svante Arrhenius (1859-1927) in a book from 1907, thus at the time when the discussion of Darwinian evolution was intense.

It is with ideas as with organisms. A lot of seeds are sown, but only a few will grow, and amongst the living things being evolved from them, most are weeded out through the struggle for existence, and only a few will remain living. In a similar way, ideas that are most successfully corresponding to nature are gradually selected (Arrhenius 1907. pp. 175-176, my translation from Swedish).

As we can see, Arrhenius associates the selection of ideas to the Darwinian concept of struggle for existence, in other words, to natural selection. However, there is an important difference in that, in the field of scientific evolution, the selection is intentional.

I would like to mention in passing that Arrhenius is the discoverer of the coupling between atmospheric carbon dioxide and the greenhouse effect; an issue of current interest in our own time.

The school courses in mathematics and physics, as traditionally given in many classrooms, are usually organized so that the sequence of subjects more or less parallels the historical evolution, whether this parallel is made explicit or not. This observation is in concordance with the Jean Piaget's principle of genetic epistemology.

By means of this concept, Piaget suggested that there is a parallelism between the progress made in the logical and rational organization of knowledge found in the history of science and the corresponding formative psychological processes developed in children (Piaget 1970 p. 13). Piaget reported, together with his co-author, an investigation of the relationship between the child's mental development and the cultural and scientific history, restricted to the fields of mathematics and physics. In these fields, the authors observed surprising coincidences in method as well as content and, as they emphasized, very striking common features. What is interesting about this observation is that children construct old concepts without ever being taught about them (Piaget and Garcia 1983 p. 292). Piaget's observation is especially interesting for the present model, because it emphasizes the important role of children in the process of scientific evolution. This, as we may recognize, is in line with the present model, accentuating as it does the importance of the developmental process in evolution. Especially, condensation is such a process.

The evolution of science is connected to the intentional education and the effect of school systems. These occurrences have certainly implied a change towards earlier acquisition, i.e., condensation, of the central concepts of science – indeed significant conditions for the rapid evolution of science.

Moreover, I think that the progress of scientific evolution can be accounted for by a self-reinforcing feedback process, in which the results of scientific research successively are transmitted to school and university curricula, thus enhancing the possibilities of continued progress.

I have illustrated the contribution of science to the superexponential increase of complexity in Figure 5. Of course, one must be aware of the fact that the assessment of complexity is performed by means of different methods in the different fields, and therefore conclusions of the pattern shouldn't be drawn too far. Yet, despite this limitation, the present analysis of scientific progress contributes to the suggested overarching view of evolution as an integrated process of biological, cultural and scientific evolution.

The Human Species – a Unique Species

It is, I claim, the cultural evolutionary process that contributes most to the superior complexity level of mankind. This part of the evolutionary process, according to my assessment of complexity, contributes to as much as half of the total evolutionary complexity developed on our planet. I have discussed this interpretation at length in previous publications (Ekstig 2010 A, 2015). According to the model of evolution proposed in the present work, it is the latest appearing species that takes the position at the highest level of complexity. In our own time, this species is the human species. This gives us reason to place the human species at the highest position in the hierarchy of living creatures. This conclusion concurs with the intuitive though controversial notion of man as the summit of evolution.

In fact, this contentious notion can be traced back to the medieval idea of the Great Chain of Being, indeed even to Aristotle. It is interesting to note that this obsolete idea comprises a hierarchical but not a time dimension, whereas many scientists of today, for instance Dawkins in *The Ancestor's Tale* (Dawkins 2004 A), arrange life on a temporal but not on a hierarchical dimension. The present model, as we have seen, attempts to describe the evolutionary process by the application of both these dimensions.

The notion of mankind as a unique and superior species is a highly controversial issue not frequently expressed in scientific literature. However, a couple of examples may be mentioned. Thus, Carl Sagan eloquently emphasizes our unique intelligence:

We are a thinking species. That's what we are good at. We're not faster than other animals, we're not better camouflaged, we don't dig better, swim better. We think better. And because of our hands we build better. That's our peculiar genius and the chief reason for the success of the human species (Sagan 1989).

Likewise, Daniel Dennett in his broad survey over Darwinism boldly states:

People ache to believe that we humans are vastly different from all other species—and they are right! We are different. We are the only species that has an *extra* medium of design preservation and design communication: culture. That is an overstatement; other species have rudiments of culture as well, and their capacity to transmit information “behaviorally” in addition to genetically is itself an important biological phenomenon, /.../ but these other species have not developed culture to the takeoff point the way our species has (Dennett 1995 p. 338).

Furthermore he points out that cultural evolution operates many orders of magnitude faster than genetic evolution, and this is part of its role in making our species special (ibid p. 339). As we have seen, such a different rate of evolutionary changes is a central component of the present model. In his ambition to strengthen human dignity, American philosopher George Kateb fervently articulates the superiority of mankind amongst all species:

We human beings belong to a species that is what no other species is; it is the highest species on earth—so far. /.../ All other species are more alike than humanity is like any of them; a chimpanzee is more like an earthworm than a human being, despite the close biological relation of chimpanzees to human beings.

The small genetic difference between humanity and its closest relatives is actually a difference in capacity and potentiality that is indefinitely large, which actually means that it can never be fully measured (Kateb 2011 p. 17).

It is interesting to note that Kateb's intuitive comparison between earthworms, chimpanzees and human beings actually coincides surprisingly well with my own assessment of complexity, implying that the evolution of human culture contributes with as much as half of the total value of complexity. I just hope that my model of the evolutionary process may increase our appreciation of mankind's dignity.

Finally, in pointing out citations in favour of the notion of mankind as a unique and superior species I would like to refer to the last sentence in Dawkins's seminal book *The Selfish Gene* that reads:

We, alone on earth, can rebel against the tyranny of the selfish replicators. (Dawkins 1976 p. 215).

One may wonder why so many scientists so fervently reject the depiction of animal evolution as following a direction towards successively higher levels, whereas lay people, as I have understood, see it as intuitively obvious. As I have pointed out, it seems that as to this issue, scientists refer to the current opinion that there is no measure of evolutionary progress.

Maybe also there is an apprehension that if one accepts to regard different species as living at different levels, one might as well consider different putative human races at different levels; if at all the notion of races is meaningful. I ardently emphasize that the present model of the evolutionary process doesn't endorse such an interpretation. I find support of this statement in a discussion by Dawkins (2004 B) saying that the great majority of human genetic variation is to be found within races, not between them (*ibid.* p. 76). This means that the horizontal lines in the diagram of Figure 6 cannot be interpreted as putative human races. This has also to do with the fact that animal species are reproductively isolated from each other whereas, as to culture, there is no such limitation. On the contrary, many human cultures have always exchanged ideas and knowledge in a way that has made their differences hard to classify. Genes are stuck to their particular species whereas memes are floating freely within and between societies.

CONCLUSION

In the present work, I give an all-embracing perspective on processes of the evolution of life and culture on earth. Thus I see the evolutionary processes as occurring in two dimensions; those of complexity and time. By means of the inclusion of complexity, evolution gets a hierarchical order.

The problem with the concept of complexity is generally seen in the absence of its definition. I have evaded this difficulty by suggesting an operational definition built on a procedure of measuring complexity by means of a combination of empirical data from the evolutionary and developmental processes. In this procedure I made use of the process of condensation of developmental traits, a process that I found to proceed regularly and independently of the haphazard external contingencies. In this way I found that the performed

measurement of complexity revealed a regular and rapidly accelerating growth over a great part of the evolutionary history, explained in part as a result a self-reinforcing feedback process in complexity. In this way the wide-spread though intuitive notion of a rapidly growing pace of evolution is supported.

Natural selection is generally considered as the key process of organic evolution which, according to the common interpretation, adapts populations to their environments, thus producing enduring though irregular evolutionary changes. In the present work, I investigate complementary possibilities of an understanding of the large-scale evolutionary process.

Thus I suggest a new kind of natural selection that diverges from the traditional form in that it acts independently of environmental contingencies. This form of natural selection has its application in the process of individual development that is found to be regularly shortened. Together with the development process, the large-scale evolutionary process is found to follow a simple general pattern.

Furthermore, I have found that natural selection is insufficient in explaining the large-scale course of evolution as characterized by a rapid acceleration of its growth. This feature is instead explained as a result a self-reinforcing feedback process in complexity.

Likewise, I have found natural selection to be insufficient in explaining the human cultural and scientific manifestations of evolution. In these fields I suggest an explanation by application of two processes: a self-reinforcing feedback process and intentional selection.

My analysis of evolution in the light of complexity has led me to the construction of a new form of a Tree of Life, displaying complexity versus time. This illustration of evolution comprises stagnant species as well as the stepwise rapid bursts of novel species implying that the latest emerging species dwells at the highest level of complexity. At present this species is the human species rendering us the position as the summit of evolution on earth.

REFERENCES

- Alexander, R. D. 1989. Evolution of the human psyche. In: P. Millar and C. Stringer (Eds.), *The human revolution: Behavioral and biological perspectives on the origins of modern humans* (pp. 455-513). Princeton: Princeton University Press.
- Arrhenius, S. 1907. *Människan inför världsgåtan*. [Human beings face the world]. Hugo Gebers: Stockholm.
- Arthur, W. 2011. *Evolution: A Developmental Approach*. Wiley-Blackwell.
- Arthur, W. 2002. The emerging conceptual framework of evolutionary developmental biology. *Nature*, 415, 757-764.
- Balinsky, B. I. 1975. *An Introduction to Embryology*. W. B. Saunders Company, Philadelphia.
- Blackmore, S. 1999. *The Meme Machine*. Oxford: Oxford Univ. Press.
- Blum, H. F. 1968. *Time's arrow and evolution*. Princeton University Press. Princeton, NJ.
- Bonner, J. T. 1988. *The Evolution of Complexity*. Princeton University Press. Princeton, NJ.
- Bowler, P. J. (1989) *Evolution, the History of an Idea*. Berkeley: University of California Press.
- Carroll, S. B. 2005. *Endless forms most beautiful: The new science of evodevo and the making of the animal kingdom*. W. W. Norton. London.
- Carroll, S. B. 2008. Evo-Devo and an Expanding Evolutionary Synthesis: A Genetic Theory of Morphological Evolution. *Cell*. 134, 1, 25-36.

- Conway Morris, S. 2013. *Life: the final frontier of complexity?* In: Lineweaver, C. et al. (2013).
- Crespi, B. 2004. Vicious circles: positive feedback in major evolutionary and ecological transitions. *Trends in Ecology and Evolution*. Vol. 19 No. 12.
- Darwin, C. 1837. Available at http://darwin-online.org.uk/EditorialIntroductions/vanWyhe_notebooks.html.
- Darwin, C. 1859. *On the Origin of Species by Means of Natural Selection*. Murray. London.
- Dawkins. R. 1976. *The Selfish Gene*. Oxford: Oxford Univ. Press.
- Dawkins. R. 1988. *The Blind Watchmaker*. Penguin Books, London.
- Dawkins. R. 1992. *Progress*. In: Fox Keller et al. (1992).
- Dawkins, R. 2004 A. *The Ancestor's Tale*. Weidenfeld and Nicolson.
- Dawkins, R. 2004 B. *A Devil's chaplain*. A Mariner Book. New York.
- Dennett, D. C. 1991. *Consciousness Explained*. Little, Brown and Company, Boston.
- Dennett, D. C. 1995. *Darwin's Dangerous Idea*. New York: Simon and Schuster. p. 390.
- Diamond. J. 1997. *Guns, Germs, and Steel: The Fates of Human Societies*. W. W. Norton and Company, Inc. New York.
- Ekstig, B. 1994. Condensation of developmental stages and evolution. *BioScience* 44(2):158-164.
- Ekstig, B. 2007. A Unifying Principle of Biological and Cultural Evolution and its Implications for the Future. *World Futures*, 63: 98-106.
- Ekstig, B. 2010 A. Complexity and evolution: A study of the growth of complexity in organic and cultural evolution. *Foundations of Science* 15: 263-278.
- Ekstig, B. 2010 B. Biological and Cultural Evolution in a Common Universal Trend of Increasing Complexity. *World Futures* 66: 435-448.
- Ekstig, B. 2012. Superexponentially Accelerating Evolution. *World Futures: The Journal of Global Education* 68:1, 40-48.
- Ekstig, B. 2015. Complexity, Natural Selection and the Evolution of Life and Humans. *Foundations of Science*. 20: 175-187. Open access: <http://link.springer.com/article/10.1007/s10699-014-9358-y>.
- Fox Keller, E. and Lloyd, E. 1992. (Eds.) *Keywords in Evolutionary Biology*. Cambridge: Harvard University Press.
- Gould, S. J. 1977. *Ontogeny and phylogeny*. Harvard University Press.
- Gould, S. J. 2002: *The Structure of Evolutionary Theory*. Harvard College.
- Hall, B. K. 2002. Evolutionary Developmental Biology. In: Minugh-Purvis and McNamara: *Human evolution through developmental change*. The John Hopkins University Press, Baltimore.
- Harari, Y. N. 2012. *Sapiens. A Brief History of Mankind*. Harper Collins Publishers: New York.
- Jinha, A. E. 2010. Article 50 Million: An Estimate of the Number of Scholarly Articles in Existence. URL: <http://www.stratongina.net/files/50millionArifJinhaFinal.pdf/>.
- Kalinka A. T. and P. Tomancak. 2012. The evolution of early animal embryos: conservation or divergence? *Trends in Ecology and evolution*, Volume 27, Issue 7, pp. 385s-393.
- Kateb, G. (2011) *Human dignity*. The Belknap Press of Harvard University Press.
- Kauffman, S. 1993. *The Origins of Order*. New York. Oxford University Press.
- Kauffman, S. 2013. Evolution beyond Newton, Darwin, and entailing law: the origin of complexity in the evolving biosphere. In: *Lineweaver, Charles et al. (2013)*.

- Kurzweil, R. 2005. *The singularity is near: When humans transcend biology*. Penguin Books Ltd.: New York.
- Lineweaver, C. H., Davies, P. and Ruse, M. 2013. (Eds.) *Complexity and the Arrow of Time*. Cambridge: Cambridge Univ. Press.
- Maynard Smith, J., Szathmáry, E. 1995. *The Major Transitions in Evolution*. Oxford University Press: Oxford.
- McKinney, M. L. and K. J. McNamara. 1991. *Heterochrony*. New York. Plenum Press.
- McNamara, K. J. 2012. Heterochrony: the evolution of development. *Evolution: Education and Outreach*, 5(2). pp. 203-218.
- McNamara, K. J. 2002. What is heterochrony? In: Minugh-Purvis and McNamara: *Human evolution through developmental change*. The John Hopkins University Press, Baltimore.
- McShea, D. W. 2001. The hierarchical structure of organisms: A scale and documentation of a trend in the maximum. *Paleobiology*. 27(2).
- McShea, D. and Brandon, R. 2010. *Biology's First Law*. Chicago: The University of Chicago Press.
- Mitchell, M. 2009. *Complexity, a guided tour*. Oxford University Press.
- Müller G. B. 2007. Evo-devo: extending the evolutionary synthesis. *Nature Reviews Genetics* 8, 943-949 | doi:10.1038/nrg2219.
- Nagy, B. et al. 2010. *Superexponential Long-term Trends in Information Technology*. URL: <http://www.santafe.edu/~jdf/papers/Superexponential.pdf>.
- Piaget, J. 1970. *Understanding Causality*. W.W. Norton and Co. New York.
- Piaget, J. and R. Garcia. 1983. *Psychogenèse et Histoire des Sciences*. [Psychogenesis and History of Science]. Flammarion, Paris.
- Popper, K. 1972. *Objective knowledge: An evolutionary approach*. Oxford University Press.
- Raff, R. A. 1996. *The Shape of Life*. Chicago: The University of Chicago Press.
- Raff, R. A. and G. A. Wray. 1989. Heterochrony: Developmental mechanisms and evolutionary results. *Journal of Evolutionary Biology*, Volume 2, Issue 6, pages 409-434.
- Reece, J. B. et al. 2011. *Campbell Biology*. Boston: Pearson.
- Rice S. H. 1997. The analysis of ontogenetic trajectories: when a change in size or shape is not heterochrony. *Proc. Nat. Acad. Sci.*; 94:907-12.
- Richards, R. 1992. *The Meaning of Evolution*. The University of Chicago Press, Chicago.
- Sagan, C. 1978. *The Dragons of Eden*. Ballantine Books, New York.
- Sagan, C. 1989. "Why scientists should popularize science." *Am. J. Phys.* 57 (4).
- Stearns, S. C. 1992. *The Evolution of Life Histories*. New York: Oxford University Press.
- Tills, O., Rundle, S. D., Salinger, M., Haun, T., Pfenninger, M. and Spicer, J. I. 2011. A genetic basis for intraspecific differences in developmental timing? *Evol. Dev.*; 13:542-8.
- Trivers, R. (1976) In: Dawkins. R. *The Selfish Gene*. Oxford: Oxford Univ. Press. p. vii.
- Wilson, E. O. [1992] 1994. *The Diversity of Life*. Penguin Books.
- Wilson, E. O. 1998. *Consilience*. Alfred A. Knopf, New York.
- Zimmer, C. 2002. *Evolution*. The Random House Group Limited, London.

Chapter 13

NATURAL SELECTION AND DIABETES MELLITUS

Svetlana Shtandel*

V. Danilevsky Institute for Endocrine Pathology Problems
Kharkov, Ukraine

ABSTRACT

Advances in modern medicine enable a change in the tension of intragroup selection in human populations. Thus, implementation of insulin for type 1 diabetes mellitus (DM) treatment considerably lowered the selection tension for this symptom and converted it from the sub-lethal to the one with a lowered adaptability. Increasing variety of type 1 DM and type 2 DM is being observed recently in different populations. Moreover, recently the heterogeneity of type 1 DM and type 2 DM has also been observed. The investigation was aimed to study the influence of the selection on the evolution of DM clinical forms. Global implementation of insulin therapy into type 1 DM treatment caused the prevalent increase of this disease. Currently, there is a positive selection trend for type 2 DM, which is the original cause for the prevalence increase within the population, and the negative one for type 1 DM determines its prevalence within the population approximately on the same level. Intra-population change of gene frequencies, susceptible for type 1 and 2 DM, predetermined the development of such DM clinical forms as LADA (latent autoimmune diabetes in adults). It also resulted in the increasing number of patients with an absolute insulin deficiency of type 2 DM, which is a more complicated form of this DM type. Polymorphisms association, changing the immune response and forming the susceptibility to type 1 (C1858T of the gene *PTPN22*, A49G of the gene *CTLA4*) and type 2 (E23K of the gene *KNJ11* participating in insulin insufficiency formation) with different DM forms illustrates the result of decreasing the selection tension against type 1 DM after insulin therapy implementation into the public health services practice.

Keywords: Type 1 and 2 diabetes mellitus, latent autoimmune diabetes of adults, natural selection, relative adaptability, prevalence in population

* Corresponding Author's Email: shtandel@mail.ru.

INTRODUCTION

Population dynamics factors that change the gene frequencies include: a) recurrent mutations; b) natural selection; c) migration and d) casual fluctuations [1]. It is known, that the genes and chromosomes mutations are the unique source of all genetic variability, although frequency of their occurrence is very low. As the mutations process is very slow, they by themselves change the population genetic structure with a very low speed [2]. Selection is the principle driving force of the evolution. Modern definition of selection is generally used as a differential reproduction of various genes variants, i.e., carriers of certain traits, have more chances to survive and have geniture, than the carriers of other traits [2]. The central concept of the selection theory is adaptability. Only the different reproduction speed of individuals with various genotypes is important for selection. The reproductive ability of a particular genotype in comparison with the norm is called the Darwinian adaptability of this genotype [1]. The adaptability level depends on the decrease of an individual viability and because of those or other pathologies, impossibility to survive till the reproductive age, and also the fertility decrease of any individual [3].

I. I. Shmalgauzen, who has contributed greatly into the problem of evolution studying and search of its solving, wrote: "As phenotypes are the carriers of viability and objects of the natural selection – the individual development course could not bur influence the evolution... the most important thing - an organism as such with its active struggle for its life is not visible in the genetic theory of natural selection." And, according to I.I. Shmalgauzen, phenotypes are the objects of natural selection, or, it can be said, that natural selection influences the phenotypes. Purpose of the thesis about selection influence on phenotypes assumes the necessity of relation between the selection and the organism properties [4].

Natural selection concept is a discordant problem of evolutionary human genetics. Despite the popularity of a hypothesis of "neutral evolution," most of the scientists believe that selection has played the principle role in evolution of species and has generated all the biological diversity of human populations [5]. While considering the large human populations, the following types of selection are distinguished: 1) intragroup, based on inter-individual adaptability differences (differential reproduction of genotypes) and 2) intergroup, which takes into consideration the differences in average adaptability of populations (differential natural growth of particular groups) [6]. There is a couple of difficulties, caused by applying the biological adaptability concept in respect to a human, measured by number of viable geniture, which does not consider the human society social organization specificity. The modern medicine allows to "correct" the phenotype manifestations of some hereditary pathologies, to create the adaptive environment for the genotypes, which in more severe constraints would be eliminated by selection, and in thus to increase their adaptability. This phenomenon was called "a dysgenic effect of medicine" [7, 1]. Change of the intragroup tension - for example, by successful treatment of multifactorial diseases (MFD) - leads to change of the distribution curve of the disease susceptibility and, hence, to its distribution increase. However, the quantitative estimation of such changes is complicated.

In the modern literature, the issues of selection in human populations [7-9] were actively studied. For this purpose, the methodology of selection intensity estimation (I_t) and its components, bound with differential pre-reproductive mortality (I_m) and differential fertility (I_f), is applied (developed by J. Crow [10]). However, meaning of selection in human

populations is being disputed till now. Thus, some authors assume, that selection in the modern urbanized populations is practically absent [11], the others - that available data do not demonstrate the selection weakening in the modern urban populations [12]. Up to half of human gene pool is not reproduced in the succeeding generation because of the embryos death, fetal deaths, neonatal mortality, death-rate before reproductive age, celibacy and sterile marriages in developed countries [13, 14]. Factors of economic progress and social development of the society influence the selection orientation and intensity in populations, but do not stop its action. In different populations the components of differential fertility and differential mortality considerably differ from each other [8, 15-17].

It should be noted that on different stages of mankind history, selection intensity and orientation were exposed to considerable fluctuations. Ancient populations of hunters and collectors were characterized by a moderate level of pre-reproductive mortalities ($I_m = 0.58$) and fertility (5-6 descendants; $I_f = 0.34$), selection intensity ($I_t = 1.11$), which were not changed significantly during the time [6]. Selection intensifying was noticed in populations of the ancient farmers when the population density increased. Maximal values of Crow indexes, known from all published data, were observed at early stages of urbanization. In the end of XIX century in the Polish community of Pittsburgh (USA) the differential mortality component reached 2.98 (75% of children did not survive till the reproductive age), the fertility component reached 0.99, and the index of total selection was 6.92. In Moscow at the same time, the pre-reproductive mortality reached 60.0%, and the index of total selection was 3.15 [6, 8, 13]. Such high values of pre-reproductive mortalities were explained by a high mortality from infectious diseases, high population density, absence of medical aids and severe life conditions of the majority of citizens. Only in XX century, as a result of social progress and successes of public health services, there was a sharp reduction of pre-reproductive mortalities in the developed countries. Simultaneously this process was accompanied by birth rate decreasing [6, 16]. Growth of population genetic burden is a predicted consequence of the selection intensity depression [6, 18]. Numerous studies, devoted to neonatal mortality rising, increasing of individuals part with various congenital defects of development and immune system, mental and cardiovascular diseases were confirmed this thesis [12].

Particularly, thanks to development of medical practice, in economically developed countries, selection by elimination of nonviable individuals moves towards the early stages of ontogenesis. However, the biological selection on the later stages of ontogenesis in these countries is going on, though not by means of elimination, but in the form of retardation. After all, it is obvious, that self-preservation means both individual surviving, and surviving in the sequence of generations. Thus, for realization of biological selection it is enough that some individuals survive more successfully, producing more geniture than others [19].

One of the forms of selection action is elimination. On the average, in humans approximately 50% of zygotes are eliminated, 15% of embryos till birth and near 5.0% of children dye at birth, about 3% more do not survive till maturity [20, 21]. Biological selection by elimination takes place also at the later ontogenetic stages, especially in underdeveloped countries, in which at the centuries border there lived 4.4 billion persons, 3/5 of which lived without elementary hygienic conditions, and 1/5 had no access to modern medicine [21]. In such conditions, biological selection by elimination is unviable (for example, selection on the infections resistance).

From everything we know about selection in modern society, one can make two conclusions: 1) Biological selection in human population really reacts by means of selective

elimination and inhibition. 2) However, in society there is no biological selection as an independent, determinative factor of evolution, irrespective of any social patterns. With appearance of labor, the biological selection is not getting weaker, but changing its final orientation, submitting to the social factors and social production. Finally, the social production realities define competition and selection course, elimination and inhibition in the modern society, i.e., are expressed as social selection which is carried out on “basis” of biological selection and consequently has biological consequences for society, biological “equivalents” [22]. Depending on the social production development level, the biological selection acts either via elimination at all stages of ontogenesis, or more generally, in the form of inhibition. Negative consequences of selection weakening through elimination, which are expressed in the economically developed countries in accumulation of “genetic burden,” are corrected by the social production: development of medical practice [23] and “incoming of fresh blood” by means of immigration.

One of the basic achievements in biology is a new clear recognition, that the following is required for all biological traits: a) to explain precisely how the trait “works”; to explain from the evolutionary point of view, for what the trait exists [24]. Medical researches concentrated on features of the organism functioning and on immediate factors, explaining why in some people disease occurs, and in others - not. The Darwinian medicine states the other, evolutionary question. Why every organism, in more or less extent is susceptible for diseases? For what appendix and wisdom teeth exist? Why our coronal arteries are so narrow? Why the mammary gland cancer occurs so often now? From the first sight, the answer seems to be simple. Natural selection is a random process; therefore it cannot lead any trait to the highest perfection [25, 26]. However, the latest, more careful analysis revealed some other reasons, caused by evolution, of the organism remaining vulnerable to diseases: new environmental factors, our organism is not adapted for; the compromise construction, which makes us more vulnerable to diseases, but induces the general benefit; microorganisms, which evolve faster, than people do; and also the defensive traits, such as pain and cough, which are similar to illness, but actually they are the protective mechanisms, generated in the course of natural selection. There is a number of works, studying how the evolutionary approach creates the base for MFD understanding [27, 28].

It is known that the food consumption was irregular at early stages of human development. In those conditions the possibility to form deposits of fat tissue and carbohydrates in the case of their big consumption was an important adaptive mechanism, allowing to survive in the periods of poor nutrition and starvation. It is also essential, that at the first stages of human development, human’s life was characterized by an exact allocation of men and women roles. Men went for hunting; women stored the hearth side, gave birth and brought up children. Thus, men were supposed to have stronger muscles that caused the primary development of the muscular tissue. For geniture feeding, women had to be adapted to keep energy reserves, i.e., to have developed fatty tissue that served as a protection of newborns against starvation at the time of poor nutrition. During later periods, food obtaining was carried on no longer by hunting, but by development of agriculture and cattle breeding, that also demanded great physical expenses. There was no abundance of nutrition for the majority of people. The certain balance between caloric intake and energy consumption was maintained. In a very short time period, from the point of view of evolution, the major part of mankind started to get plentiful nutrition without starvation periods and muscular energy expenses, related to severe life conditions. One can assume that a modern human, and firstly a man, who usually has less developed fatty tissue

in comparison to a woman, obtains ischemic heart disease and myocardial infarction as a payment for satiety in combination with his muscular energy small demand [29].

Wide distribution of a particular genotype in a population takes place in the case when this genotype gives to its owner selective advantages against the individuals not possessing it, for example, the ability to survive in extreme conditions, in particular in conditions of hunger. In conditions of constant food excess, insulin apparatus is continually under increased loading due to changes in balance between caloric intake and energy consumption. As per J.V. Neel theory, as it was mentioned above, it is presupposed [28], that there are special genes (thrifty genes), allowing a person to use effectively the limited food resources, that leads to DM development in conditions of nutrition abundance. This theory has found its confirmation in works of W. Knowler et al. [30] and P. Zimmet et al. [31], who showed the rapid increase of 2 type DM frequency among the indigenous population of Northern America and Pacific Islands in conditions of urbanization and nutrition abundance.

Globalization processes, existing in modern society, lead to increase of outbreeding, panmixia, genes mixing, and segregation burden increase. Significant part of a genome is invariable, monomorphic and conservative. This genome fraction is vitally important and any mutation in it is eliminated by natural selection. Along with this, the variable part of a genome is functionally less important. Therefore, variability, so-called genetic polymorphism, is the minor variability, connected with the adaptation processes and accommodation to certain environmental conditions. Levels of individual heterozygosis in representatives of the same or different populations can considerably differ [32]. In total, the whole genome includes about 4 million of such substitutions (polymorphisms). About 2.5 million of polymorphisms account for sense, coding part of a genome. Genetic polymorphisms spectra depend on geographical conditions, diet, racial (ethnic) belonging, etc. and they are provoked by the natural selection. In certain conditions, they can contribute to development of specific diseases or, on the contrary, inhibit them [33].

Thus, the problem of selection influence on MFD prevalence variation and evolution of its clinical forms is not studied enough and leaves a great number of questions opened.

The modern level of medicine allows to “correct” the phenotypical manifestations of many hereditary pathologies and create an adaptive environment for genotypes, which, because of any adequate therapy absence, would be eliminated by the natural selection, and, thus, to increase their adaptability. In due time, this phenomenon has been called “the dysgenic effect of medicine” [1, 7]. Selection of the polygenic diseases belongs to the intragroup selection. Change of the intragroup selection tension (for example, by successful MFD treatment) causes a change of the distribution curve of susceptibility to this disease and, consequently, increases its prevalence. However, the quantitative estimation of such changes is complicated. From the point of view of population genetics, the hereditary diseases are divided into diseases, associated with impaired reproductive ability, and diseases, where the reproductive ability is not impaired, either because of the defects insignificance, or because they are manifested only after the reproductive period completion. Frequency and prevalence of the first group of diseases are determined by frequency of the corresponding mutations occurrence. The other situation takes place in case of diseases, not affecting the reproduction [1]. This group includes the most widespread MFD, such as DM, schizophrenia, cardiovascular diseases, obesity etc. However, till now, the selection of diseases, not affecting the reproduction, in the conditions of “the dysgenic effect of medicine” is practically not being studied, and the major part of works

is devoted to either the intergroup selection, or the inborn hereditary pathologies selection [34-36].

DM both 1, and 2 types are polygenic MFD [37-42]. According to J.V. Neel hypothesis [28], type 2 DM occurs as a result of special genes existence (thrifty genes), allowing men to use effectively limited food resources, which lead to the disease development in the abundance of food conditions. Thus, if this was the case, type 1 DM was a sub-lethal trait before the insulin therapy introduction into the health care practice. This type of DM is an autoimmune disease, in the course of which the acute lymphocytic insult leads to pancreatic β -cells destruction with subsequent development of the absolute insulin insufficiency [43]. Beginning at pre-pubertal or pubertal age, not supported by the insulin therapy, resulted with the lethal outcome approximately in 0.5-1 year after manifestation. Thus, possibility to leave any genotype was practically reduced to zero. Global implementation of the insulin therapy at the end of 40th – beginning of 50th years into the health care practice has prolonged patients' life, allowed to keep genotype, which resulted in weakening of the selection tension and led to the disease prevalence increase within the population.

Nowadays, a rapid growth of type 2 DM prevalence [44] becomes perceptible worldwide. Besides the environmental factors, promoting growth of this disease prevalence, most likely, the increase of population frequency of gene complexes of susceptibility to this type of diabetes is also necessary for such increasing of type 2 DM frequency in populations. In addition, there should be noted the ubiquitous spreading of type 2 DM severe clinical variant with development of the absolute insulin insufficiency (AID), caused by depletion of residual secretory function of pancreatic β -cells and their subsequent apoptosis [45]. According to the literature, approximately 40% of patients with type 2 DM require the insulin therapy (IT) [46]. There was shown an increased family accumulation of type 2 DM in patients with type 2 DM with AID, and the genetic analysis results demonstrated that the long non-insulin-dependent (LSDID) form can formally be considered as “a less burdened” one, and type 2 DM with AID - as “a more burdened” one [47].

Is it also impossible to ignore the considerable prevalence among the clinical forms of DM variants (2-12% of all the cases) of such its form as the latent autoimmune diabetes of the adults – LADA, whose manifestation is similar to type 2 DM. The subsequent slow development of autoimmune aggression against the pancreatic β -cells, insulin-dependence and necessity of the insulin therapy became the foundation for distinguishing of this disease form as the type 1 DM subtype in classification of 1999 [48]. On the basis of the investigation of patients with the determined diagnosis of type 2 DM and availability of the auto antibodies against GAD, several authors concluded, that this form of DM has an autoimmune nature, but differs from type 1 DM by the rate of the absolute insulin dependence development and by prevalence of alleles, predisposed to type 1 DM [49, 50]. Genetic analysis of this DM form, which was carried out by us, has shown its genetic independence and considerable number of common genes in determination of clinical variants of the disease course (the % of common genes with type 1 and 2 DM was 65.3 and 66.1%, respectively) [51]. Such heterogeneity of DDM clinical variants of DM course can be the consequence of the global implementation of insulin into the health care practice, which sharply reduces the selection tension against type 1 DM that led to increasing of prevalence of gene complexes of susceptibility to the 1th type of disease within the population.

The aim of the investigation was to study the selection influence on heterogeneity of DM clinical forms and the prevalence dynamics within the population.

MATERIALS AND METHODS

To accomplish the assigned tasks, the obstetric anamnesis and presence of polymorphisms of C1858T of *PTPN22*, 49A/G of *CTLA4* and E23K *KCNJ11* genes were studied in DM patients, treated in V. Danilevsky Institute for endocrine pathology problems in 1997 – 2014.

The obstetric anamnesis data of 2106 women after 45 without any symptoms of the investigated diseases were used as a control group. The obstetric anamnesis data are obtained during examinations on different enterprises in Kharkov.

To determine the relative adaptability parameters and to calculate the selection coefficients, the data of obstetric anamnesis were used. The latter included the number of childbirths, pregnancies, spontaneous abortions, extra-uterine pregnancies, survived and deceased geniture till age of 25. It was studied either by questioning or by the hospital histories of the women with completed reproductive period (aged after 45). The number of the investigated women is represented in Table 1.

The determination of C1858T polymorphism of *PTPN22* gene was carried out on 85 patients with type 1 DM, 140 patients with type 2 DM, 115 persons with LADA and 11 healthy Kharkov citizens. The data on C1858T of *PTPN22* gene polymorphism of 242 healthy Kharkov citizens and 296 patients with type 1 DM, received as a result of the inspection in 2004, was kindly submitted for the further analysis by the research worker of Institute of parasitology and biomedicine (Granada, Spain), Maria Ivanovna Fedets. The author expresses a deep gratitude to M.I. Fedets for the data provided. The analysis results were published in the papers of M.I. Fedets as a co-author. Determination of 49A/G polymorphism of *CTLA4* gene was carried out on 64 patients with type 1 DM, 127 patients with type 2 DM, 109 persons with LADA and 75 healthy Kharkov citizens. Determination of E23K polymorphism of *KCNJ11* gene was conducted on 47 patients with type 1 DM, 101 patients with type 2 DM, 83 persons with LADA and 44 healthy Kharkov citizens. Characteristic of the investigated persons is presented in the Table 2.

The “accumulated” population frequency of type 1 and 2 DM in Kharkov region was calculated on the basis of data about patient’s age at DM beginning and a proband age, contained in 1303 histories (857 – type 2 DM, 446 – type 1 DM), available in the district hospitals, as well information about the overall population of the district [52].

Table 1. Characteristics of the women with the investigated obstetric anamnesis

	Control	Type 1 DM	Type 2 DM
N	2106	281	538
Age (years)	57.47 ± 0.16	52.52 ± 0.53	60.62 ± 0.46
Age of the DM beginning (years)	-	20.79 ± 1.84	43.51 ± 0.50
BMI, kg/m ²	24.08 ± 0.12	23.58 ± 0.36	30.64 ± 0.60
WHR	0.76 ± 0.02	0.71 ± 0.06	0.90 ± 0.01

Table 2. Characteristics of persons, whose C1858T of *PTPN22*, 49A/G of *CTLA4* and E23K of *KCNJ11* genes polymorphisms were analyzed

	Control	Type 1 DM	Type 2 DM	LADA
Age (years)	35.40 ± 0.90	37.10 ± 0.70	53.94 ± 0.47	52.19 ± 1.45
Age of the DM beginning (years)	-	23.15 ± 0.94	44.53 ± 0.51	45.80 ± 1.34
Age of the insulin therapy beginning (years)	-	23.15 ± 0.94	53.14 ± 0.81	48.06 ± 1.37
Duration of effective per oral anti-hyperglycemic therapy (years)	-	-	10.48 ± 0.43	2.57 ± 0.32

Sampling was formed according to the initial visit in 2007. The indexes of «accumulated» population frequency of type 1 and 2 DM of 1973 were estimated based on the data about age of DM beginning from 682 clinical records (402 type 2 DM and 280 – type 1 DM), received from archive of Kharkov antioiter dispensary during 1973. The data from 213 hospital histories, which were also received from archive of Kharkov antioiter dispensary, were used for calculation of this index for type 1 DM in 1995. The sampling was formed according to the initial visit of Kharkov antioitrogenic clinic in 1995. The structure of the clinical variants of type 2 DM course both in 1984, and in 2007 was determined from the hospital histories data, received from the archive of Kharkov antioiter dispensary and V. Danilevsky Institute of endocrine pathology problems. The dynamics of DM prevalence was estimated according to the official statistics data [52-58].

Relative adaptability (w) and selection coefficient (s) were calculated for selection direction determination [2]. Thus, relative adaptability (w) was estimated as a result of fertility and survival rate, and $s = 1 - w$. The selection direction was determined as difference between selection coefficients within the population and among the patients ($\Delta s = S_{\text{population}} - S_{\text{patients}}$).

The fertility component was determined by calculation of an average geniture number per one woman, for ill and healthy women, with the subsequent division of an average geniture number per each group by a greater average geniture number in comparison groups.

Survivability component was determined by calculation of the geniture part that survived till 25 in the comparison groups with subsequent division of the survived geniture part for each group on a greater part of the survived geniture in comparison groups.

Values of «accumulated» population frequency were calculated according to [59] on the following formula:

$$q_g = q_0 + q_n \prod_{i=0}^{i=n-1} (1 - q_i),$$

where $q_0, q_1, q_2, \dots, q_n$ – values of morbidity accordingly in the initial (q_0), 1, 2, ...n age-dependent interval.

$$q_n = \alpha_{n+s} p_{n+s}$$

where α_{n+s} is a part of persons, who have got ill at the age of n among the patients, whose age was $n+10$ years at the moment of research; p_{n+s} is the disease prevalence among the persons of age of $n+10$.

DNA extraction was conducted on whole blood, using DNA extraction set “DNK-sorb-V” kit (Moscow, the Russian Federation).

C1858T polymorphism of *PTPN22* gene is determined by the polymerase chain reaction using *RsaI* restrictase [60]. A 218-bp fragment containing the single nucleotide polymorphism C1858T of *PTPN22* gene was amplified using direct ACTGATAATGTTGCTTCAACGG and reverse TCACCAGCTTCCTCAACCAC primers. The reaction mixture contained 20 ng of genomic DNA, 1.2 µl of 10x PCR buffer, 1.2 µl of dNTP (1.25 mmol/l), 0.3 µl of each primer (20 pmol/µl), 0.6 µl of DMSO, and 0.1 µl of *Taq* polymerase (“Sibenzyme” company) in a 12-µl reaction mixture. Amplification conditions: initial denaturation for 2 min at 94°C, followed by 35 cycles of 94°C for 30 s, 30 s at 60°C, and 30 s at 72°C and an additional extension of 2 min at 72°C. Amplification product (12 µl) was incubated using 10 units of *RsaI* restrictase (“Sibenzyme” company) at temperature 37°C for 12 h. Restricted products were electrophoresed on 2.5% agarose gel. The mutant 1858T allele loses the restriction sequence and consists of one 218 bp fragment. 1858C allele restriction results in two fragments of 176 bp and 46 bp.

A49G polymorphism of *CTLA4* gene was amplified with the help of direct 5'-GCTCTACTTCCTGAAGACCT and reverse AGTCTCACTCACCTTTGCAG primers. The reaction mixture contained 25 ng of genomic DNA, 2.5 µl of 10x PCR buffer, 0.6 µl of dNTP (2.5 mmol/l), 0.3 µl of each primer (20 pmol/l) and 0.2 µl of *Taq* polymerase (“Sibenzyme” company) in 25 µl of the reaction mixture. Amplification conditions: initial denaturation for 4 min at 94°C, the following steps 58°C - 45 s, 45 s at 72°C, and 45 s at 94°C (30 cycles) and the final extension for 4 min at 72°C. Amplification product (10 µl) was incubated with the restriction enzyme *BbvI* at 65°C for 1 h. Normal allele 49 A loses the restriction sequence and consists of a 162 bp fragment, the mutant 49G allele restriction produces 88 bp and 74 bp fragments [61].

E23K polymorphism of *KCNJ11* gene was amplified with the help of direct GAATACGTCCTGACACGCCT and reverse GCCAGCTGCACAGGAAGGACAT primers. The reaction mixture contained 25 ng of genomic DNA, 2.5 µl of 10x PCR buffer, 0.6 µl of dNTP (2.5 mmol/l), 0.3 µl of each primer (20 pmol/l) and 0.2 µl of *Taq* polymerase (“Sibenzyme” company) in 25 µl of the reaction mixture. Amplification conditions: initial denaturation for 3 min at 95°C, the following steps 95°C – 1 min, 1 min at 62°C, and 1 min at 72°C (35 cycles) and the final extension for 5 min at 72°C. Amplification product (10 µl) was incubated with the restrictase *Ban II* at 37°C for 2 h and 65°C for 10 min. The mutant 23K allele loses the restriction sequence and consists of a 218 bp fragment. 23E allele restriction results in 178 bp and 40 bp fragments [62]. Evaluation of relative risk of the disease development at polymorphism carriage (Odds ratio) was conducted according to [63].

Statistical analysis. Data are expressed as either mean±SD or percentages. The normality of the distribution of variables was tested with the Kolmogorov-Smirnov test. Comparison of variables between the control and diseases groups was carried out with the Student's test or the chi square test. Comparison of percentage variables was carried out with the Fisher test. A p-value equal to or less than 0.05 was considered to be statistically significant.

RESULTS

The obstetric anamnesis data, which formed a basis for calculation of relative adaptability indexes, are presented in Table 3. The compared groups of healthy and type 1 DM women differed by number of pregnancies. With the similar women behaviour related to the family planning, this fact might be an evidence of selection effect on the impregnation stage. Unfortunately, this question in the investigated groups remained opened, that does not allow to make an unambiguous conclusion about difference in influence of natural selection among healthy women and patients with type 1 DM during impregnation. Number of pregnancies in healthy women significantly exceeds those of type 1 DM, though the average number of childbirth is approximately equal within the population and the investigated groups. This phenomenon can be explained by the fact that women with this diagnosis, knowing about the problems, associated with pregnancy with type 1 DM, are doing their best to maintain the pregnancy.

The values of spontaneous abortions and extra-uterine pregnancies in the compared groups had no significant differences. The obtained results can prove that on the embryogenesis stage there was no selection influence on 1 and 2 type DM. Women distribution by number of childbirths is presented in the Table 4.

The obtained data show the significant distribution difference in childbirths number of healthy women and women with DM. Thus, the significant difference in distribution of childbirths number between women with type 1 and 2 DM is also noted ($\chi^2 = 15.319$; $df = 9$; $p = 0.004$). It is shown that number of women among patients with type 1 DM, who had two and more childbirths, was significantly higher, than of the healthy ones ($\chi^2 = 34.770$; $df = 1$; $p = 0.000$) – 45.68 and 42.40%, respectively. Analogous result was observed also at type 2 DM, ($\chi^2 = 28.678$; $df = 1$; $p = 0.000$) – number of type 2 DM patients and healthy women who had two and more childbirths, was 55.39 and 42.40%, accordingly. According to Table 3, patients with type 2 DM left more children, than the healthy ones and woman with type 1 DM. It should be noted, that women, who have deceased till 45, were excluded from the sampling of patients with type 1 DM because it could overstate some indexes of average number of childbirths in this group.

Table 3. Fertility indexes and geniture survival rate

Parameter	Population	Type 1 DM	Type 2 DM
Pregnancy, ($\bar{x} \pm S_x$)	4.06 ± 0.07	$2.80 \pm 0.17^*$	4.01 ± 0.11
Childbirth, ($\bar{x} \pm S_x$)	1.41 ± 0.02	1.47 ± 0.06	1.59 ± 0.04
Spontaneous abortions, ($\bar{x} \pm S_x$)	0.08 ± 0.01	0.12 ± 0.02	0.10 ± 0.02
Extra-uterine pregnancies, ($\bar{x} \pm S_x$)	0.03 ± 0.01	0.03 ± 0.01	0.03 ± 0.01
Childless women, (%)	12.35 ± 0.72	13.88 ± 2.07	10.24 ± 1.31
Children, survived till 25, (p_s)	0.972	0.877*	0.966
Children, deceased till 25 (p_s)	0.028	0.123*	0.034

Remark* - significance of differences in comparison to the population ($p < 0.001$).

Table 4. Women distribution by number of childbirths

Childbirth number	Control, n = 2106	Type 1 DM, n = 281	Type 2 DM, n = 538
0	260	39	55
1	953	113	185
2	735	96	243
3	125	23	44
4	17	9	8
5	6	1	0
6	4	0	1
7	4	0	1
8	0	0	1
11	1	0	0
χ^2		18.236	36.746
P		0.001	0.000

Analysis of the geniture survival rate in healthy women and mothers with type 2 DM did not demonstrate any high mortality among patients' geniture, being a little lower than in children of healthy women (Table 3). At the same time, analysis of the geniture survival rate in women with type 1 DM and healthy women registered an increased mortality among the geniture of women with type 1 DM in comparison to the children of the healthy women. The obtained data are consonant with the data on increased perinatal mortality in newborns, whose mothers had DM (without diabetes types separation), presented in literature [64, 65].

Indexes of relative adaptability and selection coefficients, calculated based on the obstetric anamnesis, are shown in Table 5. The presented results demonstrate the relative adaptability reduction in the group of patients with type 1 DM.

Table 5. Relative adaptability and selection coefficients

Group	Relative adaptability (w)			Selection coefficient (s)
	Fertility component	Survival rate component	Total adaptability	
Population	0.959	1.000	0.959	0.041
Type 1 DM	1.000	0.901	0.901*	0.099*
Population	0.887	1.000	0.887	0.113
Type 2 DM	1.000	0.988	0.988	0.012*

Remark* - significance of differences in comparison to population ($p < 0.001$).

Selection coefficient value in type 1 DM group is more than twice higher, than among the healthy women of Kharkov ($F = 13.22$; $p < 0.001$), that shows existence of selection against ($\Delta s = -0.058$) this DM type nowadays (Figure 1).

Due to global application of the insulin therapy into health care practice since the beginning of the 50th years of the 20th century, the selection tension against type 1 DM was reduced. For rather short period of time, it has decreased practically from 100.0% (patients left no geniture) to 9.9% that, certainly, had to affect the population frequency of the disease.

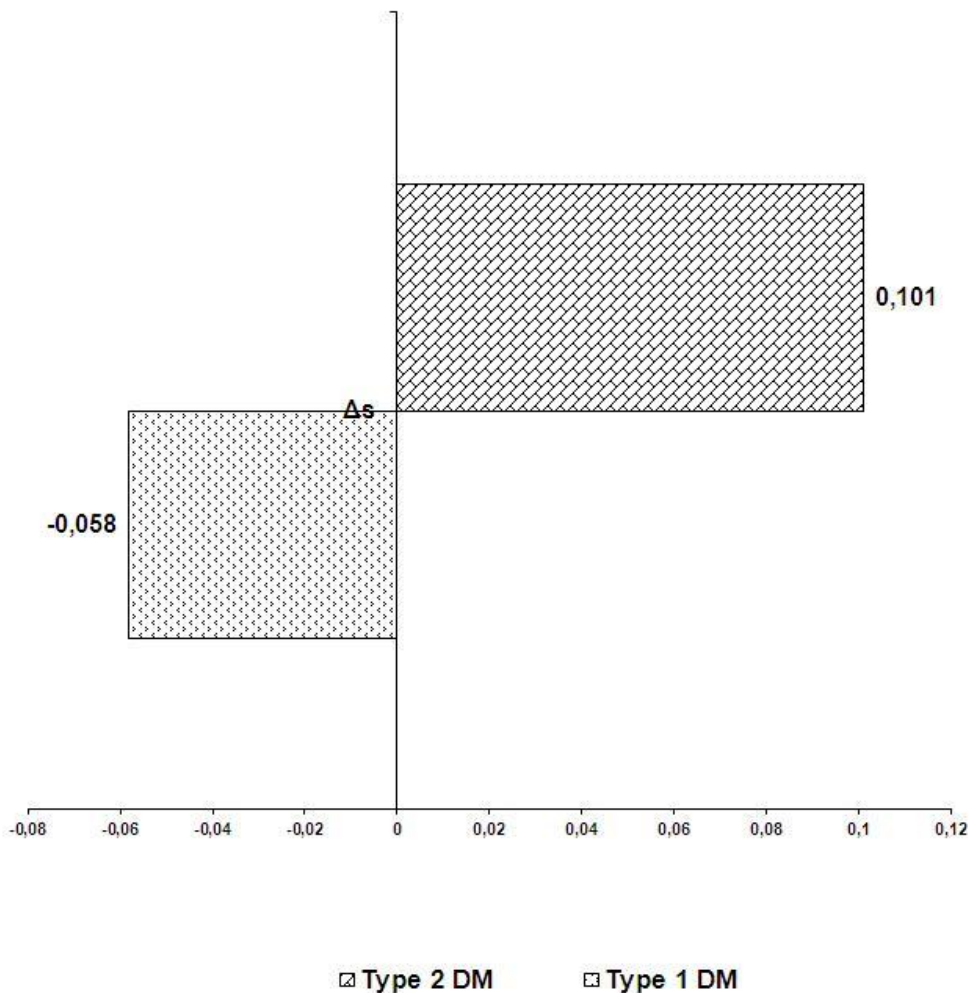


Figure 1. Selection direction of type 1 and 2 DM.

Changes of type 1 DM prevalence among the population of Kharkov area from 1973 to 2007 are presented on Figure. 2.

As expected, because of global implementation of insulin into the therapy of 1 type disease in the 50th years of the last century, approximately one generation later (from 1973 to 1995) the prevalence of type 1 DM has grown – from 0.16 to 0.32% in Kharkov region ($F = 1633.47$; $p < 0.001$). From 1995 till 2007 growth of this disease type prevalence was practically stopped and was 0.32-0.36%. And in recent years from 2009 to 2013, the decrease in population prevalence of similar disease from 0.36 to 0.26% is even noted. It can be explained by the fact that growth of genes frequencies of susceptibility to 1 type DM, caused by the insulin therapy introduction, reached its maximum and now the reduced adaptability of this trait leads to decreasing of its population frequency due to low genotype survival rate.

It should be noted, that in official statistical collections, which data were used in our work, LADA is considered as a variant of type 1 DM and number of patients with this form is summed up with number of patients with a classical type 1 DM. It can also influence the indexes of type 1 DM prevalence.

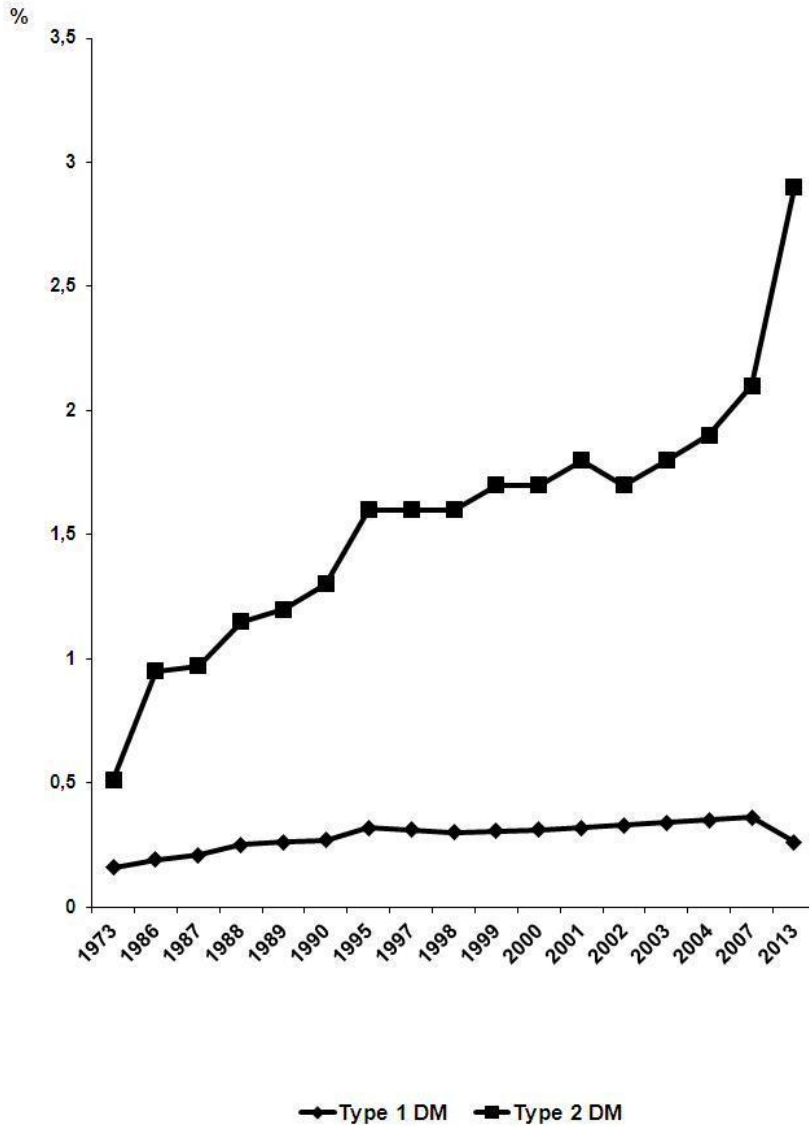


Figure 2. Type 1 and 2 DM prevalence in population of the Kharkov region in 1973-2013.

At calculation of prevalence, as the relation of patients number to the population number without accounting the demographic population structure, the prevalence increase can be caused by the population aging, because in the older age groups, as it was before, during the pre-insulin era, the previous cases are accumulated. For a more exact characteristic of dynamics of disease prevalence within the population, the age disease estimates are used, based on which «the accumulated population frequency» is calculated. To estimate the dynamics of DM occurrence probability during life, in population of Kharkov region, the indexes of «accumulated population frequency» were calculated for 1973, 1995 and 2007. The age indexes of «accumulated population frequency» of type 1 and 2 DM among the population of Kharkov region are presented on the Figure 3.

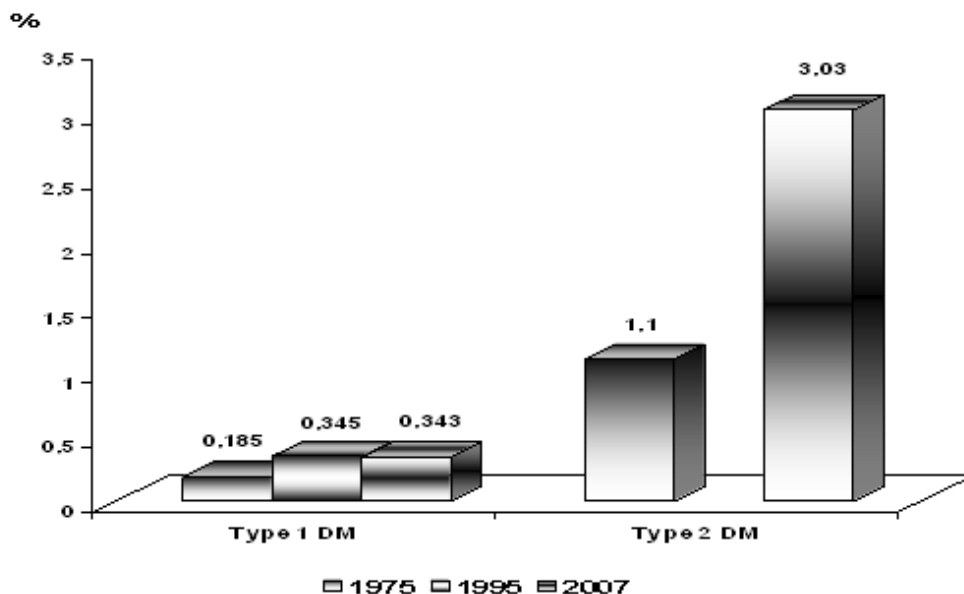


Figure 3. Indexes of «accumulated population frequency» of type 1 and 2 DM among Kharkov region citizens, %.

For type 1 DM, some growth of «accumulated population frequency» from 1973 to 1995 by 1.86 time is shown ($F = 0.116$; $p > 0.05$), and even its small decline from 1995 to 2007 from 0.345 to 0.343% ($F = 0.0001$; $p > 0.05$) is shown.

Existent selection against type 1 DM gives the grounds to assume that it is not necessary to expect a rapid growth of type 1 DM prevalence in populations.

Data, provided in Table 5 indicate the higher relative adaptability of patients with type 2 DM in comparison with a healthy phenotype. The value of selection coefficient in type 2 DM was much lower, than in the healthy women of Kharkov region ($F = 93.06$; $p < 0.001$). Thus, the conducted research shows that a positive orientation of type 2 DM selection (Figure 1) takes place ($\Delta s = 0.101$) – patients leave more geniture, than the healthy people, at their practically identical survival rate (Tables 3, 4). This, in its turn, increases the frequency of susceptibility to this type of disease genes in the population.

The obtained data are confirmed by the evidence on a significant increase (almost four times) of type 2 DM prevalence in population of Kharkov region ($F = 31518.6$; $p < 0.001$) for thirty four years in 1973-2007 (Figure 2.).

From 1973 to 2007 the «accumulated population frequency» of type 2 DM increased in 2.75 times ($F = 5.21$; $p < 0.05$). Therefore, the selection in favor of type 2 DM allows to forecast the further growth of this type of disease prevalence in populations and, from genetics point of view, explains the reasons, promoting a rapid growth of this type DM prevalence in the world.

The increase of frequency of both 1 and 2 type DM susceptibility genes in the population, caused by a positive selection direction (for type 1 DM to the middle of 90th years of the 20th century), can also be reason, causing the clinical heterogeneity of this disease.

Such aspect of selection action is confirmed both by the features of genetic heterogeneity of type 2 DM, and dynamics of clinical forms structure in all cases of type 2 DM. Recently, ubiquitous growth of specific severity of the disease clinical course with development of AID

among all patients with type 2 DM [46] is noted. So, compared to 1984, the number of patients with AID among all patients with type 2 DM were up to 22.65% in Kharkov region, in 2007 this index has grown to 33.48% ($\chi^2 = 10.729$; $p = 0.001$). The conducted researches [47] have shown that out of two studied forms of type 2 DM, one is more complicated. According to the genetic analyses results, the LSNID form can possibly be formally considered as “a less complicated,” and type 2 DM with AID - as “a more complicated.”

Hence, the population frequency increase of genes of susceptibility to type 2 DM, caused by a positive selection direction, raises probability of their high concentration in a proband, and thus promoting the increase of AID prevalence among all cases of the disease.

The frequency increase of genes of susceptibility to 1 type disease became the possible effect of selection tension decrease against type 1 DM, caused by introduction of insulin therapy into the health care practice. Combination of population frequency growth of genes of susceptibility to type 1 and 2 DM and increase in probability of their occurrence in a proband has defined appearance of such DM form, as LADA, among all variants of the disease course. This is evidenced by the genetic analysis results, which have showed that inheritance of LADA is described by the polygenic threshold model parameters, in its inheritance the essential role belongs to the genetic factors, there are nonlinear genetic interactions, and influence of genes number is possible with the expressed effect in determination of this DM form. Existence of a large number of common genes of susceptibility to type 1 and 2 DM and LADA (65.3 and 66.1%, respectively) defines the features of clinical course of this DM form - manifestation is similar to type 2 DM: the subsequent fast development of autoimmune aggression to the pancreatic β -cells, insulin-dependence and insulin therapy necessity [51].

Taking into account, that modern level of researches embraces not only the genetic analysis data, but also molecular genetics methods, we studied the distribution features of the single nucleotide polymorphisms genes, determining the susceptibility to both type 1 and 2 DM. Taking into consideration, that the HLA system antigens, which generally determine susceptibility to type 1 DM, are strongly linked and badly combined, it was reasonable to study the immune response genes polymorphisms, which play the secondary role in type 1 DM development, localized on different chromosomes and which are well combined in the population in the process of crossing. As an example of such polymorphisms, distributions of mutations of C1858T of *PTPN22* gene and 49A/G of *CTLA4* gene in the patients with type 1 and 2 DM, LADA and healthy Kharkov citizens were studied.

The gene *PTPN22* (protein tyrosine phosphatase non-receptor type 22) is located on the 1p13.3-p13.1 chromosome. This gene encodes the lymphoid-specific LYP phosphatase, which suppress the T-lymphocytes activation. Two variants of *PTPN22* gene - C1858 and T1858 - differ by an important part of amino-acid sequence, which is responsible for LYP association with Csk-kinase (negative regulation). When in the 620th position (T1858 allele) arginine is replaced by tryptophan, the connection with Csk-kinase is not provided. It leads to T-lymphocytes signal function disorder and promotes the development of autoimmune diseases susceptibility, in particular, to type 1 DM [66]. Today it is known, that C1858T polymorphism of *PTPN22* gene is associated with type 1 DM in many world populations: in Ukrainian [67], Italian [68], Swedish and Finnish [69].

The *CTLA4* (cytotoxic T-lymphocyte associated antigen-4) gene also defines the susceptibility to type 1 DM. It is localized on the chromosome 2q33, between two T-lymphocytes genes-activators: the receptor activator gene (CD28) and gene inducing the costimulant (ICOS), contains 4 exons and three introns. The receptor isoform (full-length

isoform-flCTLA4), synthesized in the activated T-lymphocytes is coded by 4 exons: the leader protein is coded by the exon 1, the ligand-binding protein – by the exon 2, the membrane-spanning domain – by the exon 3 and the cytoplasmic domain – by the exon 4. More than 30 points of the mono-nucleotide substitutions in different areas of *CTLA4* gene are known. One of important mono-nucleotide substitutions is a single nucleotide polymorphism 49A/G in the first exon (replacement of adenine by guanine in the 49th position causes replacement of threonine by alanine in the 17th codon of amino-acid sequence of the leader peptide), leading to decrease of CTLA4 protein functional activity. It is established, that the protein product of *CTLA4* gene takes part in T-lymphocytes activity regulation and plays the important role in the autoimmune processes development. [70]. Researches of different populations have revealed both availability [71-76] and absence of [77-79] the associations of type 1 DM with 49A/G polymorphism of *CTLA4* gene.

Among the single polymorphisms, determining the development of susceptibility to type 2 DM, it was logical to investigate the mutation, which is responsible for development of insulin insufficiency in a person. Polymorphism of E23K of *KCNJ11* gene belongs to such mutations. *KCNJ11* gene is localized on the 11th chromosome in area of p15.1. This gene encodes synthesis of Kir6.2 protein (Potassium inward rectifier 6.2), which is a part of the potassium channel in cells, capable to activation, and creates pores for transportation of potassium ions with the cells. The channel closing is necessary for insulin secretion by β -cells. Majority of mutations of this gene are missens-mutations, which have the dominant-negative effect, leading to decreasing of IK1 stream, reduction of repolarization and increase of action potential duration. This polymorphism takes part in the insulin insufficiency development in patients with type 2 DM and is associated with its development in various populations [80].

Table 6. Frequencies of genotypes alleles of C1858T polymorphism of *PTPN22* gene in DM patients and healthy Kharkov citizens

Index	Control, N = 253		Patients		χ^2	Significance of differences (p)	OR (95% CI)
	N	%	n	%			
Type 1 DM, N = 381							
Genotype: C/C	185	73.10	228	59.84	11.231	0.000	0.55 (0.55-1.09)
C/T	66	26.10	123	32.28	2.502	0.114	1.35 (0.08-1.62)
T/T	2	0.80	30	7.87	13.674	0.000	10.73 (0.73-11.83)
Type 2 DM, N = 140							
Genotype: C/C	185	73.10	79	57.30	9.230	0.002	0.48 (0.47-1.12)
C/T	66	26.10	52	35.88	3.527	0.060	1.67 (0.80-1.95)
T/T	2	0.80	9	6.90	9.385	0.002	8.62 (0.54-11.97)
LADA, N = 115							
Genotype: C/C	185	73.10	54	47.30	21.479	0.000	0.33 (0.39-0.97)
C/T	66	26.10	43	37.30	4.092	0.043	1.69 (0.78-2.01)
T/T	2	0.80	18	15.5	30.346	0.000	23.29 (8.30-46.57)

Table 7. Distinctions in frequencies of genotypes and alleles of C1858T polymorphism of *PTPN22* gene in DM patients

Compared groups	Index	χ^2	Significance of differences (p)
Type 1 DM – type 2 DM	Genotype: C/C	0.362	0.547
	C/T	0.877	0.349
	T/T	0.150	0.698
Type 1 DM - LADA	Genotype: C/C	5.466	0.019
	C/T	0.818	0.366
	T/T	5.257	0.022
LADA – Type 2 DM	Genotype: C/C	1.906	0.167
	C/T	0.009	0.922
	T/T	4.741	0.029

The comparative analysis of genotypes distribution with C1858T of *PTPN22* gene polymorphism among all compared groups (Table 6) revealed the statistically significant difference in distribution of healthy individuals and patients with LADA, type 1 and 2 DM by genotypes with C1858T of *PTPN22* gene - $\chi^2 = 47.063$, $p = 0.000$.

Investigation of distribution of C1858T genotypes of *PTPN22* gene has shown the significant homozygotes association of this polymorphism with LADA, type 1 and 2 DM. It should be noted, that homozygous carriers of this polymorphism occur significantly more often among patient with LADA than among the ones with 1 and 2 type DM. Significant distinctions in genotypes frequencies between the patients with type 1 and 2 DM was not revealed (Table 7).

Thus, the investigation of association of single nucleotide of C1858T polymorphism of *PTPN22* gene has shown its evident association with LADA, type 1 and 2 DM.

The obtained data overlap with existing works on this polymorphism investigation in populations of Ukraine, the USA, Poland, Finland, Sweden, Norway and others [67-69]. For example, in Sweden populations [69] the association of this polymorphism with all DM clinical forms was shown: frequencies of carriage of mutant homozygote T/T in patients with type 1, 2 DM, LADA and in control were 0.3, 3.8, 2.8 and 4.3% accordingly. The most expressed association of this polymorphism was observed with LADA. The investigation results allow to assume that development of such DM form as LADA is caused more by changes in the genes, controlling normal immunological homeostasis, and that determines the features of course of this disease form even in all genes of HLA system absence, which are necessary for type 1 DM development.

Additionally, we studied the distribution of polymorphism of 49A/G of *CTLA4* gene, which causes the immune disorders, in patients with different clinical variants of DM.

The results obtained are presented in Table 8.

These results are concordant with the data about 49A/G of *CTLA4* gene polymorphism association with type 1 DM in different populations of the world [240]. To determine the frequency of this polymorphism in different clinical variants of DM course in Kharkov population, we carried out the comparative analysis of frequencies of 49A/G *CTLA4* gene polymorphism among the patients with type 1, 2 DM and LADA.

Table 8. The frequencies of genotypes with 49A/G *CTLA4* gene polymorphism in DM patients and healthy Kharkov citizens

Index	Control, N = 75		Patients		χ^2	Significance of differences (p)	OR (95% CI)
	N	%	n	%			
Type 1 DM, N = 64							
Genotype: A/A	29	38.67	12	18.75	5.644	0.014	0.37 (0.31-1.41)
A/G	34	45.33	29	45.31	0.028	0.866	1.00 (0.51-1.95)
G/G	12	16.00	23	35.94	6.266	0.012	2.95 (0.72-3.56)
Type 2 DM, N = 127							
Genotype: A/A	29	38.67	26	20.47	6.986	0.008	0.41 (0.36-1.28)
A/G	34	45.33	63	49.61	0.195	0.659	1.19 (0.61-1.91)
G/G	12	16.00	38	29.92	4.187	0.041	2.24 (0.69-2.93)
LADA, N = 109							
Genotype: A/A	29	38.67	22	20.18	6.681	0.010	0.40 (0.35-1.30)
A/G	34	45.33	41	37.61	0.800	0.371	0.73 (0.48-1.58)
G/G	12	16.00	46	42.20	12.943	0.000	3.83 (0.87-3.89)

The genotypes distribution analysis among all compared groups has shown the reliable difference in patients with LADA, type 1 and 2 DM in comparison with the healthy Kharkov citizens ($\chi^2 = 17.853$; $df = 6$, $p = 0.001$).

There was shown a significant association of G/G mutant homozygotes with type 1, 2 DM and LADA ($OR_{type\ 1\ DM} = 5.61$; $OR_{type\ 2\ DM} = 4.27$; $OR_{LADA} = 7.30$). It should also be noted, that homozygous carriers of 49A/G of *CTLA4* gene polymorphism appeared more often among the patients with LADA than among the patients with type 2 DM (Table 9).

No significant distinctions in the genotypes frequencies in patients with type 1 and 2 DM are revealed. The comparative analysis results of E23K polymorphism of *KCNJ11* gene distribution, participating in development of insulin deficiency and determining the type 2 DM susceptibility, among the patients with LADA, type 1 and 2 DM are given in Table 10.

Table 9. Distinctions in frequencies of genotypes and polymorphism alleles of 49A/G *CTLA4* gene in DM patients

Compared groups	Index	χ^2	Significance of differences (p)
Type 1 DM – Type 2 DM	Genotype: A/A	0.008	0.929
	A/G	0.166	0.684
	G/G	0.458	0.498
Type 1 DM - LADA	Genotype: A/A	0.001	0.975
	A/G	0.698	0.403
	G/G	0.425	0.515
LADA – Type 2 DM	Genotype: A/A	0.011	0.915
	A/G	2.953	0.086
	G/G	3.834	0.049

Table 10. The frequencies of genotypes and alleles of E23K polymorphism of *KCNJ11* gene in DM patients and healthy Kharkov citizens

Index	Control, N = 44		Patients		χ^2	Significance of differences (p)	OR (95% CI)
	n	%	n	%			
Type 1 DM, N = 47							
Genotype: E/E	32	72.73	8	17.02	26.410	0.000	0.08 (0.02-0.12)
E/K	9	20.45	16	34.04	1.479	0.224	2.01 (0.52-3.50)
K/K	3	6.82	23	48.94	17.743	0.000	13.10 (0.83-15.26)
Type 2 DM, N = 101							
Genotype: E/E	32	72.73	19	18.81	36.744	0.000	0.87 (0.15-1.28)
E/K	9	20.45	33	32.67	1.670	0.196	1.89 (0.57-3.06)
K/K	3	6.82	49	48.51	21.309	0.000	12.88 (2.88-15.44)
LADA, N = 83							
Genotype: E/E	32	72.73	20	24.10	26.150	0.000	0.12 (0.03-0.91)
E/K	9	20.45	20	24.10	0.059	0.808	1.23 (0.45-2.66)
K/K	3	6.82	43	51.82	23.285	0.000	14.69 (3.92-19.20)

Genotypes distribution comparison of E23K of *KCNJ11* gene among compared groups revealed significant association of polymorphic homozygote carriers with type 1, 2 DM and LADA (OR_{type 1 DM} = 13.10; OR_{type 2 DM} = 12.88; OR_{LADA} = 14.69).

Table 11. Distinctions in frequencies of genotypes and alleles of E23K polymorphism of *KCNJ11* gene in DM patients

Compared groups	Index	χ^2	Significance of differences (p)
Type 1 DM – Type 2 DM	Genotype: E/E	0.001	0.973
	E/K	0.001	0.982
	K/K	0.017	0.897
Type 1 DM - LADA	Genotype: E/E	0.519	0.471
	E/K	1.027	0.311
	K/K	0.017	0.895
LADA – Type 2 DM	Genotype: E/E	0.478	0.489
	E/K	1.243	0.265
	K/K	0.088	0.767

It should also be noted, that the homozygous carriers of E23K of *KCNJ11* gene mutation appear more often among the patients with LADA, when comparing to the patients with type 1 and 2 DM (Table 10). There are no significant differences in genotypes frequency among the patients with type 1, 2 DM and LADA (Table 11).

Figure 4 shows distribution of polymorphisms 49A/G of *CTLA4* and C1858T of *PTPN22* genes in patients with different clinical variants of DM course.

The significant distinctions of distribution of genotypes with 49A/G of *CTLA4* and C1858T of *PTPN22* genes ($\chi^2 = 23.485$; $df = 16$; $p = 0.000$) among the compared groups of patients were shown. The pair-wise comparison of DM clinical variants shown the significant distinctions between the patients with LADA and type 2 DM ($\chi^2 = 10.622$; $df = 8$; $p = 0.031$). There was found no significant distinctions in this polymorphisms between the patients with LADA and type 1 DM; and type 1 and 2 DM ($\chi^2 = 4.098$; $df = 8$; $p = 0.393$; $\chi^2 = 3.970$; $df = 8$; $p = 0.410$, accordingly). The frequency of carriage of homozygote GG/TT genotypes in the patients with LADA was significantly higher than such in the patients with type 2 DM (8.30 and 1.60%, accordingly, $\chi^2 = 4.425$, $p = 0.035$).

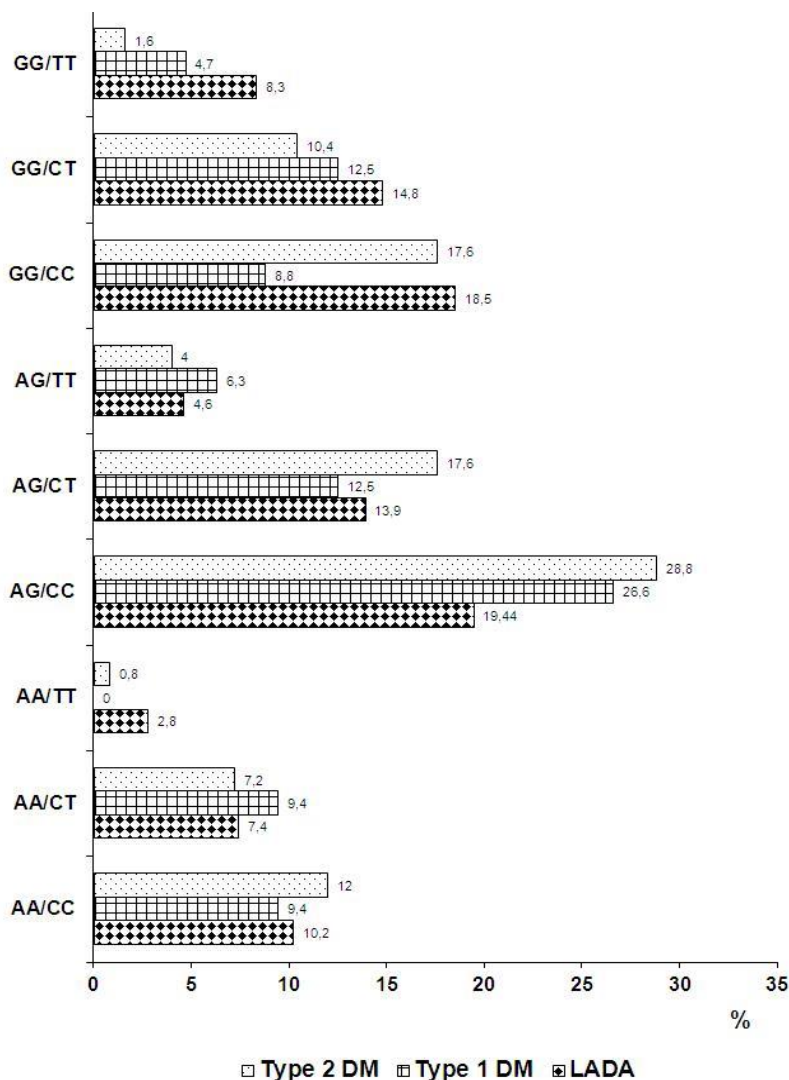


Figure 4. Distribution of polymorphisms 49A/G of *CTLA4* and C1858T of *PTPN22* genes in patients with different clinical variants of DM course.

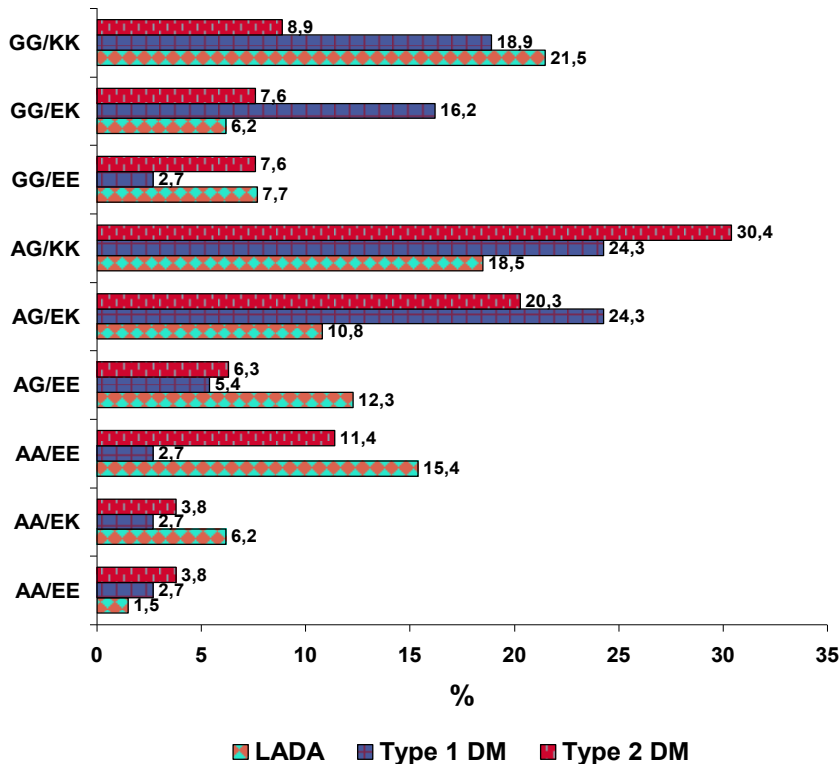


Figure 5. Distribution of 49A/G polymorphisms of *CTLA4* gene and E23K of *KCNJ11* gene in patients with different clinical variants of DM course.

Distribution of patients by carriage of polymorphisms of 49A/G of *CTLA4* gene and E23K of *KCNJ11* gene is presented on Figure 5.

The significant distinctions in distribution of genotypes with 49A/G of *CTLA4* gene and E23K of *KCNJ11* gene ($\chi^2 = 19.978$; $df = 16$; $p = 0.000$) among the compared groups of patients were shown. The pair-wise comparison of DM clinical variants shown the significant distinctions between the patients with LADA and type 1 DM ($\chi^2 = 12.065$; $df = 8$; $p = 0.017$); between the patients with LADA and type 2 DM ($\chi^2 = 10.976$; $df = 8$; $p = 0.027$). The significant distinctions in this polymorphisms between the patients with type 1 and 2 DM were not found ($\chi^2 = 7.859$; $df = 8$; $p = 0.097$). The frequency of carriage of homozygote GG/KK genotypes in patients with LADA was significantly higher than the one in the patients with type 2 DM (21.54 and 8.86%, accordingly, $\chi^2 = 3.840$, $p = 0.050$).

Distribution of patients by carriage of polymorphisms of C1858T of *PTPN22* gene and E23K of *KCNJ11* gene is presented on Figure 6.

The significant distinctions in genotypes distribution with C1858T of *PTPN22* gene and E23K of *KCNJ11* gene ($\chi^2 = 11.146$; $df = 16$; $p = 0.025$) among the compared groups of patients were shown. The pair-wise comparison of DM clinical variants shown the significant distinctions between the patients with LADA and type 2 DM ($\chi^2 = 5.233$; $df = 8$; $p = 0.264$). The significant distinctions in this polymorphisms between the patients with LADA and type 1 DM, also between the patients with type 1 and 2 DM were not found ($\chi^2 = 5.928$; $df = 8$; $p = 0.205$; $\chi^2 = 6.303$; $df = 8$; $p = 0.178$, accordingly).

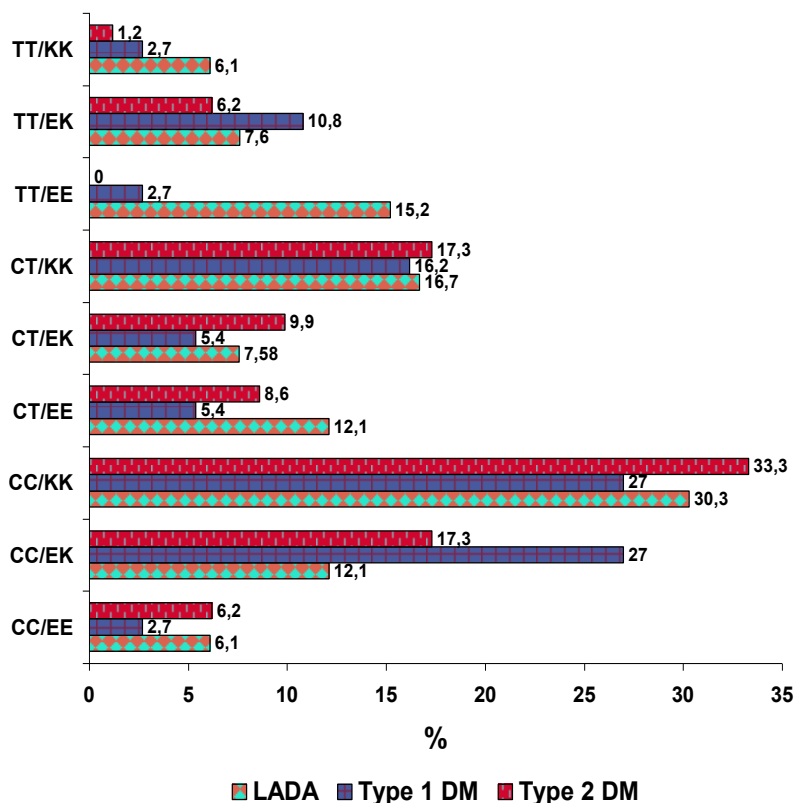


Figure 6. Distribution of polymorphisms of C1858T of *PTPN22* gene and E23K of *KCNJ11* gene in patients with different clinical variants of DM course.

Distribution of patients by carriage of polymorphisms of 49A/G of *CTLA4* gene, C1858T of *PTPN22* gene and E23K of *KCNJ11* gene is presented on Figure 7.

The significant distinctions in distribution of the genotypes with C1858T of *PTPN22* gene and 49A/G of *CTLA4* gene, and E23K of *KCNJ11* gene ($\chi^2 = 56.923$; $df = 46$; $p = 0.000$) among the compared groups of patients were shown. The pair-wise comparison of DM clinical variants shown the significant distinctions between all investigated groups: the patients with LADA and type 1 DM ($\chi^2 = 23.978$; $df = 23$; $p = 0.000$); the patients with LADA and type 2 DM ($\chi^2 = 19.341$; $df = 23$; $p = 0.000$); the patients with type 1 and type 2 DM ($\chi^2 = 21.071$; $df = 20$; $p = 0.000$); that indicates the genetic independence of all these clinical DM forms.

However, if the genetic independence of type 1 and 2 DM is well-proven, the genes of susceptibility to these disease types are selected, a question about LADA genetic determination and reasons for its occurrence remain open to date. It is assumed for today, that LADA is the variant of type 1 DM course [81]. The molecular-genetic research, conducted by us, shown LADA genetic independence, confirming the results of genetic [51] analysis, obtained before.

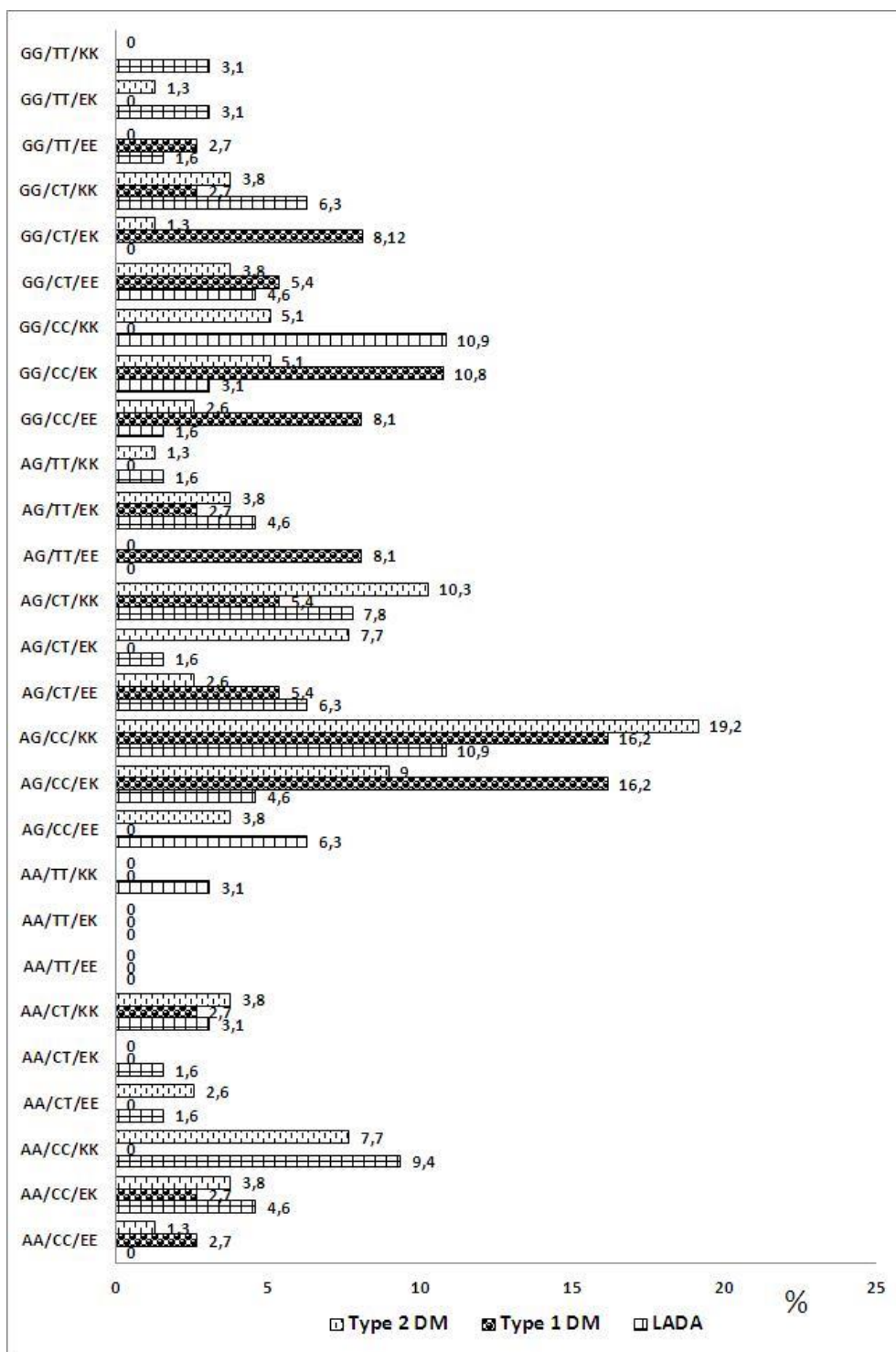


Figure 7. Distribution of polymorphisms of 49A/G of *CTLA4* gene, C1858T of *PTPN22* gene and E23K of *KCNJ11* gene in patients with different clinical variants of DM course.

Absence of significant distinctions among the investigated single gene polymorphisms in the patients with type 1 and 2 DM completely corresponds to the genetic analysis results, which indicated the presence of 59% common genes in type 1 and 2 DM determination [41]. Taking into account the fact that type 1 and 2 DM are polygenetic diseases, not only the genes, which caused one or another DM type development, the genes combination is significant too. Selection tension against type 1 DM reduction, caused by the insulin therapy introduction, and positive direction of genes of susceptibility to type 2 DM selection resulted not only in increase of the genes of susceptibility to type 1 and 2 DM in population, but also in their combination variants increase. It leads, in particular, to growth of such genetically independent DM form as LADA and increases its prevalence among the DM patients.

Thus, the conducted research showed the selection role in changing of prevalence of type 1 and 2 DM in population and in evolution and in origin of DM clinical forms.

DISCUSSION

In this work the peculiarities of different DM types selection, which appear in different ontogenetic periods and differ in morbidity and genetic determination, were studied. Thus, if type 1 DM, without an adequate insulin therapy, results in fatal outcome and, to some extent, can be partly associated with the factors, regulating the population amount, type 2 DM, vice versa, promoted the increase of survival rate of carriers of this trait in the past historical periods with food deficiency. According to J.V. Neel conception, the special genes (thrifty gene), which allowed a man to use effectively the limited food resources, results in obesity and type 2 DM development in the conditions of food abundance [28].

The conducted research showed, that selection in favour of diseases, which commonly start in a post-reproductive period, to which type 2 DM belongs, exists till the present time. The reasons, determining the positive direction of type 2 DM selection, remain unknown till today and, probably, can be of interest for further researches. The selection positive direction was accompanied by significant growth of their prevalence in population over a generation.

Peculiarities of type 1 DM selection were influenced by the modern medicine, which succeeded in compensating the disease with the insulin therapy. This enabled a patient to survive and produce genotype. Type 1 DM in the “pre-insulin period” belonged to the sub-lethal traits. Its manifestation usually occurs in the pre-pubertal or pubertal period, and then during half a year – a year such patients died from the hyperglycemic coma. The insulin therapy allowed to correct the disease course significantly, however, the negative consequences cannot be completely neutralized till today, that changed type 1 DM from sub-lethal trait into the trait of reduced adaptability. The peculiarities of clinical course of type 1 DM result in lower genotype survival rate that determines the low relative adaptability of disease and, accordingly, the negative selection direction.

The insulin therapy introduction in 50th years of 20th century into the common health care practice has led to decrease of selection tension against type 1 DM and increase of the population prevalence of this type of disease. Growth of type 1 DM prevalence proceeded up to 1995, approximately for two generations (fifty years). After 1995 the rapid growth of type 1 DM prevalence has stopped, apparently, the existent selection against type 1 DM tension played a certain role in it. Consequence of the ongoing processes is not only increased prevalence of

type 1 DM in population, but also the increased clinical DM heterogeneity. Increase in frequency of genes of susceptibility to type 1 DM in population, caused by selection tension reduction, was not only the basis for increase of population prevalence of this disease type, but also promoted probability of genes occurrence in proband, determining development of both DM types. The result of this phenomenon could be a formation and growth of prevalence in DM structure of such clinical form of disease as LADA [51], which heritability is described by the polygene threshold model parameters. In its inheritance the essential role belongs to the genetic factors (59.2%), there are nonlinear genetic intraloci and interallelic ($GD = 1.2\%$) interactions and influence of number of genes with evident effect in determination of this form of disease is possible. Although LADA is an independent DM genetic form, but, according to Ch. Smith model testing, there is approximately the same amount of common genes of susceptibility to type 1 and 2 DM (65.3 and 66.1%, accordingly) in its genetic control system.

Conducted researches of polymorphisms of C1858T of *PTPN22* gene and 49A/G of *CTLA4* gene, associated with type 1 DM in patients with different DM clinical forms, show that those polymorphisms are associated with type 2 DM and with LADA. Analogical data on associations of this polymorphisms with type 1 and 2 DM and LADA were obtained also in other populations in different countries of the world [67-72]. It should be noted, that the insulin therapy introduction into the common health care practice for correction of type 1 DM course was observed in all countries of the earth. And such disease compensation allowed to reduce considerably the selection tension against this type of DM in all countries. It increased the frequencies of genes of susceptibility to type 1 DM in all populations. Therefore, the association of polymorphism of C1858T of *PTPN22* gene and type 1 DM is not local, but observed in different countries. The model, presented before, of type 1 DM inheritance [41] indicated the linked factors, determining the hereditary susceptibility to this type of disease (high values of epistatic components). It is presently known, that antigens of the HLA system are the major contributors to type 1 DM development [42]. The HLA system antigens are codominantly inherited and are linked factors [82]. Other genes of susceptibility to type 1 DM, such as *CTLA4*, *PTPN22*, *IL2RA*, *IFIH1*, *IL2*, *PTPN2* and *KIAA0350*, are localized on different chromosomes [42]. Taking into account that these genes are linked considerably less, than antigens of the HLA system, it is logically to assume that *CTLA4*, *PTPN22*, *IL2RA*, *IFIH1*, *IL2*, *PTPN2* and *KIAA0350* genes can form different combinations with the genes of susceptibility to type 2 DM in a proband. From this point of view, it is understandable that LADA patients more frequently than the ones with type 1 DM have such gene polymorphisms as C1858T *PTPN22* and 49A/G *CTLA4* genes. The observed phenomenon corresponds to ideas of numerous authors, that LADA has an autoimmune nature, but differs from type 1 DM by type of insulin dependence and by distribution of susceptibility to DM alleles [49]. Frequency of C1858T of *PTPN22* gene and 49A/G of *CTLA4* gene polymorphisms in type 2 DM is considerably less than in LADA and is analogical to the one in type 1 disease, although carriage of this polymorphisms is not significant in pathogenesis of type 2 DM. It can be explained by selection tension reduction against type 1 DM, caused by the insulin therapy implementation, accompanied by increase of population genes frequency, causing the type 1 DM development. It can logically explain the fact of association of C1858T of *PTPN22* and 49A/G of *CTLA4* genes polymorphisms with type 2 disease, because the selection tension reduction promoted the occurrence probability of the genes of susceptibility to type 1 DM in patients with type 2 DM. However, the presence of one or another genes polymorphism, determining the immune response disorders in patients with type 2 DM, does not cause any autoimmune aggression

against the pancreatic islet cells, as the latter is caused by a combined effect of multiple mutations, causing immune response disorder.

Because the genetic control system LADA, in according the heredity model, has a bigger part of common genes with type 2 DM, it was logical to study an association of LADA, type 1 and 2 DM with polymorphism of E23K of *KCNJ11* gene, which determines the development of insulin insufficiency in a proband which is a gene, taking part in the susceptibility to type 2 DM forming [42]. The obtained results were analogical to our data about the association of polymorphisms of C1858T of *PTPN22* gene and 49A/G *CTLA4* gene and DM. The significant association of this mutation with all clinical variants of DM course (Table 10) was also shown. Thus, the conducted molecular-genetic researches confirmed positions of LADA and type 1 and 2 DM inheritance models about existence of common genes in their genetic control systems.

The distribution of the patients with different clinical variants of DM course by the polymorphisms of C1858T of *PTPN22*, 49A/G of *CTLA4* and E23K of *KCNJ11* genes shown the significant distinctions ($\chi^2 = 56.923$; $df = 46$; $p = 0.000$) in the distribution of patients with different clinical variants of disease course by the genotypes, which were investigated. The pair-wise comparison of patients groups revealed the significance of distribution of patients by the carriage of polymorphisms, which were studied among all investigated groups, that proves the genetic independence of all these DM forms. Conducted molecular-genetic research showed LADA genetic independence, confirming our genetic analysis results [51]. In addition it should be noted, that larger, than among the patients with type 1 and 2 DM, number of homozygote carriers of all three mutant alleles is characteristic for LADA. The obtained data confirm Hamaguchi assumption, that LADA has an autoimmune nature, but differs from type 1 DM by types of the insulin dependence and distribution of alleles, susceptible to DM [49]. And exactly the high frequency of mutant homozygote carriers of immune response genes among the patients with LADA, and only partial availability of the HLA genes system in them, which is forming the susceptibility to type 1 DM, determine late development of this DM form and its less severe course. The high percentage of homozygotic carriers of mutations, forming susceptibility to type 1 and 2 DM, is a bright confirmation of assumption about the selection influence on evolution of DM clinical forms. Exactly the high percent of homozygotic carriage of polymorphic alleles, simultaneously by all three investigated positions in LADA patients, illustrates the position that this form of DM occurs as a result of selection action because of frequencies increase of the genes of susceptibility to both type 1 and 2 DM in population.

CONCLUSION

Thus, natural selection has defined the prevalence dynamics and influences formation of the polygenic diseases heterogeneity, to which DM belongs.

REFERENCES

- [1] F. Vogel, A. G. Motulsky, *Human genetics. Problems and Approaches*, Springer-Verlard, Berlin Heindelberg, New York, Tokio (1986).
- [2] F. J. Ayala, J. A. Kiger, *Modern genetics*, University of California, Davis, Second Edition (1984).
- [3] Y. P. Altukhov, Y. G. Rychkov, *Journal of common biology*, **31**, 507 (1970).
- [4] Shmal'gauzen, *Cybernetic questions of biology*. Science, Novosibirsk (1968).
- [5] Y. P. Altukhov, *Bulletin of genetics and selectors society*, **8**, 40 (2004).
- [6] O. L. Kurbatova, *Materials of the third anthropological reading to the 75year from the day of birth of academician V.P. Alekseeva «Ecology and demography of man in the past and present»* in press (2004) "Demographic genetics of megapopulation: changeability of selection parameters," The third anthropological reading to the 75year from the day of birth of academician V.P.Alekseeva «Ecology and demography of man in the past and present» 15-17, November. 2004 (Moscow, Russia). pp.259-262.
- [7] WF Bodmer and LL Cavalli-Sforza, *Genetics, Evolution, and Man*, WH Freeman and Company, San Francisco (1976).
- [8] N. Kucher, O. L. Kurbatova, *Genetics*, **22**, 304 (1986).
- [9] L. A. Atramentova, L. I. Fedchun and S. A. Povolockiy, *Genetics*, **29**, 520 (1993).
- [10] J. F. Crow, *Human Biol.* **30**, 1 (1958).
- [11] N. P. Dubinin, I.I. Karpec and V.N. Kudryavcev, *Genetics, behaviour, responsibility*, Politizdat, Moscow (1982).
- [12] Y. P. Altukhov, *Genetic processes in populations*, IKC "Akademkniga", Moscow (2003).
- [13] L. S. Penrose, *Proc. Roy. Soc. Ser. B.*, **159**, 93 (1963).
- [14] V. M. Kornac'kiy *Problems of health of society and the life duration*, Kiev (2006).
- [15] O. L. Kurbatova, *Theses of lectures of First (third) Russian congress of medical geneticists* in press (1994). "Some tendencies of transformation of adaptive genetic-gemographic structure of multinational populations of Russia and CIS." The First (third) Russian congress of medical geneticists, 14-16, Dec. 1994, (Moscow, Russia) p.112.
- [16] Y. E. Dubrova, T. M. Karafet, R. I. Sukernik and T. V. Gol'tsova, *Genetics*, **26**, 122 (1990).
- [17] L. I. Bol'shakova and A. A. Revazov, *Genetics*, **24**, 340 (1988).
- [18] O. L. Kurbatova and E. Y. Pobedonosceva, Genetic-demographic processes at urbanization: migration, outbreeding and marriage assortativeness. In: Y. P. Altukhov *Heredity of man and environment*. Science, Moscow, **2**, pp.7–22 (1992).
- [19] R. L. Berg and S. N. Davidenkov, *Heredity and inherited diseases of man*, Leningrad (1971).
- [20] V. Surin, *News*, № 224, (2003).
- [21] *On the threshold of XXI century: Report about the world development of 1999/2000*. 23.
- [22] Y. Vnitskikh, *New ideas in philosophy*, **13**, 2004 at http://www.psu.ru/psu/files/1641/09_Vnitskih.doc.
- [23] V. S. Baranov, *Soros educational magazine*. № 3, 3 (1999).
- [24] E. Mayr, *The growth of biological thought: diversity, evolution; and inheritance*, Cambridge: Harvard University Press (1985).
- [25] G. W. Williams and R. M. Nesse, *Q. Rev. Biol.*, **66**, 1 (1991).

- [26] R. M. Nesse and G. C. Williams, *Why we get sick? The new science of Darwinian medicine*, Vintage Books, New York (1996).
- [27] R. M. Nesse, *World Psychiatry*, 7 (2002).
- [28] J. V. Neel, *Am. J. Hum. Genet.*, **14**, 4, 353 (1962).
- [29] C. Liberman, *Russian cardiologic journal*, 35 (2002).
- [30] M. L. Harris, W. C. Hadden, W. C. Knowler and P. H. Bennett, *Diabetes*, 523 (1987).
- [31] J.-M. Ekoé, P. Zimmet, and D. R. R. Williams, *The epidemiology of diabetes mellitus: an international perspective*, Wiley and Sons, Chichester, UK (2001).
- [32] Y. P. Altukhov, *Man*, № 6, 2003 at <http://vivovoco.rsl.ru/VV/PAPERS/MEN/GENETICS.HTM>.
- [33] V. S. Baranov and E. V. Baranova, *Ecological genetics*, 22, (2004).
- [34] E. K. Ginter, A. A. Revazov and A. N. Petrin, Influence of urbanization on the load of the inherited diseases in populations. In press Y. P. Altukhov “*Heredity of man and environment*.” Moscow: Science, 2, 22, (1992).
- [35] V. N. Serov, G. M. Burduli and O. G. Frolova, *The reproductive losses*. Moscow: Success (1997).
- [36] N. Zhilka, T. Irkina and V. Teshenko, *The condition of reproductive health in Ukraine (medico-demographic survey)*. Kiev (2001).
- [37] N. B. Kerimi, A. S. Sergeev, A. G. Mazovetskii and T. L. Kuraeva, *Genetica*, **20**, 166, (1984).
- [38] S. Ghosh and N. J. Schork, *Diabetes*, **45**, 1, (1996).
- [39] M. Mc Carthy, P. Froguel and G. A. Hitman, *Diabetologia*, **37**, 959, (1994).
- [40] S. Shtandel, S. Finogenova and L. Atramentova, *Diabetologia*, **41**, Suppl. 1, A104, (1998).
- [41] S. A. Shtandel, L. A. Atramentova, S. A. Finogenova and A. R. Gevorkyan, *Cytology and genetics*, **34**, 34, (2000).
- [42] T. M. Frayling and M. I. McCarthy, *Diabetologia*, **50**, 2229, (2007).
- [43] M. I. Balabolkin, E. M. Klebanova and V. M. Kreminskaya *Treatment of diabetes mellitus and his complications: Guidance for doctors*, “Medicine,” Moscow, (2005).
- [44] P. Zimmet and D. McCarty, *IDF Bulletin*, 39, (1995).
- [45] U.K. Prospective Diabetes Study (UKPDS) Group. Intensive blood-glucose control with sulphonylureas or insulin compared with conventional treatment and risk of complications in patients with type 2 diabetes (UKPDS 33), *Lancet*, **352**, 837 (1998).
- [46] M. Mkrtum'yan, *The attending physician*, № 10, 32, (2003).
- [47] S. A. Shtandel, *Natural selection of metabolic syndrome basic clinical components in modern terms*. (2012). Ch. M. Lopez Garsia and P. A. Perez Gonzalez, “*Handbook on Metabolic Syndrome. Classification, Risk Factors and Health Impact*.” New York: Nova Biomedical. 333, (2012).
- [48] Report of Expert Committee on the Diagnosis and Classification of Diabetes Mellitus. *Diabetes Care*, **25**, Suppl. 1, 5, (2002).
- [49] K. Hamaguchi, Y. Kusuda, N. Abe and T. Sakata, *Chin. J. Pathophysiol*, **17**, 768, (2001).
- [50] K. J. Basile, V. C. Guy, S. Schwartz and S. F. Grant, *Curr Diab Rep.*, **14**, 550, (2014).
- [51] S. A. Shtandel' and T. M. Tikhonova, *Cytology and Genetics*, **42**, 358, (2008).
- [52] N. D. Tron'ko and A. D. Chernobrov, *The basic indexes of activity of endocrinologic service of Ukraine for 2007*. Kiev, 31 tabl, (2008).

- [53] N. D. Tron'ko and A. D. Chernobrov, *The basic indexes of activity of endocrinologic service of Ukraine for 2000*. Kiev, 30 tabl, (2001).
- [54] N. D. Tron'ko and A. D. Chernobrov, *The basic indexes of activity of endocrinologic service of Ukraine for 2001*. Kiev, 30 tabl, (2002).
- [55] N. D. Tron'ko and A. D. Chernobrov, *The basic indexes of activity of endocrinologic service of Ukraine for 2002*. Kiev, 30 tabl, (2003).
- [56] N. D. Tron'ko and A. D. Chernobrov, *The basic indexes of activity of endocrinologic service of Ukraine for 2003*. Kiev, 30 tabl, (2004).
- [57] N. D. Tron'ko and A. D. Chernobrov, *The basic indexes of activity of endocrinologic service of Ukraine for 2013*. Kiev, 31 tabl, (2014).
- [58] *The basic indexes of the specialised endocrinologic help to the population of the Ukrainian Soviet Socialist Republic in 1989 – 1990*. The Kiev Scientific research institute of endocrinology and metabolism, Kiev, 30 tabl, (1991).
- [59] Y. Boyarskiy Population and methods of his study. *Statistica*, Moscow (1975).
- [60] W. Zheng, *Diabetes*, **54**, 906, (2005).
- [61] H. Donner, H. Rau, P.G. Walfish, T. Siegmund, R. Finke, J. Herwig, K. H. Usadel and K. Badenhoop, *Journal of Clinical Endocrinology and Metabolism*, **82**, 143, (1997).
- [62] Ch. Ezenwaka, R. Kalloo, M. Uhlig, R. Schwenk and J. Eckel, *Journal of Endocrinology*, **185**, 439, (2005).
- [63] M. Bland and D. G. Altman, *BMJ*, **320**, 1468, (2000).
- [64] N. G. Goyda, *Scientific basis of the development of the medicosanitary help system to women with an extragenital pathology*. Avtoessay of DMed thesis: 14.02.03. National medical university by name of O.O. Bogomol'tsa. Kiev, (2000).
- [65] E. D. Cherstvoy, G. I. Kravcova and G. I. Lazyuk, *Diseases of fruit, new-born and child*. Higher school, Minsk (1991).
- [66] N. Bottini, L. Musumeci, A. Alonso, S. Rahmouni, K. Nika, M. Rostamkhani, J. MacMurray, G. F. Meloni, P. Lucarelli, M. Pellecchia, G. S. Eisenbarth, D. Comings and T. Mustelin, *Nat Genet.*, **36**, 337, (2004).
- [67] M. Fedetz, F. Matesanz, A. Caro-Maldonado, I. I. Smirnov, V. N. Chvorostinka, T. A. Moiseenko and A. Alcina, *Tissue Antigens*, **67**, 430, (2006).
- [68] Petrone, C. Suraci, M. Capizzi, A. Giaccari, E. Bosi, C. Tiberti, E. Cossu, P. Pozzilli, A. Falorni, R. Buzzetti and NIRAD Study Group, *Diabetes Care*, **31**, 534, (2008).
- [69] Cervin, V. Lyssenko, E. Bakhtadze, E. Lindholm, P. Nilsson, T. Tuomi, C. M. Cilio and Leif Groop, *Diabetes*, **57**, 1433, (2008).
- [70] D. Abramov, I. I. Dedov, D. Y. Trophimov, M. N. Boldyreva, T. L. Kuraeva and L. P. Alekseeva, *Diabetes Mellitus*, № 3: (2007), http://dmjournal.ru/_mod_files/_upload/SD2007_3_2.pdf.
- [71] M. Krokowski, J. Bodalski, A. Bratek, P. Machejko and S. Caillat Zucman, *Diabetes Metab.*, **24**, 241, (1998).
- [72] D. A. Chistiakov, K. V. Savost'anov and V. V. Nosikov, *BMC Genet.*, **2**, (2001), <http://www.biomedcentral.com/1471-2156/2/6>.
- [73] C. Guja, S. Marshall, K. Welsh, M. Merriman, A. Smith, J. A. Todd and C. Ionescu Tirgoviste, *J. Cell Mol. Med.*, **6**, 75, (2002).
- [74] Fajardy, A. Vambergue, C. Stuckens, J. Weill, P. M. Danze and P. Fontaine, *Eur. J. Immunogenet.*, **29**, 251, (2002).

- [75] M. Mochizuki, S. Amemiya, K Kobayashi, Y. Shimura, T. Ishihara, Y. Nakagomi, K. Onigata, S. Tamai, A. Kasuga and S. Nanazawa, *Diabetes Care*, **26**, 843, (2003).
- [76] F. K. Kavvoura and J. P. Ioannidis, *Am. J. Epidemiol.*, **162**, 3, (2005).
- [77] T. Yanagawa, T. Maruyama, K. Gomi, M. Taniyama, A. Kasuga, Y. Ozawa, M. Terauchi, H. Hirose, H. Maruyama and T. Saruta, *Autoimmunity*, **29**, 53, (1999).
- [78] O. Cinek, P. Drevnek, Z. Sumnk, B. Bendlová and S. Kolousková, *Eur. J. Immunogenet.*, **29**, 219, (2002).
- [79] Dedov, L. I. Kolesnikova and T. P. Bardimova, *Polymorphism of CTLA4 (49 A/G) gene in patients of buryat population with the 1 type diabetes Materials* in press (2006). The 13th International Congress on circumpolar health, June 12-16 2006. (Novosibirsk) p. 63.
- [80] Y. Y. Li, *Mol. Biol. Rep.*, 141, (2013).
- [81] Dedov and M. V. Shestakova, *Diabetes mellitus: diagnostics, treatment, prophylactic*. Moscow (2011).
- [82] V. V. Nosikov and Y. A. Seregin, *Molecular biology*, **42**, 867, (2008).

Chapter 14

GENETIC DRIFT AMONG NATIVE PEOPLE FROM SOUTH AMERICAN GRAN CHACO REGION AFFECTS INTERLEUKIN 1 RECEPTOR ANTAGONIST VARIATION

Cecilia Inés Catanesi* and Laura Angela Glesmann

Laboratory of Genetic Diversity, IMBICE, La Plata, Argentina

ABSTRACT

Genetic variation is generally responsible for ethnic differences in certain diseases, including inflammatory processes. The antagonist of cytokine IL-1, IL-1Ra, has been widely studied among Caucasian and African populations for genetic polymorphisms, and interethnic differences have been documented. However, the variation and genotype distribution of polymorphisms from these genes among South American Amerindians are thus far unknown. We present the results for a VNTR located in the IL-1Ra second intron, in a sample of 169 individuals belonging to 5 Native American populations from Argentina and Paraguay, identified as native according to their self designation, and their geographic location. We also compare this data with the results obtained from a sample of non-native Argentinian people. Among the five known alleles of the VNTR, we found only two (alleles 1 and 2) in the native populations from Gran Chaco, and heterozygosity was 19%. The allele 2 which is considered proinflammatory (IL-1Ra * 2) has been found in homozygosity at a considerable frequency among native individuals. However, the association of this allele with inflammatory disease previously demonstrated for other populations of the world, might not be acting in the same way for native people, probably due to local adaptation. This would indicate that the allele 2 will probably not have a negative influence on individuals of native origin who have homozygous genotype 2-2. On the contrary, few records on inflammatory disease are available for the native people.

It seems that the increment on allele 2 is not related to any adaptive process but to genetic drift, that changes randomly the allele frequencies of different genetic regions along the genome. The effect of genetic drift has already been demonstrated with genetic markers located in autosomes, X and Y chromosomes. These results indicate that we must be very cautious when studying populations that passed a process of genetic drift, which can

* Corresponding Author's Email: ccatanesi@imbice.gov.ar.

become a confounding factor in epidemiological studies. This information will contribute to a future understanding of the association of this polymorphism with disease, and its incidence in different ethnic groups.

TEXT

The native human populations inhabiting the American territory have arrived to the continent through different migratory events, and have been involved in social and cultural interactions with other human groups. This exchange can be partially reflected in their current genetic structure. However, other processes than genetic flow have played an important role in the microevolutionary change of these populations, and genetic drift might be, even nowadays, one of the most important processes acting on them.



Figure 1. Geographic location of the phytogeographic Gran Chaco region.

Before the arrival of the European colonizers, the Gran Chaco region possessed a rich cultural and linguistic diversity. But the initial contact between Native Americans and Europeans five hundred years ago, initiated a dramatic reduction of the native populations, as happened in other parts of the American continent (Martínez Sarasola 2004). Many groups

were completely wiped out, while others have introduced some genetic admixture from non-native groups (Mulligan et al. 2004) giving rise through the centuries to the current Latin American populations, which share three main ancestral origins: Native American, European, and African (Salzano & Bortolini 2002, Rondón et al. 2008). In this chapter we present the analysis of some extant native groups inhabiting the South American Gran Chaco region, and the genetic drift process that can be found through genetic analysis. Information on these populations is not always concordant with the conclusions on their origin and history (Torroni et al. 1993, Horai et al. 1993, Zago et al. 1995), in part due to the genetic drift process, which generates a strong differentiation among tribes.

The Gran Chaco is a plain, very wide phytogeographic region of approximately 100.000 km², which includes part of Bolivia, Paraguay, and northeast of Argentina (Figure 1). About 5.000 years ago it was colonized by several nomadic native groups of hunting, fishing and gathering habits: Abipon, Mocoví, Qom, Mbya, and Pilagá belonging to Guaycurú linguistic group, Wichí, Chorote, and Chulupí of the Mataguayo linguistic family, and the Lule-Vilela group in the western part, corresponding to current Argentinian territory. In the northeast, the Ayoreo and Lengua, live in the region that belongs to Paraguay. These ethnic groups did not practice written language, therefore the historical record begins with the arrival of Spanish colonizers, and before that moment, information can be obtained only from oral transmitted tales (Martínez-Sarasola 2004; Tissera, 2008). These native groups show an important diversity of spoken languages as a result of complex interactions and interchanges (Jurado Medina et al. 2014).

BACKGROUND

Molecular biology offers a wide variety of coding and non coding (neutral) DNA markers for analyzing the diversity among individuals of a population, and among populations of a region. There is a background of information on DNA markers for Gran Chaco populations including several SNPs (single nucleotide polymorphisms), STRs (short tandem repeats, also called microsatellites), and insertions-deletions. SNPs are small changes in the DNA sequence affecting only one nucleotide. Usually, SNP mutation rates are moderate to low. STRs are short sequences tandemly repeated, they are highly variable and highly informative for studying genetic diversity and evolutionary processes. Insertion-deletion markers are sequences that are either present or absent, and can extend from only one or few nucleotides to several hundreds of them. An important amount of information has been obtained from non coding markers in Y chromosome, mitochondrial DNA, autosomes and X chromosome, while much fewer information is available from coding regions, such as blood antigens and HLA genes (Goicoechea et al. 2001, Dejean et al. 2004).

UNIPARENTAL Y CHROMOSOME VARIABILITY

Male specific region of Y chromosome has been widely studied throughout the world, allowing to disentangle the genetic structure of different populations. The diversity of Gran

Chaco native people has been analyzed through uniparentally inherited markers localized in Y chromosome, SNP and STR polymorphisms.

For the native populations of Gran Chaco, the search of autochthonous Q1a3a haplogroup has been firstly focused among different Chaco groups to determine either a native or a non-native origin of this chromosome. This haplogroup is determined by M3 SNP marker. Among those considered native, a set of STR markers has been usually genotyped to determine the diversity of each analyzed group.

An analysis on three different ethnic groups, Pilaga, Wichí, and Toba on SNP and STR markers showed genetic drift as a powerful evolutionary force for these seasonal hunters living in small bands (Demarchi and Mitchell 2004). The Native American-specific M3 marker was carried by 76.9% of the individuals, with a moderately high intergroup variation based on Y-chromosome STRs (10.7%).

In this way, two populations of Wichí origin living in Formosa province, 70 kilometers away from each other, were analyzed by Ramallo et al. (2009), showing a frequency of the Q1a3a haplogroup of 72,7% and 81,6%, and a rich lineage variability regionally distributed. Allelic variation was also non coincident. Their nomadic way of life and their habit to live in small groups has been clearly a strong force driving to genetic drift. Moreover, the Wichí people show a distribution of partialities called “Wichí ethnic complex” with certain linguistic differences (Braunstein 2006, Ramallo et al. 2009).

However, a small sample belonging to another ethnic group, did not show such influence. The Mocoví people living south to Gran Chaco, in Santa Fe province, were also analyzed for Y markers including two SNPs (M3 and M346) and ten STRs (Glesmann et al. 2011). The M3-T transition was present in 52% of the individuals, and STR haplotype diversity was 99.69%. In this case, the Mocoví Y-chromosomes still retain an interesting variability, with some of the M3-T haplotypes not found in other Amerindian groups. This considerable amount of haplotype variability is likely to be originary from this population.

X CHROMOSOME VARIABILITY

Due to its particular mode of inheritance, and its lower recombination rate compared to autosomes, the patterns of genetic variation of different types of markers specifically located in the X chromosome can result highly informative for population studies. Indeed, its special characteristics give a chance for analyzing genetic variability from a different point of view (Bourgeois et al. 2009, Ribeiro Rodrigues et al. 2009, 2011).

A study on X chromosome variation was performed including five Chaco ethnic groups: from Argentina, Wichí and Chorote from Salta province, and Mocoví from Santa Fe; from Paraguay: Lengua and Ayoreo (Catanesi et al. 2007). This analysis showed significant differences among these populations, with high F_{st} and R_{st} values, probably due to the drift process. The differentiation was related to the geographic location of populations, grouping those from Salta together, and those from Paraguay together, with Mocoví resulting separated from the rest (Catanesi et al. 2007).

A more recent study included another Wichí population living in a region called “Impenetrable” due to its hard climatic conditions and dense vegetation, in the Chaco province, and a Mocoví population (the same that was analyzed for Y chromosome) living in the south

of Gran Chaco (Glesmann et al. 2013). A high proportion of homozygotes and a marked linkage disequilibrium was found especially in Wichí, with differences in modal alleles and frequencies between both populations. On the other hand, a higher proportion of variation was observed in Mocoví. It has been reported that the Mocoví people belonging to the studied population are currently taking part of the neighbor non-native society through work and education.

This social integration might be responsible for a cultural change among Mocoví (Franceschi and Dasso, 2010). Although the genetic flow between Mocoví and non-natives might be occurring (Citro, 2006), a more important process of genetic drift may be reflected in the X chromosome variation reported, especially among Wichí. The Wichí people from Chaco province not only live isolated from other native and non-native groups, but also display an irregular distribution in small bands along the territory. Their geographic isolation and the extreme environmental conditions may be considered as the major factors contributing to the population differentiation (Glesmann et al. 2013).

UNIPARENTAL MITOCHONDRIAL VARIABILITY

Native Americans share five different maternally inherited mitochondrial DNA haplogroups: A, B, C, D, and X (Schurr et al. 1990, Torroni et al. 1993, Santos et al. 1996), which are present in different proportions depending on the demographic background of each population (Bailliet et al. 2004, Avena et al. 2012, Wang et al., 2008, Pauro et al. 2010, Yang et al. 2010, Motti et al. 2013).

An unbalanced proportion of male and female native uniparental lineages has been clearly described, as a consequence of mixed marriages of European males with native women through the last 5 centuries, thus making maternal haplogroups of native origin to widespread in different regions (García and Demarchi, 2006, 2009, Pauro et al. 2010, Yang et al. 2010).

Interestingly, a loss of mitochondrial variability has been described for certain Gran Chaco communities, including Wichí from Chaco province and Ayoreo from Paraguay. The random action of genetic drift or a bottleneck effect has been proposed as responsible for this reduction (Demarchi et al. 2001, Demarchi and Mitchell 2004).

AUTOSOME NON CODING VARIATION

Studying different genomic compartments has contributed to the understanding of the evolutionary processes occurring among South American Gran Chaco native populations. Different autosomal STR markers have shown in general a drop in the number of alleles and, consequently, an excess in the proportion of homozygote individuals. Variation in modal alleles for each native population, and a drastic reduction in allele number were found, particularly among Wichí (Tourret et al. 2000; Catanesi et al. 2001). As a consequence of genetic drift, a relatively poor correlation with geographic location of the tribes was more notable when analyzing autosomal variation (Zago et al. 1996; Tourret et al. 2000; Catanesi et al. 2001) than X chromosome variation (Catanesi et al. 2007).

AUTOSOME CODING VARIATION IN THE INTERLEUKINE 1 RECEPTOR ANTAGONIST

Coding genetic variation is generally responsible for ethnic differences in certain diseases. Natural selection changes allele frequencies according to environmental conditions, thus modifying the diversity of a population under selection (Hedrick 2000). Information on coding regions is scarce for native populations from Gran Chaco region, but the HLA region has been reported to present a low variability of dinucleotide STRs for three Chaco tribes: Wichí, Chorote, and Toba (Dejean et al. 2004), and metabolic genes have also been studied (Bailliet et al. 2007).

The inflammatory processes can be more prevalent among individuals who belong to a certain ethnic group, thus needing a specific medical treatment. Interleukin-1 (IL-1) is a cytokine secreted by macrophages and other cell types, playing an important role in the inflammatory response of virtually all cells and organs, as a major pathogenic mediator inducing pain (Dinarello et al., 2012; Gabay et al., 2010; Sims and Smith, 2010, Garlanda et al. 2013). The IL-1 gene family comprises several genes including three closely related IL-1A, IL-1B, and IL-1Ra encoding respectively two proinflammatory cytokines IL-1 α , IL-1 β , and their natural antagonist, the IL-1 receptor antagonist (IL-1Ra). The latter blocks IL-1 action by competing for its receptor (Dinarello 2009). The gene coding IL-1Ra maps the long arm of human chromosome 2 (band q14-21) (Steinkasserer et al. 1993, Grover et al. 2006). This gene presents a polymorphic VNTR in intron 2, with an 86 base pair repeated motif responsible for differences in the levels of expression of the receptor, which is reportedly associated with autoimmune diseases, ischemic stroke, and other pathologies (Worrall et al. 2007). The VNTR presents 5 allelic variants corresponding to 2, 3, 4, 5, and 6 copies of the 86 bp repeat (Tarlow et al. 1993, Grover et al. 2006). Three potential protein binding sites are located nearby this polymorphism, therefore the number of repeats may influence the level of gene transcription and posterior translation to a protein product.

Studies on North American U.S. population proposed the association of the allele IL-1Ra*2 of this VNTR with chronic inflammatory processes and pain (Joos et al. 2001, Foster et al. 2004), and this allele has also been associated as a risk factor for various autoimmune diseases (Rider et al. 2000, You et al. 2007, Havemose-Poulsen et al. 2007).

Since genetic diversity at the immune system is primarily important for human population's survival, the variation of this polymorphism in a group of native people inhabiting the South American Gran Chaco was explored.

A sample of 169 Amerindians from the Gran Chaco region -including Argentina and Paraguay, and a sample of 107 non-Amerindian Argentineans mainly of European ancestry (from Misiones and Buenos Aires provinces, Argentina), were analyzed. The Amerindians, identified as native according to their self designation, and their geographic location, comprised individuals belonging to five native groups from Gran Chaco: Lengua (from Paraguay, n = 36), Ayoreo (from Paraguay, n = 41), Chorote (from Santa Victoria Este, Salta province, Argentina, n = 20), Wichí (also from Santa Victoria Este, n = 39), and Mocoví (from Colonia Dolores, Santa Fe province, Argentina, n = 33). A small sample of 26 non-Amerindian Argentinians of full Japanese ancestry was also genotyped, however this group was not polymorphic for this VNTR, showing only homozygote individuals for allele 4, probably due to the small number of individuals analyzed, therefore this sample was not included in the comparative analysis.

The VNTR was amplified using the primers Fw: CTCAGCAACACTCCTAT, and Rv: TCCTGGTCTGCAGGTAA, in a MPI-Evo02 Thermocycler (La Plata, Argentina). Cycling conditions included 36 cycles of 94° 40 sec., 57° 1 min., and 72° 1 min., with an initial denaturation of 94° 2 min., and a final extension of 72° 5 min. Alleles were defined in 2% agarose electrophoresis, as in Foster et al. (2004). After post-gel DNA staining with GelGreen™ (Biotium, Hayward, CA), the electrophoretic bands were visualized in an image analyzer GelDocXR (Biorad, USA).

Allele frequencies, gene diversity, exact test of Hardy-Weinberg equilibrium, molecular variance (AMOVA), and pairwise genetic distances measured as Wright index F_{st} and R_{st} , were analyzed with Arlequin v. 3.5 (Excoffier et al. 2010). The F_{st} index estimates the amount of genetic differentiation between populations, by comparing total heterozygosity (five populations together) to each population heterozygosity. Genetic distances were represented using a matrix of distance MDS (multi dimensional scaling) with the program NTSYS 2.1 (Exeter) using R_{st} Slatkin's estimation from allele frequencies.

We found three out of the five known alleles of this VNTR (Table 1). The allele 2 was more frequent among the native people than in European and North American populations, and the allele 5 was found only in non-native European ancestry people.

Table 1. Sample size and genotype frequencies observed for IL1-Ra VNTR. Alleles are named by the number of repeats

Population	Sample size	Genotype Frequencies					
		2/2	2/4	2/5	4/4	4/5	5/5
Lengua	36	0.111	0.278	-	0.611	-	-
Ayoreo	41	0.024	0.195	-	0.780	-	-
Chorote	20	0.2	0.1	-	0.7	-	-
Wichí	39	0.205	0.154	-	0.641	-	-
Mocoví	33	0.061	0.182	-	0.757	-	-
Buenos Aires	104	0.038	0.010	-	0.875	0.019	0.058
Misiones	77	0.065	0.078	0.013	0.740	0.013	0.091
Japanese	26	-	-	-	1	-	-

Total native heterozygosity was 19%, while non-native people showed a much lower value, 2,5%. Genotypic distribution did not fit Hardy-Weinberg equilibrium among Chorote and Wichí Amerindians (Chorote: observed heterozygosity = 0.100, expected heterozygosity = 0.385, $P = 0.00278 \pm 0.00005$; Wichí: observed heterozygosity = 0.154, expected heterozygosity = 0.410, $P = 0.00019 \pm 0.00001$). It is noteworthy that the VNTR did not fit populational equilibrium in two cases, mostly due to the presence of higher frequencies of homozygotes for allele 2 than expected (Table 2). The genotyping of this marker was rechecked for those homozygotes, and such genotypes were confirmed.

Pairwise F_{st} analysis showed significant values for non-native (Buenos Aires + Misiones) compared to native populations, except for Ayoreo, and also between Wichí and Ayoreo (Table 3, data below diagonal). On the other hand, for pairwise R_{st} estimations non significant values were observed (Table 3, data above diagonal). R_{st} value between Wichí and non-native gave a limit probability of $P = 0,5$, showing a tendency to differentiation.

Table 2. Allele frequencies for the five Amerindian ethnic groups (frequencies +/-S.D.)

Allele	Mocoví	Lengua	Ayoreo	Chorote	Wichí
2	0.152+/-0.044	0.250+/-0.051	0.122+/-0.036	0.250+/-0.069	0,282+/-0,051
4	0.848+/-0.044	0.750+/-0.051	0.878+/-0.036	0.750+/-0.069	0,718+/-0,051

Table 3. Population pairwise Fst and Rst values below and above the diagonal, respectively. Significant results are in bold ($\alpha = 0.05$)

Population pairwise Rst							
Population pairwise Fst		Non-native	Wichí	Chorote	Ayoreo	Lengua	Mocoví
	Non-native		0.01555	-0.00278	-0.00242	0.00310	-0.00898
	Wichí	0.12312		-0.01662	0.06532	-0.01088	0.03460
	Chorote	0.09339	-0.01662		0.04015	-0.01983	0.01135
	Ayoreo	0.00760	0.06532	0.04015		0.04112	-0.01008
	Lengua	0.09222	-0.01088	-0.01983	0.04112		0.01524
	Mocoví	0.01772	0.03460	0.01135	-0.01008	0.01524	

An analysis of molecular variance (AMOVA) (Table 4) showed a significant 3.77% differentiation between native and non-native people ($P = 0.03773$).

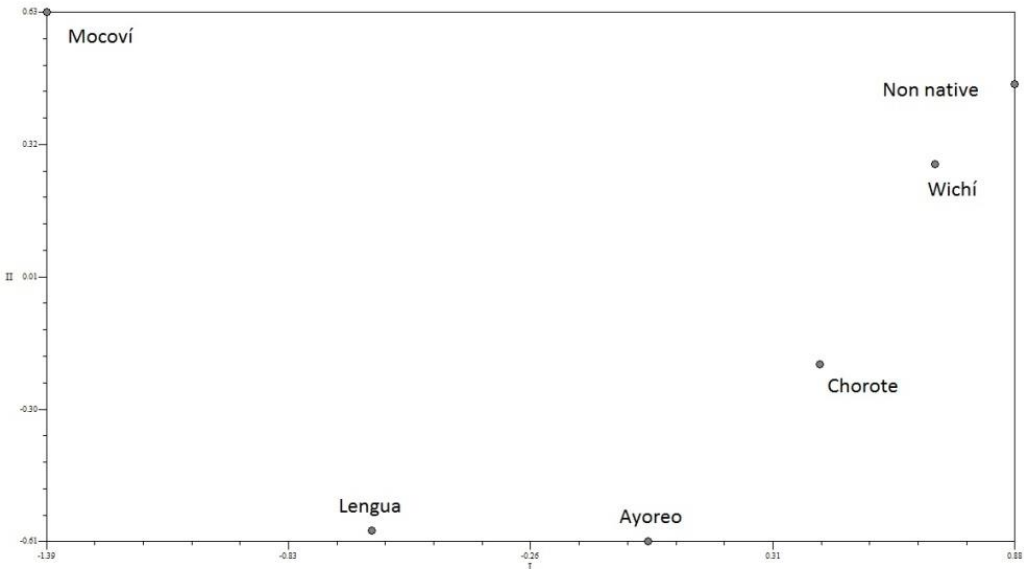


Figure 2. Multidimensional scaling (MDS) obtained from Rst distance matrix (Stress value = 0.0000). Non-native sample includes populations from Buenos Aires and Misiones.

Multidimensional map obtained from Rst distances (Figure 2) did not show any clustering, but the distribution of native populations showed certain agreement with their geographic distribution, with Mocoví (the population in a southernmost location) positioned distant from

all other populations, Lengua near Ayoreo, and Chorote near Wichí. It is interesting to remark the close position of Wichí and non-native people, consistent with a migrational process. However, F_{st} values suggest the migrational process is more likely occurring among these native populations.

Table 4. AMOVA Analysis (distance method: F_{st}). Samples for testing genetic structure were grouped as native and non-native populations $\alpha = 0.05$

	Sum of squared	Variance components	Percentage of variation
Among groups	2,143	0.00565	3.77
Among populations within groups	1.443	0.00329	2.20
Within populations	76.869	0.14079	94.03
Total	80.455	0.14973	100

DISCUSSION

The existence of an enormous drift in American native people, acting on small subpopulations and generating variation from one population to another, has been emphasized. The stochastic accumulation of differences through the time often increases differences among populations, while intrapopulation variation decreases enormously (Cavalli-Sforza et al. 1996).

Our results on a VNTR polymorphism within a coding region might be interpreted in different ways. On the one hand, natural selection might be acting on these populations favouring proinflammatory allele 2 in order to increase a particular inflammatory-immune pathway, since many individuals belonging to native Gran Chaco populations are still living under strict environmental conditions. However, it is more likely to interpret these results as the effect of finite population size making variability to be lost rapidly.

The decrease in heterozygosity of each population compared to the diversity of the pooled native sample taken together, and the low variability observed within populations are consistent with a genetic drift process (Holsinger 2015).

When populations are isolated from one another, as it happens especially with Wichí people, they will tend to diverge from one another as a result of genetic drift. This tendency operates much faster in populations with few individuals, driving them to divergence (Hedrick 2000).

The influence of genetic drift could be detrimental to perform association genetic studies for specific diseases, because case-control and/or cohort studies might drive to erroneous conclusions when comparing populations which are substructured and in disagreement with Hardy-Weinberg equilibrium (Acosta et al. 2012).

However, the association of this allele with inflammatory disease previously demonstrated for other populations of the world, might not be acting in the same way for native people, probably due to local adaptation. This would indicate that the allele 2 probably does not influence negatively on individuals of native origin who have homozygous genotype 2-2. On the contrary, few records on inflammatory disease are available for the native people (Trujillo Miriam, personal communication). It seems that the increment on allele 2 is not related to any

adaptive process but to genetic drift, that changes randomly the allele frequencies of different genetic regions along the genome. The effect of genetic drift has already been demonstrated with genetic markers located in autosomes, X and Y chromosomes (see above).

These results indicate that we must be very cautious when studying populations that passed through a process of genetic drift, which can become a confounding factor not only in evolutionary studies, but in epidemiological studies. In the first case, it could be considered as a positive selection effect; in the second one, the existence of a genetic association with certain phenotype could be falsely interpreted (Solovieva et al. 2004).

This information will contribute to a future understanding of the association of this polymorphism with disease, and its incidence in different ethnic groups.

REFERENCES

- Acosta O, Solano L, Huerta D, Oré D, Sandoval J, Figueroa J, Fujita R. 2012. Variabilidad genética de la respuesta inflamatoria. I. Polimorfismo-511 C/T en el gen IL1 β en diferentes subpoblaciones peruanas. [Genetic variability of the inflammatory response. I. Polymorphism-511 C / T in the IL1 β gene in different Peruvian subpopulations]. *Anales de la Facultad de Medicina*, Vol. 73, Núm. 3.
- Avena S., Via M., Ziv E., et al. (2012). Heterogeneity in Genetic Admixture across Different Regions of Argentina. *PloS ONE* 7(4): e34695.
- Bailliet G, Rothhammer F, Carnese FR, Bravi CM, Bianchi NO. 1994. Founder mitochondrial haplotypes in Amerindian populations. *Am. J. Hum. Genet.* 54:27-33.
- Bailliet G., Santos M., Alfaro E., Dipierri J., Demarchi D., Carnese F., Bianchi N. (2007). Allele and genotype frequencies of metabolic genes in Native Americans from Argentina and Paraguay. *Mutat. Res. Mar* 5;627(2):171-7.
- Bourgeois S, Yotova V, Wang S, et al. 2009. X-Chromosome Lineages and the Settlement of the Americas. *Am. J. Phys. Anthropol.* 140: 417–428.
- Braunstein, J. 2006. *El signo del agua. Formas de clasificación étnica Wichí. Definiciones étnicas, organización social y estrategias políticas en el Chaco y la Chiquitania*. [The water sign Forms of Wichí ethnic classification. Ethnic definitions, social organization and political strategies in the Chaco and Chiquitania]. Editor: Isabelle Combes. Actes et Memoire de l'Institute Français d'Etudes Andines 11.
- Catanesi C, Martina P, Giovambattista G, Zukas P, Vidal Rioja L. 2007. Geographic structure in Gran Chaco Amerindians based on five X-STRs. *Hum. Biol.* 79 (4): 463-474.
- Catanesi C.I., Méndez M.G., Salceda S.A., Vidal Rioja L.A. 2001. *Perfil morfométrico-molecular de tres poblaciones nativas sudamericanas*. Milenio. [Morphometric-molecular profile of three South American native populations. Millennium]. M. Amanda Caggiano Editor, Chivilcoy (Argentina) 127-130. ISBN 987-98845. Edición no periódica con referato.
- Catanesi C.I., Turret M.N., Vidal Rioja L.B. 2001. Genetic diversity at LPL and CD4 microsatellite loci in South American populations. *Journal of Basic and Applied Genetics* XIV, 1: 41-46. Editada por la Sociedad Argentina de Genética (SAG). Argentina. ISSN BAG 1666-0390.

- Cavalli-Sforza, L., P. Menozzi, A. Piazza. 1996. In: *The History and Geography of Human Genes*. Princeton NJ: Princeton University Press.
- Citro S. 2006. *La Fiesta del 30 de Agosto entre los mocovíes de Santa Fe*. [The Festival of August 30 between the Mocovíes of Santa Fe]. Buenos Aires, Facultad de Filosofía y Letras, Universidad de Buenos Aires.
- Dejean CB, Crouau-Roy B, Goicoechea AS, Avena SA, and Carnese FR. 2004. Genetic variability in Amerindian populations of Northern Argentina. *Genetics and Molecular Biology*, 27, 4, 489-495.
- Demarchi DA, Mitchell RJ. 2004. Genetic Structure and Gene Flow in Gran Chaco Populations of Argentina: Evidence from Y-Chromosome Markers. *Human Biology*, v. 76, no. 3, pp. 413–429.
- Demarchi DA, Panzetta-Dutari GA, Motran CC, Marcellino AJ. 2001. Mitochondrial DNA haplogroups in Amerindian populations from the Gran Chaco. *Am. J. Phys. Anthropol.*, Vol. 115, N. 3: 199-203.
- Dinarello CA: Immunological and inflammatory functions of the interleukin-1 family. *Annu. Rev. Immunol.* 2009, 27:519-550.
- Dinarello CA, Simon A, van der Meer JW. Treating inflammation by blocking interleukin-1 in a broad spectrum of diseases. *Nature Rev Drug Discov.* 2012; 11:633–652.
- Excoffier L Lischer H. 2010. Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows. *Mol. Ecol. Resources*, Vol. 10, Issue 3: 564–567.
- Franceschi ZA and Dasso MC. 2010. *Etno-Grafías, La escritura como testimonio entre los Wichí. El universo Wichí: historia y cultura*. [Etno-Grafías, Writing as a testimony among the Wichí. The Wichí universe: history and culture]. Ed. Corregidor, Buenos Aires 336 pp.
- Foster D.C., Szenski T.M., Stodgell C.J. 2004. Impact of genetic variation in interleukin-1 receptor antagonist and melanocortin-1 receptor genes on vulvar vestibulitis syndrome. *The Journal of Reproductive Medicine*, 49 (7): 503-509.
- Gabay C, Lamacchia C, Palmer G. IL-1 pathways in inflammation and human diseases. *Nature Rev. Rheumatol.* 2010; 6:232–241.
- García A, Demarchi DA. 2009. Incidence and distribution of Native American mtDNA haplogroups in central Argentina. *Hum. Biol.* 81(1):59-69.
- Garlanda C, Dinarello CA, and Mantovani A. 2013. The interleukin-1 family: back to the future. *Immunity*. 2013 December 12; 39(6): 1003–1018.
- Glesmann L.A., Martina P.F., Méndez M.G., Catanesi C.I. 2011. High genetic variation in Y chromosome patterns of the Mocoví population. *Rev. del Museo de Antropología*, vol. 4: 179-186.
- Glesmann L.A., Martina P.F. Catanesi C.I. 2013. Genetic variation of X-STRs in the Wichí population from Chaco province, Argentina. *Human Biology*. Wayne State Univ. Press, Detroit. Paper 32.
- Goicoechea AS, Carnese FR, Dejean CB, Avena SA, Weimer TA, Franco MH, Callegari-Jacques S, Estalote A, Simoes ML, Palatnik M, Salomoni P and Salzano FM. 2001. Genetic relationships between Amerindian populations of Argentina. *Am. J. Phys. Anthropol.* 115:133-143.
- Grover S, Tandon S, Misra R, Aggarwal A. 2006. Interleukin-1 receptor antagonist gene polymorphism in patients with rheumatoid arthritis in India. *Indian J. Med. Res.* 123(6): 815-820.

- Havemose-Poulsen A., Sørensen L.K., Bendtzen K., Holmstrup P. 2007. Polymorphisms within the IL-1 gene cluster: effects on cytokine profiles in peripheral blood and whole blood cell cultures of patients with aggressive periodontitis, juvenile idiopathic arthritis, and rheumatoid arthritis. *J. Periodontol.* 78: 475-492.
- Hedrick PW. 2000. *Genetics of population*, 2nd. Ed. Jones and Bartlett Publishers. USA, 553 pp.
- Holsinger KE. 2015. Mutation, migration and genetic drift. In: *Lecture notes in population genetics*. Department of Ecology & Evolutionary Biology, University of Connecticut 333 pages.
- Horai S, Kondo R, Nakagawa-Hattori Y. 1993. Peopling of the Americas, founded by four major lineages of mitochondrial DNA. *Mol. Biol. Evol.* 10:23-47.
- Joos L., McIntyre L., Ruan J., Connett J.E., Anthonisen N.R., Weir T.D., Paré P.D., Sandford A.J. 2001. Association of IL-1 β and IL-1 receptor antagonist haplotypes with rate of decline in lung function in smokers. *Thorax*, 56: 863-866.
- Jurado Medina LS, Ramallo V, Calandra H, Lamenza G, Braunstein J, Salceda S, Bailliet G. 2014. Gran Chaco paternal lineages, a DNA approach. *Folia Historica del Nordeste* N° 22, Resistencia, Chaco, IIGHI - IH- CONICET/UNNE: 187-202.
- Martínez Sarasola C. 2004. *Las comunidades que ocupaban nuestro territorio en el siglo XVI. Nuestros Paisanos los Indios. Vida, historia y destino de las comunidades indígenas en la Argentina.* [The communities that occupied our territory in the sixteenth century. Our Paisanos los Indios. Life, history and destiny of the indigenous communities in Argentina]. Editorial EMECE, Buenos Aires, Argentina. I:77-82.
- Motti JMB, Muzzio M, Ramallo V, Rodenak Kladniew B, Alfaro EL, Dipierri JE, Bailliet G, Bravi CM. 2013. Origen y distribución espacial de linajes maternos nativos en el noroeste y centro oeste argentinos. [Origin and spatial distribution of native maternal lineages in the northwest and central west of Argentina]. *Rev. Arg. de Antropología Biológica*, Vol. 15, N. 1: 03-14.
- Mulligan C., Hunley K., Cole S., Long J. 2004. Population genetics, history, and health patterns in native Americans. *Annual Review of Genomics and Human Genetics* 5: 295-315.
- Pauro M, García A, Bravi CM, Demarchi DA. 2010. Distribución de haplogrupos mitocondriales autóctonos en poblaciones rurales de Córdoba y San Luis. [Distribution of allochthonous mitochondrial haplogroups in rural populations of Córdoba and San Luis]. *Rev. Argentina de Antropología Biológica* Vol.12, N.1: 47-55.
- Ramallo V, Santos MR, Muzzio M, Motti JMB, Salceda S, Bailliet G 2009. Linajes masculinos y su diversidad en comunidades Wichí de Formosa. *Revista del Museo de Antropología*, Facultad de Filosofía y Humanidades [Male lineages and their diversity in Wichí communities of Formosa. *Magazine of the Museum of Anthropology*, Faculty of Philosophy and Humanities] – Universidad Nacional de Córdoba – Argentina, 2: 67-74.
- Ribeiro-Rodrigues E., de Jesus Brabo Ferreira Palha T., Auler Bittencourt E., Ribeiro-dos-Santos A., Santos S. 2011. Extensive survey of 12 X-STRs reveals genetic heterogeneity among Brazilian populations. *Int. J. Legal Med.* 125:445–452.
- Ribeiro Rodrigues E., Dos Santos N., Campos Ribeiro dos Santos A., et al. 2009. Assessing Interethnic Admixture Using an X-Linked Insertion-Deletion Multiplex. *Am. J. Hum. Biol.* 21:707–709.
- Rider L.G., Artlett C.M., Foster C.B., Ahmed A., Neeman T., Chanock S.J., Jimenez S.A., and Miller F.W. 2000. Polymorphisms in the IL-1 receptor antagonist gene VNTR are possible

- risk factors for juvenile idiopathic inflammatory myopathies. *Clin. Exp. Immunol.* Jul 2000; 121(1): 47–52.
- Rondón F., Oosorio J., Peña A., Garcés H., Barreto G. 2008. Diversidad genética en poblaciones humanas de dos regiones colombianas. *Colomb Med.* 39 (Supl 2): 52-60.
- Salzano F. and Bortolini M. 2002. *The Evolution and Genetics of Latin American Populations*. Cambridge University Press, Cambridge, 512 pp.
- Schurr TG, Ballinger SW, Gan YY, Hodge JA, Merriwether DA, Lawrence DN, Knowler WC, Weiss KM, Wallace DC. 1990. Amerindian mitochondrial DNAs have rare Asian mutations at high frequencies, suggesting they derived from four primary maternal lineages. *Am. J. Hum. Genet* 46(3):613-623.
- Sims JE, Smith DE. 2010. The IL-1 family: regulators of immunity. *Nat. Rev. Immunol.* 2010; 10:89–102.
- Solovieva S, Leino-Arjas P, Saarela J, Luoma K, Raininko R, Riihimäki H. 2004. Possible association of interleukin 1 gene locus polymorphisms with low back pain. *Pain*, 2004 May;109(1-2):8-19.
- Steinkasserer A., Spurr N.K., Sim R.B. 1993. Chromosomal mapping of the human interleukin-1 receptor antagonist gene (IL-1RN) and isolation of specific YAC clones. *Inflammation Research* 38(2): C59-C60.
- Tarlow JK, Blakemore AI, Lennard A, Solari R, Hughes HN, Steinkasserer A, Duff GW. Polymorphism in human IL-1 receptor antagonist gene intron 2 is caused by variable numbers of an 86-bp tandem repeat. *Hum. Genet.* 1993 May;91(4):403-4.
- Tissera R. d. I. M. 2008. *Chaco, Historia General. Subsecretaría de Cultura, Min. De Educación, Cultura, Ciencia y Tecnología de la provincia de Chaco*. [Chaco, General History Undersecretary of Culture, Ministry of Education, Culture, Science and Technology of the province of Chaco]. Librería de la Paz, Resistencia, Chaco, Argentina, 46, 401, 404.
- Torroni A, Schurr TG, Cabell MF, Brown MD, Neel JV, Larsen M, Smith DG, Vullo CM, Wallace DC. 1993. Asian affinities and continental radiation of the four founding Native American mtDNAs. *Am. J. Hum. Genet.* 53:563-590.
- Tourret M.N., Catanesi C.I., Vidal Rioja L.B. 2000. Variability of the F13B locus in South American Populations. *Human Biology*, v.72, 4, 707-714. Editorial: Wayne State University Press, Detroit, Michigan, USA. ISSN 0018-7143.
- Wang S, Ray N, Rojas W, Parra MV, Bedoya G, Gallo C, Poletti G, Mazzotti G, Hill K, Hurtado AM, Camrena B, Nicolini H, Klitz W, Barrantes R, Molina JA, Freimer NB, Bortolini MC, Salzano FM, Petz-Erler ML, Tsuneto LT, Dipierri JE, Alfaro EL, Bailliet G, Bianchi NO, Llop E, Rothhammer F, Excoffier L, Ruiz-Linares A. 2008. Geographic patterns of genome admixture in Latin American mestizos. *PLoS Genet.* 4(3):e1000037.
- Worrall B.B., Brott T.G., Brown R.D., Brown W.M., Rich S.S., Arepalli S., Wavrant-De Vrièze F., Duckworth J., Singleton A.B., Hardy J., Meschia J.F., et al. 2007. IL1RN VNTR polymorphism in ischemic stroke: analysis in 3 populations. *Stroke*. 2007 Apr;38(4):1189-96.
- Yang NN, Mazieres S, Bravi C, Ray N, Wang S, Burley MW, Bedoya G, Rojas W, Parra MV, Molina JA, Gallo C, Poletti G, Hill K, Hurtado AM, et al. 2010. Contrasting Patterns of Nuclear and mtDNA Diversity in Native American Populations. *Ann. Hum. Genet.* 2010 Nov;74(6):525-38.

- You C., Li J., Xie X., Zhu Y., Li P., Chen Y. 2007. Association of interleukin-1 genetic polymorphisms with the risk of rheumatoid arthritis in Chinese population. *Clin. Chem. Lab. Med.* 45: 968-971.
- Zago, M.A., W.A. Silva. Jr., M.H. Tavella, S.E.B. Santos, J.F. Guerreiro and M.S. Figueiredo. 1996. Interpopulational and intrapopulational genetic diversity of Amerindians as revealed by six variable number of tandem repeats. *Hum. Hered.* 46:274-289.

Chapter 15

ALTERNATIVE SPLICING AND NEUROLOGICAL DISORDERS: FOCUS ON PARKINSON'S DISEASE

***Velia D'Agata¹, Valentina La Cognata^{1,2}
and Sebastiano Cavallaro^{1,*}***

¹Department of Biomedical and Biotechnological Sciences, Section of Human Anatomy and Histology, University of Catania, Catania, Italy,

²Institute of Neurological Sciences, Italian National Research Council, Catania, Italy

ABSTRACT

Alternative splicing (AS) is a fundamental mechanism of gene expression regulation that extremely expands the coding potential of genomes and the cellular transcriptomic and proteomic diversity. This dynamic and finely-tuned machinery is particularly widespread in the nervous system and is critical for both neuronal development and functions. Alternative splicing defects, therefore, frequently underlie neurological disorders. In this chapter, we will focus on Parkinson's disease (PD), the second most common neurodegenerative disorder worldwide. We will provide a current overview on the impact of alternative splicing in PD by representing the multiple splicing transcripts produced from the major PD-linked genes and their regulation in PD states. Furthermore, we will review the studies describing global splicing expression changes revealed by whole-genome transcriptomic approaches. We will also summarize the current knowledge about the alternative splicing modulation in PD through non-coding RNAs (miRNA and lcnRNA) molecules. Assessing the role of alternative splicing on PD pathobiology may represent a central step toward an improved understanding of this complex disease.

* Corresponding Author's Email: sebastiano.cavallaro@cnr.it (Tel: +39 095 7436634; Fax: + 39 095 7122426).

1. INTRODUCTION

The DNA-RNA-protein flowchart has been for long time considered the central dogma of molecular biology. Further steps of regulation are continuously discovered, greatly expanding this simplistic framework and revealing the complex network that controls gene expression [1]. The multilayer biological processes that regulate gene expression include several RNA-based mechanisms, such as regulation of mRNA translation efficiency, stability and localization, miRNA-mediated silencing, long non-coding RNAs modulation, nonsense-mediated decay, polyadenylation site selection and RNA editing. Among these complex circuits, the most crucial layer is represented by alternative splicing (AS), a process whereby multiple mRNA transcripts and protein isoforms with different functional properties are created from a single gene [1]. More than 90% of multi-exon human genes undergo AS [2, 3]. The main site of AS events is represented by the central nervous system [4, 5].

Besides the physiological involvement of AS during development, many evidences suggest a significant contribution of this mechanism in human pathologies, and therefore its clinical relevance is growing exponentially. It is estimated that 50% of disease-causing mutations affect pre-mRNA splicing [1].

In this chapter, we will briefly introduce the AS process and its role in the pathophysiology of the nervous system. Then, we will focus on the current knowledge about its impact on Parkinson's disease (PD), the second most common neurodegenerative disorder worldwide. Hence, the multiple transcripts generated by the major PD-linked genes and their regulation during the disease will be reported.

During last years, the rapid advancement of biotechnologies has allowed to detect simultaneously, on large scale, gene-expression changes in various pathophysiological conditions. As a result, several studies on the global expression of splicing variants in PD have been performed and will be recapitulated at the end of the chapter. Finally, the current knowledge about the epigenetic modulation of AS mechanism in PD will be summarized. Deciphering the functional role of AS transcripts and encoded protein isoforms in PD may represent an essential step toward an improved understanding of this complex disease.

2. THE ALTERNATIVE SPLICING MACHINERY

In humans and other complex metazoans, the vast majority of genes are discontinuous, in other words, they contain mRNA-encoding regions, called exons, which are interrupted by non-coding segments, the introns. This split gene structure offers a fertile ground for AS regulation, by providing incredible opportunities to enhance the transcriptome and proteome repertoire without the need for expansion of the genome.

The AS process consists in the removal of introns from the pre-mRNA transcript and joining of the coding regions in various exon-exon combinations to form a mature mRNA, which is then polyadenylated, exported to cytoplasm, and translated into protein. Four main constitutive DNA sequence motifs enable the splicing process: the 5' (GU) and the 3' (AG) splice sites, the lariat branch point, and the polypyrimidine tract. In addition to the *core* splicing elements, other nucleotide sequences within exons (also known as ESEs, or exonic splicing enhancers, and ESS or exonic splicing silencers) and introns (also known as ISEs or intronic

splicing enhancers, and ISSs or intronic splicing silencers) are critical for the correct recognition and combination of exons, by promoting or inhibiting the final assembly.

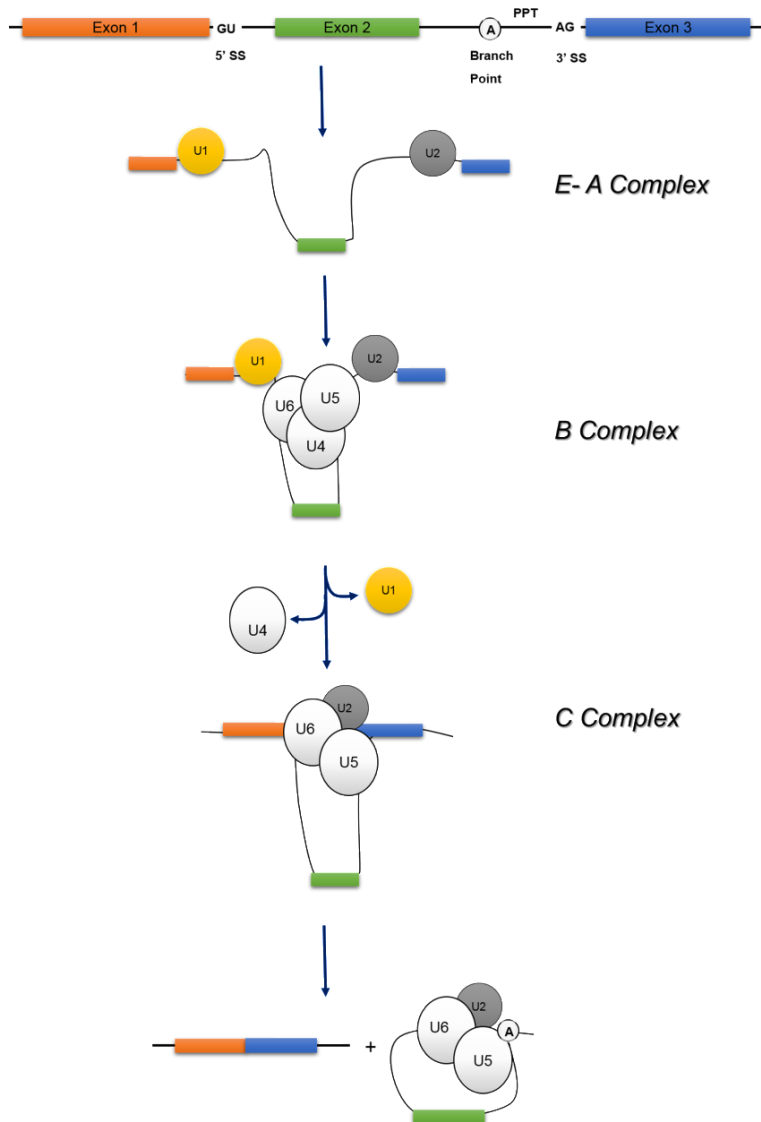


Figure 1. The alternative splicing machinery. Four main conserved DNA sequence motifs allow the splicing mechanism: the donor splice site GU (5' SS), the acceptor splice site AG (3' SS), the lariat branch point (A) located upstream of the acceptor site and the polypyrimidine tract (PPT) placed between the acceptor site and the branch point. The splicing machinery includes mainly five spliceosomal uridine-rich small nuclear ribonucleoproteins (snRNPs) (U1, U2, U4, U5, and U6), and further auxiliary RNA binding proteins. During the first step of spliceosome assembly, U1 snRNP base-pairs with the 5'-splice site of the pre-mRNA (E complex), whereas U2 base-pairs with the branch-point (A complex). Then, the tri-snRNP complex U4, U5 and U6 associates with the forming spliceosome (B complex), and both U1 and U4 are ejected. This allows U6 to replace U1 at the 5' splice site (C complex) and leads to a U6–U2 interaction that gets close together the 5'-splice site and the branch point, allowing for a transesterification step. At the end, U5 brings near the two exons, joining them through a second transesterification reaction.

Splicing is performed by the spliceosome, a multi-component modular molecular machine composed of five small uridine-rich nuclear RNAs (snRNAs) that tie with proteins to form small nuclear ribonucleoproteins (snRNPs) (U1, U2, U4, U5, and U6) (Figure 1) [6]. The splicing reaction is performed through a coordinated series of RNA–RNA, RNA–protein and protein–protein interactions. During the first step of spliceosome assembly, U1 snRNP base-pairs with the 5'-splice site of the pre-mRNA (E complex), whereas U2 base-pairs with the branch-point (A complex) (Figure 1). Then, a tri-snRNP complex, including U4, U5 and U6, associates with the spliceosome and form the B complex (Figure 1). Hence, both U1 and U4 are ejected, allowing to replace U1 with U6 at the 5' splice site (C complex) (Figure 1). The interaction of U6–U2 gets close together the 5'-splice site and the branch point, allowing a first transesterification reaction. In the end, U5 completes the process by joining the exons through a second transesterification reaction and removing the spliced-out exon [7].

The delicate and finely tuned interplay between the constitutive splicing motifs, the splicing regulatory sequences, the components of the spliceosome and other additional auxiliary RNA-binding proteins allows the occurrence of five major alternative splicing events: exon skipping/inclusion, use of alternative 3' splice site, use of alternative 5' splice site, mutually exclusive exons and intron retention (Figure 2) [2]. This combinatorial assembling of exons increases the number of transcribed mRNAs, and therefore, is able to generate numerous protein isoforms with different biological properties, protein-protein interactions and subcellular localization [8]. AS can also modify enzymatic activity, and even reverse the functional role of the isoforms: for instance, Bcl-X gene is able to produce a long form preventing cell apoptosis and a short form inducing programmed cell death. AS can also occur in non-coding sequences affecting the efficiency of mRNA translation, localization or stability through the “Nonsense-Mediated mRNA Degradation” (NMD) [9] [10].

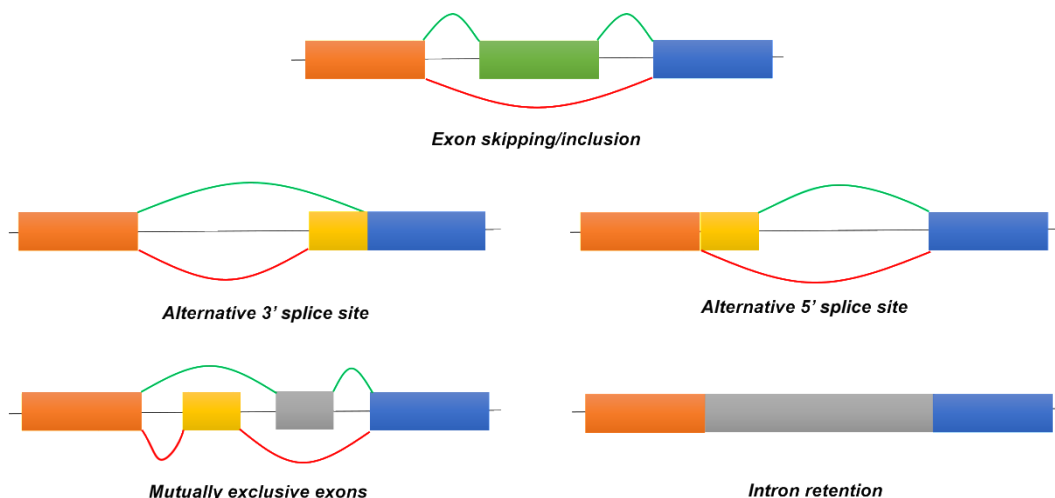


Figure 2. The alternative splicing mechanism. Five major alternative splicing events are currently known: exon skipping/inclusion, use of alternative 3' splice site, use of alternative 5' splice site, mutually exclusive exons, and intron retention. The splicing events rely on the interplay between the constitutive splicing motifs, the splicing regulatory sequences, the RNA secondary structures, the components of the spliceosome and further auxiliary RNA-binding proteins. However, how the spliceosome decides which exons to include remains currently not clear.

Overall, the AS mechanism works as an on–off switch in gene expression and represents an extremely economical mean to increase protein diversity, explaining the divergence between the estimated 24,000 human protein-coding genes and the 100,000 different proteins that are predicted to be generated [11]. It is a finely-tuned process regulated in time and space, allowing each mRNA to be expressed in specific cellular conditions [12]. The splicing molecular machinery must be both rapid and precise: this is a demanding task considering that introns are usually much larger than exons and that even a single nucleotide addition or deletion will shift the reading frame and modify the encoded protein sequence.

Two types of defects can interfere the AS regulation: mutations of the DNA elements required for correct pre-mRNA processing (called *cis*-acting mutations) or alterations of the factors that are necessary for splicing regulation (termed *trans*-acting defects). *Cis*- and *trans*-splicing aberrations represent direct causative agents of disease or can contribute to disease susceptibility or can modulate disease severity of several neurologic disorders, as we will describe in the next paragraph.

3. ALTERNATIVE SPLICING IN THE NERVOUS SYSTEM AND NEUROLOGICAL DISEASES

AS is particularly widespread in the nervous system. More than 40% of genes are alternatively spliced in brain, meaning an exceptionally high level of splicing events as compared to other tissues [13].

In recent years, our understanding of the fundamental role of AS in neural development and functions has significantly advanced, thanks to the coupling of various computational and experimental approaches [14]. Several studies have suggested that the elaborate and dynamic regulation of AS in the nervous system is critical for modulating protein-protein interactions, transcription networks, and multiple aspects of neuronal development and maintenance [14, 15]. The precise temporal and spatial expression of the highly complex repertoire of isoforms is able to perform complex tasks, such as appropriate connections among billions of neurons that occur during learning and memory processes. It has been demonstrated that the production of particular splice isoforms helps to determine the properties of many different types of neurons, and that several important events during neuronal development are under control of AS mechanisms, including cell-fate determination, axon guidance and synaptogenesis [4]. Splicing events have also been observed in brain areas during whole human lifespan, demonstrating the involvement of more than a third of expressed genes and their link to neural functions [16].

Given the importance of splicing in the nervous system circuits, it is not surprising that an extensive range of different neurological diseases is caused by splicing errors. Alzheimer's disease, Retinitis Pigmentosa, Spinal Muscular Atrophy, Muscular Dystrophy, Ataxia-telangiectasia, Neurofibromatosis, Amyotrophic Lateral Sclerosis, Fragile-X-associated Tremor/Ataxia Syndrome, Prader-Willi Syndrome, Schizophrenia and Autism disorders are directly caused or have an increased risk by either *cis*- or *trans*-acting splicing defects [1, 7, 17–20].

In this broad neurological disorders scenario, the relevance of AS in PD is not still clear, and the splicing mechanisms that regulate PD-related genes remain mostly unknown. In the next paragraphs, we will briefly introduce the genetics of Parkinson's disease, and then we will focus on the splicing regulation of PD-linked genes.

4. GENETICS OF PARKINSON'S DISEASE: BASIC CONCEPTS

Parkinson's disease (PD) is a progressive, debilitating neurodegenerative disorder and constitutes the second most common neurologic disease worldwide after Alzheimer's disease. Clinically, most patients present resting tremor, bradykinesia, stiffness of movement and postural instability. These major symptoms derive from the profound and selective loss of dopaminergic neurons of *substantia nigra pars compacta*. The pathological hallmarks of PD are round eosinophilic intracytoplasmic proteinaceous inclusions termed Lewy bodies (LBs) and dystrophic neurites (Lewy neurites) present in surviving neurons [21].

PD had been for a long time considered a sporadic non-genetic disorder. The identification of distinct genetic loci responsible for rare Mendelian forms of PD has revolutionized this view and has provided new insights to understanding the molecular pathogenesis of this disease [21]. Linkage mapping analysis, genome-wide association studies (GWAS) and next generation sequencing technologies have revealed an increasing number of locus and genes strongly linked to either dominant, recessive and X-linked forms of the disease [22, 23]. For example, mutations in *SNCA* (localized in the *PARK1* locus) and *LRRK2* (locus *PARK8*) are responsible for the development of the autosomal dominant form of PD; *PARKIN* (locus *PARK2*) *PINK1* (locus *PARK6*), *DJ1* (locus *PARK7*) are classified as typical recessive PD genes; mutations in *ATP13A2* (locus *PARK9*), *PLA2G6* (locus *PARK14*) and *FBXO7* (locus *PARK15*) cause the atypical recessive forms; while X-linked forms of PD disease derive from genetic defects in *ATP6A2* and *TAF1* genes. For the sake of completeness, we mention here further monogenic loci, not confirmed genes or risk factors genes (i.e., *PARK3*, *GBA*, *UCHL1*, *PARK10*, *GIGYF2*, *PARK12*, *HTRA2*, *PARK16*, *DNAJ*, *HLA-DR*, *GAK-DGKQ*, *SYNJ1*, *GBAP1*) [23-25]. Some additional genes have also implicated in PD (*SRRM2*, *MAO-B*, *SNCAIP* and *MAPT*) for their altered splicing regulation in PD states.

In the next paragraphs, we will describe the alternatively spliced mRNA variants of PD genes and the current scientific data demonstrating their involvement in PD pathogenesis.

5. ALTERNATIVE SPLICING OF GENES INVOLVED IN THE DOMINANTLY INHERITED FORMS OF PARKINSON'S DISEASE

Several evidences have demonstrated a link between specific genetic mutations and various forms of PD. In general, the inheritance patterns of human disorders are identified by examining the way the disorder is transmitted among the family members. In the autosomal-dominant inheritance, one mutated allele of the gene is enough to cause the disease [26].

Here, we will describe the current literature data about autosomal dominant PD-genes (*SNCA* and *LRRK2*) and the involvement of their splice variants in health and disease. Some additional genes have currently been linked to the development of dominantly inherited forms of PD, such as *VPS35*, *GBA* and *EIF4G1*. However, they will not be discussed since there are no studies that have investigated their AS involvement in the disease.

5.1. *SNCA*

Alpha-synuclein is a small, natively unfolded presynaptic protein, which aggregate to form Lewy bodies, the neuropathological hallmark lesions of PD [27]. *SNCA* was the first gene to be associated to PD family inheritance. Disease-causing mutations of *SNCA* include genetic variations in the coding regions (Ala53Thr, Ala30Pro, Glu46Lys), single nucleotide substitution in 3' UTR and dose-dependent genomic multiplications (duplications or triplications) [27-29]. Some point mutations in splice donor sites have also been reported (IVS2+9A>C) [30].

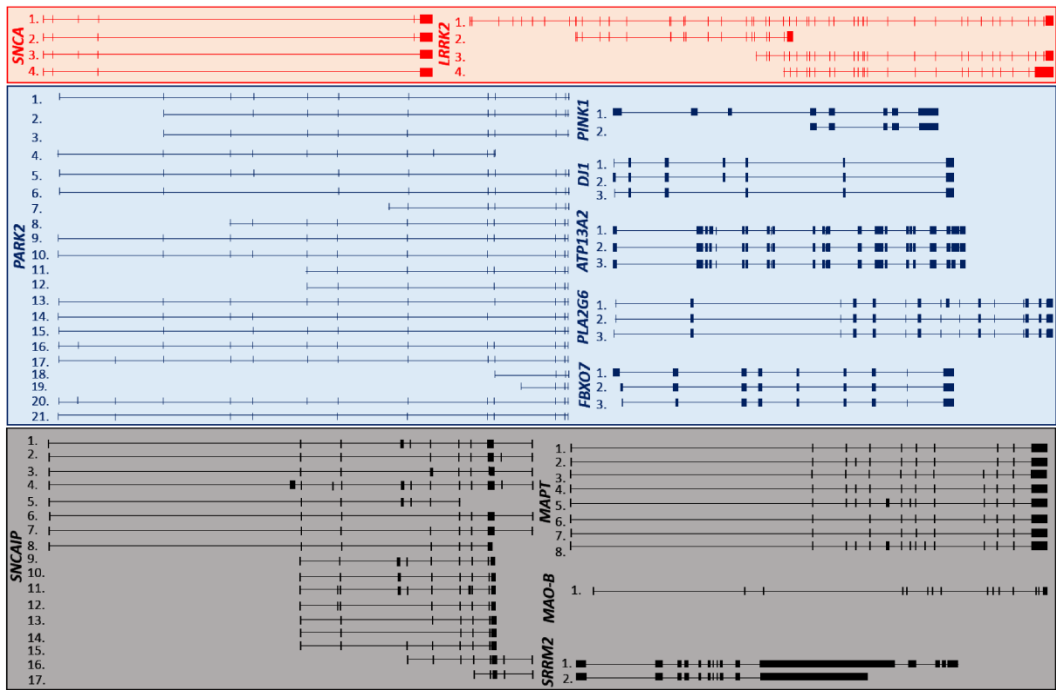


Figure 3. Exonic maps of the alternative splicing variants of PD genes. Exonic structures of the described mRNA splicing variants are reported in figure. Genes are divided in panels (Red panel: autosomal dominant PD genes; Blue panel: autosomal recessive PD genes; Grey panel: minor PD genes). Each variant is indicated with a number on the left corresponding to the Accession Number reported in Table 1. For some genes, additional mRNA transcripts have been deposited in GenBank but their full-length nature has not been determined. For the complete spectrum of collected mRNAs, the reader is referred to Genome Browser Database (<http://genome.ucsc.edu/>) or Ensemble repository (www.ensembl.org).

Table 1. Alternative splice variants of PD genes. PD genes and accession numbers of the relative splice variants are reported in Table. Genes are divided into three groups (autosomal dominant, autosomal recessive and other PD genes). Numbers in the column “#mRNA variant” corresponds to numbers in Figure 3

Gene name	# mRNA variant	Acc. Num.	Protein length	Gene name	# mRNA variant	Acc. Num.	Protein length
AUTOSOMAL DOMINANT PD GENES							
SNCA	1.	NM_000345.3	140 aa	LRRK2	1.	NM_198578.3	2527 aa
	2.	JN709868	126 aa		2.	AK127729	-
	3.	NM_007308.2	112 aa		3.	AL832453	-
	4.	JN709869	98 aa		4.	AK131537	-
AUTOSOMAL RECESSIVE PD GENES							
PARK2	1.	NM_004562.2	465 aa	PINK1	1.	NM_032409.2	581 aa
	2.	AF381282.1	274 aa		2.	BC009534	303 aa
	3.	AF381284.1	203 aa	DJ1	1.	NM_001123377.1	189 aa
	4.	BC022014.2	387 aa		2.	NM_007262.4	189 aa
	5.	NM_013987.2	437 aa		3.	AK130886	-
	6.	NM_013988.2	316 aa	ATP13A2	1.	NM_022089.3	1180 aa
	7.	AK294684.1	139 aa		2.	NM_001141973.2	1175 aa
	8.	GU345837.1	-		3.	NM_001141974.2	1158 aa
	9.	GU345838.1	143 aa	PLA2G6	1.	NM_003560.2	806 aa
	10.	GU345840.1	415 aa		2.	NM_001004426.1	752 aa
	11.	GU357501.1	-		3.	NM_001199562.1	752 aa
	12.	GU357502.1	-	FBXO7	1.	NM_012179.3	522 aa
	13.	GU361466.1	143 aa		2.	NM_001257990.1	408 aa
	14.	GU361467.1	387 aa		3.	NM_001033024.1	443 aa
	15.	GU361468.1	95 aa				
	16.	GU361469.1	51 aa				
	17.	GU361470.1	-				
	18.	GU361471.1	-				
	19.	KC357594.1	34 aa				
	20.	KC357595.1	530 aa				
	21.	KC774171.1	358 aa				

Gene name	# mRNA variant	Acc. Num.	Protein length	Gene name	# mRNA variant	Acc. Num.	Protein length
OTHER PD-RELATED GENES							
<i>SNCAIP</i>	1.	NM_005460.2	919 aa	<i>MAPT</i>	1.	NM_001203251.1	410 aa
	2.	NM_001242935.1	603 aa		2.	NM_001203252.1	410 aa
	3.	AK299687	477 aa		3.	NM_001123067.3	412 aa
	4.	BC040552	1016 aa		4.	NM_005910.5	441 aa
	5.	AK310835	-		5.	NM_016835.4	758 aa
	6.	AK304646	521 aa		6.	NM_016834.4	383 aa
	7.	BC033743	553 aa		7.	NM_016841.4	352 aa
	8.	AK298882	381 aa		8.	NM_001123066.3	776 aa
	9.	AB110788	919 aa	<i>MAO-B</i>	1.	NM_000898.4	520 aa
	10.	AB110789	859 aa				
	11.	AB110790	588 aa				
	12.	AB110791	113 aa	<i>SRRM2</i>	1.	NM_016333.3	2752 aa
	13.	AB110792	66 aa		2.	BC070050	1022 aa
	14.	AB110794	62 aa				
	15.	AB110793	88 aa				
	16.	AK001617	597 aa				
	17.	AK021944	435 aa				

SNCA maps on chromosome 4q22.1 and contains 6 exons spanning about 114 kb [29]. These six exons combine to produce the full-length transcript (named SNCA140 from the amino acid size of the protein product) and at least three additional splicing variants (SNCA126, SNCA112 and SNCA98). These latter are generated by in-frame excision of exons 3, 5, or both (Figure 3, red panel and Table 1). SNCA140, 126 and 112 are expressed in a wide spectrum of human tissues, while SNCA98 is probably a brain-specific splice variant with different expression levels in various fetal and adult brain areas [31].

The expression profile of these splicing variants has been investigated and differ in pathological conditions. All four transcripts are over-expressed in PD *frontal cortex* as compared to healthy controls, with a significant up-regulation of SNCA126 [32]. Furthermore, in the *substantia nigra* of affected patients, SNCA126, SNCA112 and SNCA98 are significantly over-expressed [33, 34], whereas in *cerebellum* are present high levels of SNCA112 and SNCA98 [33]. Different expression profiles of *SNCA* variants occur also in other forms of neurodegenerative disorders. For example, in dementia with Lewy bodies and Alzheimer's disease, both SNCA140 and SNCA126 have been found down-regulated and SNCA98 overexpressed. On the contrary, a disease specific regulation has been reported for SNCA112 variant: up-regulation in dementia with Lewy bodies and downregulation in Alzheimer's disease [32, 35, 36]. Additional RNA transcripts with an extended 3' untranslated region have been described and appear selectively linked to pathological processes [37].

Some fascinating data emerge on the SNCA112 variant. PD risk-associated SNPs within the 3' region of *SNCA* gene have been linked to higher SNCA112 level in about one hundred of *frontal cortex* samples. These data suggest the *cis*-regulatory effect of these mutations on splicing mechanism [38]. Expression of SNCA112 is also induced by some parkinsonism mimetics (MPP⁺, rotenone) and oxidant molecules [39]. However, the reason for this biological effect remains obscure.

As anticipated, *SNCA* transcripts produce distinct protein isoforms. The longest transcript SNCA140 encodes for a small protein composed of three separate domains: (1) amphipathic helices at the N-terminal domain conferring the affinity to bind membranes, (2) a central hydrophobic region, named NAC (non-Ab component), which confers the β -sheet potential, and (3) a glutamatergic carboxyl C-terminal domain that is highly negatively charged and prone to be unstructured [27]. Structural characteristics of the shorter isoforms can be predicted based on the exclusion of spliced-out exons. SNCA126 has an interruption on the N-terminal protein-membrane interaction domain [40], SNCA112 is significantly shorter in the unstructured C-terminal region [40], while SNCA98 isoform is a truncated protein consisting almost only of the central region containing NAC [31]. Recently, it has been verified *in vitro* a lower aggregation propensity of the shorter isoforms [41]. In addition, studies with electron microscopy have revealed a different structural organization of these isoforms: SNCA140 proteins aggregate in straight fibrils, SNCA126 isoforms form short fibrils arranged in parallel arrays, while SNCA98 are combined in circular structures [41]. These data may open new perspectives regarding the formation of Lewy bodies induced by alpha-synuclein.

Various functions of alpha-synuclein have been suggested, such as molecular chaperone, regulator of dopamine uptake and homeostasis, inhibitor of phospholipase D2, downregulator of p53 pathway [40] and promoter of the SNARE-complex assembling [42]. Unfortunately, nothing is known about the specific pathophysiological roles of each alpha-synuclein isoform and their relative post-translational modifications (i.e., phosphorylation, nitration, sumoylation,

oxidation, glycosylation, cleavage and ubiquitination), which are known to play significant roles in SNCA functions and regulation [40].

5.2. *LRRK2*

LRRK2 encodes a large 2,527 amino-acid multi-domain protein called leucine-rich repeat kinase 2. The protein contains multiple conserved domains such as a GTPase-like domain (Ras of complex proteins or ROC), the C-terminal of ROC (COR), a kinase domain, and several protein interaction domains (e.g., the leucine-rich repeat - LRR, the WD40 domain, the ankyrin repeat domain and the armadillo repeat region). Although the precise physiological function of *LRRK2* is unknown, it is probably implicated in different cellular functions such as neurite outgrowth, cytoskeletal maintenance, vesicle trafficking and autophagic protein degradation [43].

LRRK2 harbours the most frequent mutations linked to both autosomal dominant inherited late-onset and sporadic PD (i.e., Gly2019Ser and mutations altering codon Arg1441). In addition, several other mutations affecting splice sites have been also observed (IVS19+5_8delGTAA, IVS25-8delT, IVS27-9C>T, IVS30-6C>T, IVS31+3A>G, IVS32+14G>A, IVS33+6T>A, IVS37-9A>G, IVS38+7C>T, IVS46-14T>A, IVS46-8delT) but their pathogenetic role is still under debate [30, 44-51].

LRRK2 transcribes the full-length mRNA and some shorter transcripts, which are composed of a set of exons matching either the central or the final region of the longest form (Figure 3, red panel and Table 1).

Recently, the expression profile of the *LRRK2* splice variants has been studied in the brain of healthy humans by using exon array and RT-PCR methods [52]. Both experiments have detected in *substantia nigra* a transcript without exons 32–33, whereas in *occipital cortex*, *medulla* and *cerebellum* a variant missing exon 32 [52].

Further evidences on *LRRK2* splicing have been observed by Giesert and collaborators [53], who have identified two *LRRK2* splice-variants in mouse brain: one with skipped exon 5, mainly expressed in astrocytes, and another truncated variant terminating with an alternative exon 42a, barely detectable in microglia but extremely expressed in neurons and astrocytes. Protein-structure predictions reveal that the loss of exon 5 may generate a smaller protein with changed affinity for binding partners, whereas the alternative exon 42a may lead to modifications of its enzymatic activity and loss of WD40 domain. Interestingly, the deletion of this domain in the Zebrafish *LRRK2* ortholog (z*LRRK2*) causes a Parkinsonism-like phenotype including loss of diencephalon dopaminergic neurons and locomotion defects [54]. Further studies will need to assess the involvement of *LRRK2* alternative splice variants in PD.

6. ALTERNATIVE SPLICING OF GENES INVOLVED IN THE RECESSIVELY INHERITED FORMS OF PARKINSON'S DISEASE

Some monogenic forms of PD are inherited with an autosomal recessive pattern and, frequently, symptoms have an early onset. In these disorders, mutations in both alleles (either homozygous or compound heterozygous) cause the pathological phenotype [26].

The recessive cases of PD can be clinically classified in typical and atypical forms. Atypical parkinsonism includes a variety of neurological disorders in which patients have some clinical features of PD, but the symptoms are caused not only by neuronal loss in the *substantia nigra*, but also by additional degeneration of cells in other parts of the nervous system such as the striatum.

Here, we will describe the current literature data regarding genes involved in autosomal recessive PD and their AS regulation. We will include genes causing the typical (*PARK2*, *PINK1*, *DJ1*) and the atypical (*ATP13A2*, *PLA2G6*, *FBXO7*) forms of PD.

6.1. *PARK2*

Homozygous or compound heterozygous mutations in *PARK2* (also known as *parkin RBR E3 ubiquitin protein ligase*) are responsible for the 50% cases of autosomal recessive juvenile Parkinsonism (AR-JP), a form of early-onset Parkinsonism characterized by a prolonged response to levodopa and a benign slow course. *PARK2* mutations also explain ~15% of the PD sporadic cases with onset before 45 [55, 56] and are susceptibility factors for late-onset forms (2% of cases) [57]. Along with the about 200 mutations currently identified in *PARK2* coding region, several other point mutations in splice acceptor or donor sites (introns 1, 6, 7, 10, 12, 13 and 16) have been reported [30, 58-64].

PARK2 spans more than 1.38 Mb of genomic DNA in the long arm of chromosome 6 (6q25.2-q27) and contains 17 exons, which are alternatively spliced to produce at least 21 different splicing variants (Figure 3, blue panel and Table 1) [64]. The full length *PARK2* transcript encodes a protein of 465 amino acid [64-66] involved in numerous molecular pathways (protein turnover, stress response, mitochondria homeostasis, mitophagy, mitochondrial DNA stability, metabolism, cell growth and survival) [67]. The expression of multiple parkin isoforms, likely generated by *PARK2* variants, have been observed in different brain areas [68].

The extensive AS of *PARK2* is differently regulated both at transcript and protein level [32, 69-74]. Distinct expression pattern of *PARK2* splice variants has been observed in different tissues and organs including brain and leukocytes [73-77]. *PARK2* protein isoforms show also a differential distribution in human leukocytes [78] and aged brain [79], as well as in different rat and mouse nervous system areas (cerebral cortex/diencephalons, hippocampus, cerebellum, brainstem, striatum, spinal cord, *substantia nigra*) and peripheral tissues (heart, liver, spleen, pancreas, kidney) [80-84]. In addition, a specific expression profile of *PARK2* isoforms have been detected in different tumour tissues [85, 86].

Emerging evidences support the importance of *PARK2* splice variants expression changes in the development of some neurological disease. Differential expression of *PARK2* transcripts have been identified in *frontal cortex* of Parkinson's disease, pure dementia with Lewy bodies, common Lewy body disease and Alzheimer's disease patients as compared to healthy controls [32, 71]. Particularly, a study reports two *PARK2* splicing variants significantly overexpressed in PD [71], whereas another one describes both an increase in the expression level of a parkin splice variant with a simultaneous decrease of the full length transcript in affected PD as compared healthy subjects [72]. The differential and disease-specific expression profiles of *PARK2* transcripts suggest a role of AS mechanism in the development of neurodegenerative disorders.

6.2. *PINK1*

Mutations in *PINK1* (*PTEN-Induced putative kinase 1*) are the second most frequent cause of autosomal recessive early-onset Parkinsonism. They include homozygous or compound heterozygous changes, with a frequency variable from 1 to 9% depending based on the ethnic background [87]. The *PINK1* mutations spectrum involves non sense and missense mutations, insertions or deletions, and whole-gene or single/multiple exon copy number variants located across the entire gene [88].

PINK1 maps on the short arm of chromosome 1 (1p36.12), encompassing ~18 kb of genomic DNA. Its coding sequence is spread over eight exons. In addition to the full-length, only one shorter variant seems to exist matching with the last five exons (Figure 3, blue panel and Table 1).

Some interesting findings emerge regarding the splicing regulation of exon 7. A 23bp deletion disrupting the splice acceptor site of exon 7 has been detected in a sporadic parkinsonian patient, producing several aberrant mRNAs [89]. Moreover, whole exon 7 deletion and a novel U1-dependent 5' splice-site mutation in exon 7 have been found in a large Spanish family with PD members [90].

The *PINK1* protein is a putative serine/threonine kinase of 581 amino acids involved in mitochondrial response to cellular and oxidative stress [91]. In human brain, at least two *PINK1* isoforms are expressed: the full-length protein of ~63 kDa and the N-terminally truncated isoform of 52 kDa [87, 92-94]. An additional isoform of approximately 45 kDa has been detected, although it has not been further investigated [95].

6.3. *DJ1*

Mutations in *DJ1* (also known as *PARK7*) are the less common cause of autosomal recessive Parkinsonism (~1% of early-onset PD) [21, 96]. The first identified mutations were a large homozygous deletion and a missense mutation (L166P) in Italian and Dutch consanguineous families [97, 98]. Subsequently, other alterations including missense mutations in coding and promoter regions, frame-shifts, copy number variations [21, 99] and splice sites changes have been detected [100, 101].

DJ-1 maps on chromosome 1 (1p36.23) and covers seven exons. Two splice transcript variants, differing in the 5' donor site of exon 1, have been identified but encode the same protein. Furthermore, a third cDNA lacking exon 4 and generating a shorter isoform has been isolated (Figure 3, blue panel and Table 1).

The product of *DJ-1* is an highly conserved multifunctional protein belonging to the peptidase C56 family [102]. It acts as positive regulator of transcription, redox-sensitive chaperone, sensor for oxidative stress, and apparently protects neurons from ROS-induced apoptosis [103, 104]. Several *DJ-1* isoforms, which differ for the isoelectric point (pI), have been observed in human brain and peripheral blood [105-108]. However, it has been postulated that the different pI of each protein arises from post-translational modifications of the full-length protein [109]. The expression of these isoforms seems to be altered in PD, and therefore, they have been proposed as potential peripheral blood biomarkers of PD [109].

Interestingly, one of the major binding partners of *DJ-1* in dopaminergic neuronal cells is the splicing factor proline/glutamine-rich (SFPQ protein) [104, 110]. SFPQ, originally

identified as a polypyrimidine-tract binding protein, is part of the spliceosome C complex and is required for *in vitro* splicing of pre-mRNA [104, 110]. DJ-1 binding to SFPQ modulates its transcriptional activity and, therefore, tunes its effect on splicing regulation. DJ-1 mutations could reverberate on its downstream targets, including the splicing factor SFPQ and altering the splicing control.

6.4. *ATP13A2*

ATP13A2 is composed of 29 exons and lies on chromosome 1 covering about 26 kb of genomic DNA. Its mutations are associated with a form of recessively levodopa-responsive inherited atypical Parkinsonism (the Kufor-Rakeb syndrome) [111]. One of the first identified disease-causing mutations was a guanine-to-adenine transition in the donor splice site of exon 13, leading to the skipping of exon 13 and resulting in an deletion of part of the third transmembrane domain [112].

According to databases, at least three transcripts are expressed in humans (Figure 3, blue panel and Table 1). Splice variants 1 and 2 differ only for a nucleotide segment on exon 5, while transcript 3 lacks exons 22 and 28. The *ATP13A2* mRNA is highly expressed in the brain, particularly in the *substantia nigra* of patients with classical late-onset PD [99].

Isoform 1 is a protein of 1180 amino acids with 10 transmembrane domains. Isoform 2 contains a five amino acid deletion near the N-terminus, while isoform 3 contains two deletions, generating a highly diverged C-terminus [112]. Functional studies have shown that the isoform 1 is located in the lysosome membrane, whereas the isoform 3 protein is retained in the endoplasmic reticulum and rapidly degraded by the proteasome. In addition, both isoform 1 and 3 are eliminated via the endoplasmic-reticulum-associated degradation pathway [112]. The full length protein has been recently identified as a potent modifier of the toxicity induced by alpha-synuclein [113].

6.5. *PLA2G6*

Mutations in *PLA2G6* (*phospholipase A2 group VI*) have been recently associated to a particular parkinsonian phenotype consisting of levodopa-responsive dystonia, pyramidal signs and cognitive/psychiatric features with onset in early adulthood [114]. In particular, the c.1077G>A mutation in the terminal nucleotide of exon 7 (apparently a synonymous mutation) stands out as a cause of abnormal mRNA splicing, causing the activation of a cryptic splice site and producing a 4-bp deleted transcript with different frame-shift in leukocytes [114].

PLA2G6 maps on chromosome 22, covering 70 kb of genomic DNA. Several transcript variants have been described up to now, but the sequences of only three of them have been determined (Figure 3, blue panel and Table 1). The longest *PLA2G6* mRNA includes 17 exonic regions and encodes the 85/88 kDa calcium-independent phospholipase, known as A2 isoform a. The other two transcripts differ in the start point, both lacking of exon 9 and encoding the same protein, called isoform b. The expression profile of *PLA2G6* variants in healthy and disease states is still unknown.

6.6. *FBXO7*

Parkinsonian-Pyramidal Disease (PPD or PARK15-associated parkinsonism) is an autosomal recessive neurodegenerative disease with juvenile onset and additional pyramidal signs. It is caused by mutations in *FBXO7* (*F-box only protein 7*). The pathogenic mutations include some missense nucleotide changes (R378G, R498X, T22M) and a compound heterozygous mutation (IVS7 + 1G/T) responsible of the splice donor removal in intron 7 which leads to an aberrant *FBXO7* mRNA [115-117].

FBXO7 contains nine exons and maps on chromosome 22q12.3. It encodes a 522 amino acids protein consisting of several domains targeting proteins for ubiquitination [116]. Several mRNAs have been observed, but only three of them have been completely characterized (Figure 3, blue panel and Table 1) [115]. Transcript 1 is the longest and the most abundant form ubiquitously expressed [118], particularly in skin fibroblasts [119]. Splice variants 2 and 3, both arising from an inner alternative exon 1, differ for the start codon and produce two isoforms of 443 and 408 amino acids, respectively. Both isoforms 1 and 2 have been detected in cells [119], whereas the expression of the third one remains unknown.

7. ALTERNATIVE SPLICING OF OTHER PD GENES

In addition to the above-mentioned pathogenic genes, we will also briefly discuss some others genes (*SRRM2*, *MAO-B*, *SNCAIP*, *MAPT*) implicated in PD pathways and whose splicing regulation is altered during the disease.

7.1. *SNCAIP*

SNCAIP maps on chromosome 5q23.2 and spans about 152 kb of genomic DNA. It encodes Synphilin-1, a presynaptic protein made up of ankyrin-like repeats, a coiled-coil domain and an ATP/GTP-binding domain [120]. Synphilin-1 interacts strongly with alpha-synuclein in neuronal tissue and seems to play a role in the formation of Lewy bodies during neurodegeneration.

The most studied alternative splice variants (Figure 3, grey panel and Table 1) are synphilin-1 and 1A. The first is the full-length transcript, while the second lacks exons 4 and 5, and contains an extra exon located between exons 10 and 11. Synphilin-1A isoform is involved in the pathogenesis of PD and may contribute to the formation of Lewy bodies [121-123]. Accordingly, this protein shows enhanced aggregatory properties causing neuronal toxicity [121-123].

The mRNA expression levels of synphilin 1, 1A and other two additional synphilin variants are simultaneously overexpressed in the *frontal cortex* of PD patients as compared to healthy controls [32, 71].

7.2. *MAPT*

MAPT, which encodes the microtubule-associated protein tau [124], is located on chromosome 17q21 and contains 15 exons. It produces multiple splice transcripts (Figure 3, grey panel and Table 1) which are differentially expressed in human tissues [7]. Specifically, in the adult human central nervous system, *MAPT* splicing is able to create six tau isoforms composed of either three or four microtubule-binding repeat motifs in the C-terminal (3R and 4R-tau).

Several known mutations within and around *MAPT* exon 10 interrupt the exonic and intronic splicing elements as well as the formation of an RNA stem-loop structure at the 5' splice site (which normally functions to restrict spliceosome assembly), resulting in an altered ratio of 3R/4R isoforms [17, 19]. The alteration of this balance causes the hyperphosphorylation and aggregation of tau proteins into neurofibrillary tangles, triggering the frontotemporal dementia with Parkinsonism linked to chromosome 17 (FTDP-17) [17, 19]. These data support a direct relationship between aberrant alternative splicing of tau and neuropathology.

7.3. *MAO-B*

MAO-B gene is located on chromosome X and includes 15 exons (Figure 3, grey panel and Table 1). Several SNPs have been described in populations with different ethnic background [125]. A common SNP associated with two-fold risk of PD is the G/A dimorphism in intron 13 sequence [126-128]. This SNP does not change the coding sequence, and does not affect the consensus acceptor and donor sites. However, it creates a splicing enhancer that stimulates intron 13 removal, spliceosomal complex assembly and modifies the binding site efficiency of splicing factors [125].

7.4. *SRRM2*

Serine/arginine (SR) proteins belong to the class of alternative splicing *trans*-acting factors. One of the SR proteins is the RNA splicing factor SRRM2 (or serine/arginine repetitive matrix 2), which has been identified as the only gene that stood out as differentially expressed in multiple gene expression PD datasets [129].

SRRM2 generates two main alternative splicing transcripts: the full-length isoform containing 15 exons, and the shorter isoform, which contains 11 exons and lacks exons 12–15 (Figure 3, grey panel and Table 1). These two isoforms are differentially expressed in postmortem PD brain [129]. The shorter transcript is upregulated in *substantia nigra* but unchanged in the *amygdala* of PD patients versus healthy controls. On the contrary, the longer transcript is downregulated in both *substantia nigra* and *amygdala* of PDs as compared to controls [129]. In addition, in the peripheral blood of affected patients, the shortest isoform is overexpressed, whereas the longest one is reduced [129].

8. WHOLE-GENOME RNA EXPRESSION STUDIES REVEAL OVERALL ALTERNATIVE SPLICING CHANGES IN PD

Although a “gene by gene” approach may simplify the analysis, global AS changes in PD have to be considered. The majority of large scale gene-expression array studies in the brain of affected patients have actually looked at the full length transcript of each gene by ignoring their multiple splice variants [22]. Nevertheless, AS is significantly altered in cortical neurons of PD patients [130].

In order to specifically explore the splicing expression changes, some studies have used exon arrays which enable an improved monitoring and detection of the AS events. These studies have allowed to observe significant changes of transcript expression profiles in PD blood cells as compared to healthy controls [129, 131]. A similar study was conducted in blood samples of advanced PD patients prior to and following deep brain stimulation which efficiently improves the motor symptoms of PD [132]. This study showed that neurosurgical intervention alters transcripts profiles [132]. Specific splicing variant microarrays have also been used in PD patients in order to detect mRNAs splice variants as molecular biomarkers for an early PD diagnosis. This analysis has allowed to identify 13 splice variants with an altered expression in early-stage PD patients versus healthy controls [133, 134].

Along with microarray technology, AS variants can now be better characterized through RNA deep sequencing (RNAseq) [22]. Whole transcriptome RNAseq data have been obtained from blood leukocytes of PD patients' pre- and post- deep brain stimulation treatment [135]. This method has enabled to identify a large variety of differential splicing events pre and post treatment as well as novel human exons and junctions in protein-coding mRNA transcripts [135]. Based on this first study, RNAseq represents a promising technique to explore PD alternative splicing.

9. EPIGENETIC REGULATION OF PD ALTERNATIVE SPLICING

A growing body of evidence are supporting the role of epigenetics in the development and progression of many neurodegenerative diseases including PD. Epigenetic modifications refer to changes in gene expression through DNA methylation, histones modifications and non-coding RNAs [136]. This latter class of molecules include both long non-coding RNAs (lncRNAs) and microRNAs (miRNAs), which have been suggested as potential modulator factors of AS mechanism in PD.

lncRNAs are described as transcripts longer than 200 nucleotides, and include several spliced transcript collected in the GENCODE non-coding RNA database. lncRNAs expression profiling has been recently investigated in PD leukocytes pre- and post-deep brain stimulation via RNAseq [135]. This study has allowed to identify some lncRNAs overexpressed in PD and inversely decreased following deep brain stimulation [135]. Differentially expressed lncRNA include the spliceosome component U1, supporting the hypothesis of disease-involved splicing modulations [135].

A great number of alternatively spliced exons have also been predicted as binding sites of microRNAs (miRNAs). These are a class of very small non-coding RNA molecules (about 20-22 bp), mainly acting as posttranscriptional modulators of multiple target genes by partial

sequence complementarity. Through this mechanism, they may also influence the activity of the splicing machinery. The interaction between miRNA differential expression and alternative splicing variations in PD has been recently explored [137]. Parallel changes in miRNA profiles and their spliced targets have been detected in PD peripheral leukocytes and in PD-relevant brain regions (including the *substantia nigra* and the *frontal lobe*) through coupled analysis of small RNA sequencing data, splice junction arrays and exon arrays [137].

The characterization of existing links between non-coding mRNAs and AS modifications represents a main step to comprehend the molecular basis of PD.

CONCLUSION

AS is an extremely harmonized process, established by the combination of specific DNA sequences, intronic and exonic elements, regulatory factors and temporal and spatial signaling pathways. Mutations that alter any of these elements may modify the finely regulated splicing processes by varying the assembling or the functions of the encoded proteins and lastly generating the diseases. The comprehension of the role AS in PD may represent an essential point to decipher etiology of this disease. Future studies, both with the standard or the new large-scale methods, will offer a complete data pool of the AS events in PD and will provide new possible insights in order to improve strategies for PD treatment and diagnosis.

ACKNOWLEDGMENTS

This work was supported by the Italian Ministry of Education, Universities and Research through grant CTN01_00177_817708 and the international Ph.D. program in Neuroscience of the University of Catania. We also gratefully acknowledge Cristina Calì, Alfia Corsino, Maria Patrizia D'Angelo and Francesco Marino for their administrative and technical support.

REFERENCES

- [1] Ward, AJ; Cooper, TA. The pathobiology of splicing. *The Journal of pathology*, 2010, 220 (2), 152-63.
- [2] Yap, K; Makeyev, EV. Regulation of gene expression in mammalian nervous system through alternative pre-mRNA splicing coupled with RNA quality control mechanisms. *Molecular and cellular neurosciences*, 2013, 56, 420-8.
- [3] Calarco, JA; Zhen, M; Blencowe, BJ. Networking in a global world: establishing functional connections between neural splicing regulators and their target transcripts. *Rna*, 2011, 17 (5), 775-91.
- [4] Li, Q; Lee, JA; Black, DL. Neuronal regulation of alternative pre-mRNA splicing. *Nature reviews. Neuroscience*, 2007, 8 (11), 819-31.
- [5] Pan, Q; Shai, O; Lee, LJ; Frey, BJ; Blencowe, BJ. Deep surveying of alternative splicing complexity in the human transcriptome by high-throughput sequencing. *Nature genetics*, 2008, 40 (12), 1413-5.

- [6] Behzadnia, N; Golas, MM; Hartmuth, K; Sander, B; Kastner, B; Deckert, J; Dube, P; Will, CL; Urlaub, H; Stark, H; Luhrmann, R. Composition and three-dimensional EM structure of double affinity-purified, human prespliceosomal A complexes. *The EMBO journal*, 2007, 26 (6), 1737-48.
- [7] Tazi, J; Bakkour, N; Stamm, S. Alternative splicing and disease. *Biochimica et biophysica acta*, 2009, 1792 (1), 14-26.
- [8] Stamm, S; Ben-Ari, S; Rafalska, I; Tang, Y; Zhang, Z; Toiber, D; Thanaraj, TA; Soreq, H. Function of alternative splicing. *Gene*, 2005, 344, 1-20.
- [9] Hughes, TA. Regulation of gene expression by alternative untranslated regions. *Trends Genet*, 2006, 22 (3), 119-22.
- [10] Lejeune, F; Maquat, LE. Mechanistic links between nonsense-mediated mRNA decay and pre-mRNA splicing in mammalian cells. *Curr Opin Cell Biol*, 2005, 17 (3), 309-15.
- [11] Modrek, B; Lee, C. A genomic view of alternative splicing. *Nat Genet*, 2002, 30 (1), 13-9.
- [12] Kelemen, O; Convertini, P; Zhang, Z; Wen, Y; Shen, M; Falaleeva, M; Stamm, S. Function of alternative splicing. *Gene*, 2012.
- [13] Yeo, G; Holste, D; Kreiman, G; Burge, CB. Variation in alternative splicing across human tissues. *Genome biology*, 2004, 5 (10), R74.
- [14] Raj, B; Blencowe, BJ. Alternative Splicing in the Mammalian Nervous System: Recent Insights into Mechanisms and Functional Roles. *Neuron*, 2015, 87 (1), 14-27.
- [15] Norris, AD; Calarco, JA. Emerging Roles of Alternative Pre-mRNA Splicing Regulation in Neuronal Development and Function. *Frontiers in neuroscience*, 2012, 6, 122.
- [16] Mazin, P; Xiong, J; Liu, X; Yan, Z; Zhang, X; Li, M; He, L; Somel, M; Yuan, Y; Phoebe Chen, YP; Li, N; Hu, Y; Fu, N; Ning, Z; Zeng, R; Yang, H; Chen, W; Gelfand, M; Khaitovich, P. Widespread splicing changes in human brain development and aging. *Molecular systems biology*, 2013, 9, 633.
- [17] Licatalosi, DD; Darnell, RB. Splicing regulation in neurologic disease. *Neuron*, 2006, 52 (1), 93-101.
- [18] Singh, RK; Cooper, TA. Pre-mRNA splicing in disease and therapeutics. *Trends in molecular medicine*, 2012, 18 (8), 472-82.
- [19] Cooper, TA; Wan, L; Dreyfuss, G. RNA and disease. *Cell*, 2009, 136 (4), 777-93.
- [20] Feng, D; Xie, J. Aberrant splicing in neurological diseases. Wiley interdisciplinary reviews. *RNA*, 2013, 4 (6), 631-49.
- [21] Moore, DJ; West, AB; Dawson, VL; Dawson, TM. Molecular pathophysiology of Parkinson's disease. *Annual review of neuroscience*, 2005, 28, 57-87.
- [22] Lewis, PA; Cookson, MR. Gene expression in the Parkinson's disease brain. *Brain research bulletin*, 2012, 88 (4), 302-12.
- [23] Bonifati, V. Genetics of Parkinson's disease--state of the art, 2013. *Parkinsonism & related disorders*, 2014, 20 Suppl 1, S23-8.
- [24] Lubbe, S; Morris, HR. Recent advances in Parkinson's disease genetics. *Journal of neurology*, 2014, 261 (2), 259-66.
- [25] Trinh, J; Farrer, M. Advances in the genetics of Parkinson disease. *Nature reviews. Neurology*, 2013, 9 (8), 445-54.
- [26] Klein, C; Westenberger, A. Genetics of Parkinson's disease. *Cold Spring Harbor perspectives in medicine*, 2012, 2 (1), a008888.

- [27] Stefanis, L. alpha-Synuclein in Parkinson's disease. *Cold Spring Harbor perspectives in medicine*, 2012, 2 (2), a009399.
- [28] Pihlstrom, L; Toft, M. Genetic variability in SNCA and Parkinson's disease. *Neurogenetics*, 2011, 12 (4), 283-93.
- [29] Deng, H; Yuan, L. Genetic variants and animal models in SNCA and Parkinson disease. *Ageing research reviews*, 2014, 15C, 161-176.
- [30] Nuytemans, K; Meeus, B; Crosiers, D; Brouwers, N; Goossens, D; Engelborghs, S; Pals, P; Pickut, B; Van den Broeck, M; Corsmit, E; Cras, P; De Deyn, PP; Del-Favero, J; Van Broeckhoven, C; Theuns, J. Relative contribution of simple mutations vs. copy number variations in five Parkinson disease genes in the Belgian population. *Human mutation*, 2009, 30 (7), 1054-61.
- [31] Beyer, K; Domingo-Sabat, M; Lao, JI; Carrato, C; Ferrer, I; Ariza, A. Identification and characterization of a new alpha-synuclein isoform and its role in Lewy body diseases. *Neurogenetics*, 2008, 9 (1), 15-23.
- [32] Beyer, K; Domingo-Sabat, M; Humbert, J; Carrato, C; Ferrer, I; Ariza, A. Differential expression of alpha-synuclein, parkin, and synphilin-1 isoforms in Lewy body disease. *Neurogenetics*, 2008, 9 (3), 163-72.
- [33] McLean, JR; Hallett, PJ; Cooper, O; Stanley, M; Isacson, O. Transcript expression levels of full-length alpha-synuclein and its three alternatively spliced variants in Parkinson's disease brain regions and in a transgenic mouse model of alpha-synuclein overexpression. *Molecular and cellular neurosciences*, 2012, 49 (2), 230-9.
- [34] Cardo, LF; Coto, E; de Mena, L; Ribacoba, R; Mata, IF; Menendez, M; Moris, G; Alvarez, V. Alpha-synuclein transcript isoforms in three different brain regions from Parkinson's disease and healthy subjects in relation to the SNCA rs356165/rs11931074 polymorphisms. *Neuroscience letters*, 2014, 562, 45-9.
- [35] Beyer, K; Lao, JI; Carrato, C; Mate, JL; Lopez, D; Ferrer, I; Ariza, A. Differential expression of alpha-synuclein isoforms in dementia with Lewy bodies. *Neuropathology and applied neurobiology*, 2004, 30 (6), 601-7.
- [36] Beyer, K; Humbert, J; Ferrer, A; Lao, JI; Carrato, C; Lopez, D; Ferrer, I; Ariza, A. Low alpha-synuclein 126 mRNA levels in dementia with Lewy bodies and Alzheimer disease. *Neuroreport*, 2006, 17 (12), 1327-30.
- [37] Rhinn, H; Qiang, L; Yamashita, T; Rhee, D; Zolin, A; Vanti, W; Abeliovich, A. Alternative alpha-synuclein transcript usage as a convergent mechanism in Parkinson's disease pathology. *Nature communications*, 2012, 3, 1084.
- [38] McCarthy, JJ; Linnertz, C; Saucier, L; Burke, JR; Hulette, CM; Welsh-Bohmer, KA; Chiba-Falek, O. The effect of SNCA 3' region on the levels of SNCA-112 splicing variant. *Neurogenetics*, 2011, 12 (1), 59-64.
- [39] Kalivendi, SV; Yedlapudi, D; Hillard, CJ; Kalyanaraman, B. Oxidants induce alternative splicing of alpha-synuclein: Implications for Parkinson's disease. *Free radical biology & medicine*, 2010, 48 (3), 377-83.
- [40] Beyer, K. Alpha-synuclein structure, posttranslational modification and alternative splicing as aggregation enhancers. *Acta neuropathologica*, 2006, 112 (3), 237-51.
- [41] Bungereoth, M; Appenzeller, S; Regulin, A; Volker, W; Lorenzen, I; Grotzinger, J; Pendziwiat, M; Kuhlenbaumer, G. Differential aggregation properties of alpha-synuclein isoforms. *Neurobiology of aging*, 2014, 35 (8), 1913-9.

- [42] Burre, J; Sharma, M; Tsetsenis, T; Buchman, V; Etherton, MR; Sudhof, TC. Alpha-synuclein promotes SNARE-complex assembly in vivo and in vitro. *Science*, 2010, 329 (5999), 1663-7.
- [43] Rideout, HJ; Stefanis, L. The neurobiology of LRRK2 and its role in the pathogenesis of Parkinson's disease. *Neurochemical research*, 2014, 39 (3), 576-92.
- [44] Johnson, J; Paisan-Ruiz, C; Lopez, G; Crews, C; Britton, A; Malkani, R; Evans, EW; McInerney-Leo, A; Jain, S; Nussbaum, RL; Foote, KD; Mandel, RJ; Crawley, A; Reimsnider, S; Fernandez, HH; Okun, MS; Gwinn-Hardy, K; Singleton, AB. Comprehensive screening of a North American Parkinson's disease cohort for LRRK2 mutation. *Neuro-degenerative diseases*, 2007, 4 (5), 386-91.
- [45] Skipper, L; Shen, H; Chua, E; Bonnard, C; Kolatkar, P; Tan, LC; Jamora, RD; Puvan, K; Puong, KY; Zhao, Y; Pavanni, R; Wong, MC; Yuen, Y; Farrer, M; Liu, JJ; Tan, EK. Analysis of LRRK2 functional domains in nondominant Parkinson disease. *Neurology*, 2005, 65 (8), 1319-21.
- [46] Zabetian, CP; Samii, A; Mosley, AD; Roberts, JW; Leis, BC; Yearout, D; Raskind, WH; Griffith, A. A clinic-based study of the LRRK2 gene in Parkinson disease yields new mutations. *Neurology*, 2005, 65 (5), 741-4.
- [47] Shojaei, S; Sina, F; Farboodi, N; Fazlali, Z; Ghazavi, F; Ghorashi, SA; Parsa, K; Sadeghi, H; Shahidi, GA; Ronaghi, M; Elahi, E. A clinic-based screening of mutations in exons 31, 34, 35, 41, and 48 of LRRK2 in Iranian Parkinson's disease patients. *Movement disorders: official journal of the Movement Disorder Society*, 2009, 24 (7), 1023-7.
- [48] Di Fonzo, A; Tassorelli, C; De Mari, M; Chien, HF; Ferreira, J; Rohe, CF; Riboldazzi, G; Antonini, A; Albani, G; Mauro, A; Marconi, R; Abbruzzese, G; Lopiano, L; Fincati, E; Guidi, M; Marini, P; Stocchi, F; Onofri, M; Toni, V; Tinazzi, M; Fabbrini, G; Lamberti, P; Vanacore, N; Meco, G; Leitner, P; Uitti, RJ; Wszolek, ZK; Gasser, T; Simons, EJ; Breedveld, GJ; Goldwurm, S; Pezzoli, G; Sampaio, C; Barbosa, E; Martignoni, E; Oostra, BA; Bonifati, V; Italian Parkinson's Genetics, N. Comprehensive analysis of the LRRK2 gene in sixty families with Parkinson's disease. *European journal of human genetics : EJHG*, 2006, 14 (3), 322-31.
- [49] Grimes, DA; Racacho, L; Han, F; Panisset, M; Bulman, DE. LRRK2 screening in a Canadian Parkinson's disease cohort. *The Canadian journal of neurological sciences. Le journal canadien des sciences neurologiques*, 2007, 34 (3), 336-8.
- [50] Paisan-Ruiz, C; Nath, P; Washecka, N; Gibbs, JR; Singleton, AB. Comprehensive analysis of LRRK2 in publicly available Parkinson's disease cases and neurologically normal controls. *Human mutation*, 2008, 29 (4), 485-90.
- [51] Lesage, S; Condroyer, C; Lannuzel, A; Lohmann, E; Troiano, A; Tison, F; Damier, P; Thobois, S; Ouvrard-Hernandez, AM; Rivaud-Pechoux, S; Brefel-Courbon, C; Destee, A; Tranchant, C; Romana, M; Leclerc, L; Durr, A; Brice, A; French Parkinson's Disease Genetics Study, G. Molecular analyses of the LRRK2 gene in European and North African autosomal dominant Parkinson's disease. *Journal of medical genetics*, 2009, 46 (7), 458-64.
- [52] Trabzuni, D; Ryten, M; Emmett, W; Ramasamy, A; Lackner, KJ; Zeller, T; Walker, R; Smith, C; Lewis, PA; Mamais, A; de Silva, R; Vandrovcova, J; International Parkinson Disease Genomics, C; Hernandez, D; Nalls, MA; Sharma, M; Garnier, S; Lesage, S; Simon-Sanchez, J; Gasser, T; Heutink, P; Brice, A; Singleton, A; Cai, H; Schadt, E;

- Wood, NW; Bandopadhyay, R; Weale, ME; Hardy, J; Plagnol, V. Fine-mapping, gene expression and splicing analysis of the disease associated LRRK2 locus. *PloS one*, 2013, 8 (8), e70724.
- [53] Giesert, F; Hofmann, A; Burger, A; Zerle, J; Kloos, K; Hafen, U; Ernst, L; Zhang, J; Vogt-Weisenhorn, DM; Wurst, W. Expression analysis of Lrrk1, Lrrk2 and Lrrk2 splice variants in mice. *PloS one*, 2013, 8 (5), e63778.
- [54] Sheng, D; Qu, D; Kwok, KH; Ng, SS; Lim, AY; Aw, SS; Lee, CW; Sung, WK; Tan, EK; Lufkin, T; Jesuthasan, S; Sinnakaruppan, M; Liu, J. Deletion of the WD40 domain of LRRK2 in Zebrafish causes Parkinsonism-like loss of neurons and locomotive defect. *PLoS genetics*, 2010, 6 (4), e1000914.
- [55] Bonifati, V. Autosomal recessive parkinsonism. *Parkinsonism & related disorders*, 2012, 18 Suppl 1, S4-6.
- [56] Lucking, CB; Durr, A; Bonifati, V; Vaughan, J; De Michele, G; Gasser, T; Harhangi, BS; Meo, G; Denefle, P; Wood, NW; Agid, Y; Brice, A; French Parkinson's Disease Genetics Study, G; European Consortium on Genetic Susceptibility in Parkinson's, D. Association between early-onset Parkinson's disease and mutations in the parkin gene. *The New England journal of medicine*, 2000, 342 (21), 1560-7.
- [57] Oliveira, SA; Scott, WK; Martin, ER; Nance, MA; Watts, RL; Hubble, JP; Koller, WC; Pahwa, R; Stern, MB; Hiner, BC; Ondo, WG; Allen, FH; Jr. Scott, BL; Goetz, CG; Small, GW; Mastaglia, F; Stajich, JM; Zhang, F; Booze, MW; Winn, MP; Middleton, LT; Haines, JL; Pericak-Vance, MA; Vance, JM. Parkin mutations and susceptibility alleles in late-onset Parkinson's disease. *Annals of neurology*, 2003, 53 (5), 624-9.
- [58] Illarioshkin, SN; Periquet, M; Rawal, N; Lucking, CB; Zagorovskaya, TB; Slominsky, PA; Miloserdova, OV; Markova, ED; Limborska, SA; Ivanova-Smolenskaya, IA; Brice, A. Mutation analysis of the parkin gene in Russian families with autosomal recessive juvenile parkinsonism. *Movement disorders : official journal of the Movement Disorder Society*, 2003, 18 (8), 914-9.
- [59] Pigullo, S; De Luca, A; Barone, P; Marchese, R; Bellone, E; Colosimo, A; Scaglione, C; Martinelli, P; Di Maria, E; Pizzuti, A; Abbruzzese, G; Dallapiccola, B; Ajmar, F; Mandich, P. Mutational analysis of parkin gene by denaturing high-performance liquid chromatography (DHPLC) in essential tremor. *Parkinsonism & related disorders*, 2004, 10 (6), 357-62.
- [60] Scherfler, C; Khan, NL; Pavese, N; Eunson, L; Graham, E; Lees, AJ; Quinn, NP; Wood, NW; Brooks, DJ; Piccini, PP. Striatal and cortical pre- and postsynaptic dopaminergic dysfunction in sporadic parkin-linked parkinsonism. *Brain : a journal of neurology*, 2004, 127 (Pt 6), 1332-42.
- [61] Bertoli-Avella, AM; Giroud-Benitez, JL; Akyol, A; Barbosa, E; Schaap, O; van der Linde, HC; Martignoni, E; Lopiano, L; Lamberti, P; Fincati, E; Antonini, A; Stocchi, F; Montagna, P; Squitieri, F; Marini, P; Abbruzzese, G; Fabbrini, G; Marconi, R; Dalla Libera, A; Trianni, G; Guidi, M; De Gaetano, A; Boff Maegawa, G; De Leo, A; Gallai, V; de Rosa, G; Vanacore, N; Meo, G; van Duijn, CM; Oostra, BA; Heutink, P; Bonifati, V; Italian Parkinson Genetics, N. Novel parkin mutations detected in patients with early-onset Parkinson's disease. *Movement disorders : official journal of the Movement Disorder Society*, 2005, 20 (4), 424-31.

- [62] Bardien, S; Keyser, R; Yako, Y; Lombard, D; Carr, J. Molecular analysis of the parkin gene in South African patients diagnosed with Parkinson's disease. *Parkinsonism & related disorders*, 2009, 15 (2), 116-21.
- [63] Pankratz, N; Kissell, DK; Pauciulo, MW; Halter, CA; Rudolph, A; Pfeiffer, RF; Marder, KS; Foroud, T; Nichols, WC; Parkinson Study Group, PI. Parkin dosage mutations have greater pathogenicity in familial PD than simple sequence mutations. *Neurology*, 2009, 73 (4), 279-86.
- [64] La Cognata, V; Iemmolo, R; D'Agata, V; Scuderi, S; Drago, F; Zappia, M; Cavallaro, S. Increasing the Coding Potential of Genomes Through Alternative Splicing: The Case of PARK2 Gene. *Current genomics*, 2014, 15 (3), 203-16.
- [65] Kitada, T; Asakawa, S; Hattori, N; Matsumine, H; Yamamura, Y; Minoshima, S; Yokochi, M; Mizuno, Y; Shimizu, N. Mutations in the parkin gene cause autosomal recessive juvenile parkinsonism. *Nature*, 1998, 392 (6676), 605-608.
- [66] Matsumine, H; Saito, M; Shimoda-Matsubayashi, S; Tanaka, H; Ishikawa, A; Nakagawa-Hattori, Y; Yokochi, M; Kobayashi, T; Igarashi, S; Takano, H; Sanpei, K; Koike, R; Mori, H; Kondo, T; Mizutani, Y; Schaffer, AA; Yamamura, Y; Nakamura, S; Kuzuhara, S; Tsuji, S; Mizuno, Y. Localization of a gene for an autosomal recessive form of juvenile Parkinsonism to chromosome 6q25.2-27. *Am J Hum Genet*, 1997, 60 (3), 588-96.
- [67] Xu, L; Lin, DC; Yin, D; Koeffler, HP. An emerging role of PARK2 in cancer. *Journal of molecular medicine*, 2014, 92 (1), 31-42.
- [68] Scuderi, S; La Cognata, V; Drago, F; Cavallaro, S; D'Agata, V. Alternative splicing generates different parkin protein isoforms: evidences in human, rat, and mouse brain. *BioMed research international*, 2014, 690796.
- [69] Ikeuchi, K; Marusawa, H; Fujiwara, M; Matsumoto, Y; Endo, Y; Watanabe, T; Iwai, A; Sakai, Y; Takahashi, R; Chiba, T. Attenuation of proteolysis-mediated cyclin E regulation by alternatively spliced Parkin in human colorectal cancers. *International journal of cancer. Journal international du cancer*, 2009, 125 (9), 2029-35.
- [70] Kitada, T; Asakawa, S; Minoshima, S; Mizuno, Y; Shimizu, N. Molecular cloning, gene expression, and identification of a splicing variant of the mouse parkin gene. *Mamm Genome*, 2000, 11 (6), 417-21.
- [71] Humbert, J; Beyer, K; Carrato, C; Mate, JL; Ferrer, I; Ariza, A. Parkin and synphilin-1 isoform expression changes in Lewy body diseases. *Neurobiol Dis*, 2007, 26 (3), 681-7.
- [72] Tan, EK; Shen, H; Tan, JM; Lim, KL; Fook-Chong, S; Hu, WP; Paterson, MC; Chandran, VR; Yew, K; Tan, C; Yuen, Y; Pavanni, R; Wong, MC; Puvan, K; Zhao, Y. Differential expression of splice variant and wild-type parkin in sporadic Parkinson's disease. *Neurogenetics*, 2005, 6 (4), 179-84.
- [73] Dagata, V; Cavallaro, S. Parkin transcript variants in rat and human brain. *Neurochemical research*, 2004, 29 (9), 1715-24.
- [74] Sunada, Y; Saito, F; Matsumura, K; Shimizu, T. Differential expression of the parkin gene in the human brain and peripheral leukocytes. *Neuroscience letters*, 1998, 254 (3), 180-2.
- [75] D'Agata, V; Zhao, W; Pascale, A; Zohar, O; Scapagnini, G; Cavallaro, S. Distribution of parkin in the adult rat brain. *Progress in neuro-psychopharmacology & biological psychiatry*, 2002, 26 (3), 519-27.

- [76] D'Agata, V; Zhao, W; Cavallaro, S. Cloning and distribution of the rat parkin mRNA. *Brain research. Molecular brain research*, 2000, 75 (2), 345-9.
- [77] Solano, SM; Miller, DW; Augood, SJ; Young, AB; Penney, JB; Jr. Expression of alpha-synuclein, parkin, and ubiquitin carboxy-terminal hydrolase L1 mRNA in human brain: genes associated with familial Parkinson's disease. *Annals of neurology*, 2000, 47 (2), 201-10.
- [78] Kasap, M; Akpınar, G; Sazci, A; Idrisoglu, HA; Vahaboglu, H. Evidence for the presence of full-length PARK2 mRNA and Parkin protein in human blood. *Neuroscience letters*, 2009, 460 (3), 196-200.
- [79] Pawlyk, AC; Giasson, BI; Sampathu, DM; Perez, FA; Lim, KL; Dawson, VL; Dawson, TM; Palmiter, RD; Trojanowski, JQ; Lee, VM. Novel monoclonal antibodies demonstrate biochemical variation of brain parkin with age. *The Journal of biological chemistry*, 2003, 278 (48), 48120-8.
- [80] Horowitz, JM; Myers, J; Stachowiak, MK; Torres, G. Identification and distribution of Parkin in rat brain. *Neuroreport*, 1999, 10 (16), 3393-7.
- [81] Stichel, CC; Augustin, M; Kuhn, K; Zhu, XR; Engels, P; Ullmer, C; Lubbert, H. Parkin expression in the adult mouse brain. *The European journal of neuroscience*, 2000, 12 (12), 4181-94.
- [82] Gu, WJ; Abbas, N; Lagunes, MZ; Parent, A; Pradier, L; Bohme, GA; Agid, Y; Hirsch, EC; Raisman-Vozari, R; Brice, A. Cloning of rat parkin cDNA and distribution of parkin in rat brain. *Journal of neurochemistry*, 2000, 74 (4), 1773-6.
- [83] Huynh, DP; Dy, M; Nguyen, D; Kiehl, TR; Pulst, SM. Differential expression and tissue distribution of parkin isoforms during mouse development. *Brain research. Developmental brain research*, 2001, 130 (2), 173-81.
- [84] D'Agata, V; Grimaldi, M; Pascale, A; Cavallaro, S. Regional and cellular expression of the parkin gene in the rat cerebral cortex. *The European journal of neuroscience*, 2000, 12 (10), 3583-8.
- [85] Maugeri, G; D'Amico, AG; Magro, G; Salvatorelli, L; Barbagallo, GM; Saccone, S; Drago, F; Cavallaro, S; D'Agata, V. Expression profile of parkin isoforms in human gliomas. *International journal of oncology*, 2015.
- [86] D'Amico, AG; Maugeri, G; Magro, G; Salvatorelli, L; Drago, F; D'Agata, V. Expression pattern of parkin isoforms in lung adenocarcinomas. *Tumour biology : the journal of the International Society for Oncodevelopmental Biology and Medicine*, 2015, 36 (7), 5133-41.
- [87] Pogson, JH; Ivatt, RM; Whitworth, AJ. Molecular mechanisms of PINK1-related neurodegeneration. *Current neurology and neuroscience reports*, 2011, 11 (3), 283-90.
- [88] Nuytemans, K; Theuns, J; Cruts, M; Van Broeckhoven, C. Genetic etiology of Parkinson disease associated with mutations in the SNCA, PARK2, PINK1, PARK7, and LRRK2 genes: a mutation update. *Human mutation*, 2010, 31 (7), 763-80.
- [89] Marongiu, R; Brancati, F; Antonini, A; Ialongo, T; Ceccarini, C; Scarciolla, O; Capalbo, A; Benti, R; Pezzoli, G; Dallapiccola, B; Goldwurm, S; Valente, EM. Whole gene deletion and splicing mutations expand the PINK1 genotypic spectrum. *Human mutation*, 2007, 28 (1), 98.
- [90] Samaranch, L; Lorenzo-Betancor, O; Arbelo, JM; Ferrer, I; Lorenzo, E; Irigoyen, J; Pastor, MA; Marrero, C; Isla, C; Herrera-Henriquez, J; Pastor, P. PINK1-linked

- parkinsonism is associated with Lewy body pathology. *Brain : a journal of neurology*, 2010, 133 (Pt 4), 1128-42.
- [91] Valente, EM; Abou-Sleiman, PM; Caputo, V; Muqit, MM; Harvey, K; Gispert, S; Ali, Z; Del Turco, D; Bentivoglio, AR; Healy, DG; Albanese, A; Nussbaum, R; Gonzalez-Maldonado, R; Deller, T; Salvi, S; Cortelli, P; Gilks, WP; Latchman, DS; Harvey, RJ; Dallapiccola, B; Auburger, G; Wood, NW. Hereditary early-onset Parkinson's disease caused by mutations in PINK1. *Science*, 2004, 304 (5674), 1158-60.
 - [92] Silvestri, L; Caputo, V; Bellacchio, E; Atorino, L; Dallapiccola, B; Valente, EM; Casari, G. Mitochondrial import and enzymatic activity of PINK1 mutants associated to recessive parkinsonism. *Human molecular genetics*, 2005, 14 (22), 3477-92.
 - [93] Beilina, A; Van Der Brug, M; Ahmad, R; Kesavapany, S; Miller, DW; Petsko, GA; Cookson, MR. Mutations in PTEN-induced putative kinase 1 associated with recessive parkinsonism have differential effects on protein stability. *Proceedings of the National Academy of Sciences of the United States of America*, 2005, 102 (16), 5703-8.
 - [94] D'Amora, M; Angelini, C; Marcoli, M; Cervetto, C; Kitada, T; Vallarino, M. Expression of PINK1 in the brain, eye and ear of mouse during embryonic development. *Journal of chemical neuroanatomy*, 2011, 41 (2), 73-85.
 - [95] Lin, W; Kang, UJ. Characterization of PINK1 processing, stability, and subcellular localization. *Journal of neurochemistry*, 2008, 106 (1), 464-74.
 - [96] Lockhart, PJ; Lincoln, S; Hulihan, M; Kachergus, J; Wilkes, K; Bisceglia, G; Mash, DC; Farrer, MJ. DJ-1 mutations are a rare cause of recessively inherited early onset parkinsonism mediated by loss of protein function. *Journal of medical genetics*, 2004, 41 (3), e22.
 - [97] Bonifati, V; Rizzu, P; Squitieri, F; Krieger, E; Vanacore, N; van Swieten, JC; Brice, A; van Duijn, CM; Oostra, B; Meo, G; Heutink, P. DJ-1(PARK7), a novel gene for autosomal recessive, early onset parkinsonism. *Neurological sciences : official journal of the Italian Neurological Society and of the Italian Society of Clinical Neurophysiology*, 2003, 24 (3), 159-60.
 - [98] van Duijn, CM; Dekker, MC; Bonifati, V; Galjaard, RJ; Houwing-Duistermaat, JJ; Snijders, PJ; Testers, L; Breedveld, GJ; Horstink, M; Sandkuijl, LA; van Swieten, JC; Oostra, BA; Heutink, P. Park7, a novel locus for autosomal recessive early-onset parkinsonism, on chromosome 1p36. *Am J Hum Genet*, 2001, 69 (3), 629-34.
 - [99] Corti, O; Lesage, S; Brice, A. What genetics tells us about the causes and mechanisms of Parkinson's disease. *Physiological reviews*, 2011, 91 (4), 1161-218.
 - [100] Tarantino, P; Civitelli, D; Annesi, F; De Marco, EV; Rocca, FE; Pugliese, P; Nicoletti, G; Carrideo, S; Provenzano, G; Annesi, G; Quattrone, A. Compound heterozygosity in DJ-1 gene non-coding portion related to parkinsonism. *Parkinsonism & related disorders*, 2009, 15 (4), 324-6.
 - [101] Hedrich, K; Djarmati, A; Schafer, N; Hering, R; Wellenbrock, C; Weiss, PH; Hilker, R; Vieregge, P; Ozelius, LJ; Heutink, P; Bonifati, V; Schwinger, E; Lang, AE; Noth, J; Bressman, SB; Pramstaller, PP; Riess, O; Klein, C, DJ-1 (PARK7) mutations are less frequent than Parkin (PARK2) mutations in early-onset Parkinson disease. *Neurology*, 2004, 62 (3), 389-94.
 - [102] Lev, N; Roncevic, D; Ickowicz, D; Melamed, E; Offen, D. Role of DJ-1 in Parkinson's disease. *Journal of molecular neuroscience : MN*, 2006, 29 (3), 215-25.

- [103] Ariga, H; Takahashi-Niki, K; Kato, I; Maita, H; Niki, T; Iguchi-Ariga, SM. Neuroprotective function of DJ-1 in Parkinson's disease. *Oxidative medicine and cellular longevity*, 2013, 2013, 683920.
- [104] Xu, J; Zhong, N; Wang, H; Elias, JE; Kim, CY; Woldman, I; Pifl, C; Gygi, SP; Geula, C; Yankner, BA. The Parkinson's disease-associated DJ-1 protein is a transcriptional co-activator that protects against neuronal apoptosis. *Human molecular genetics*, 2005, 14 (9), 1231-41.
- [105] Bandopadhyay, R; Kingsbury, AE; Cookson, MR; Reid, AR; Evans, IM; Hope, AD; Pittman, AM; Lashley, T; Canet-Aviles, R; Miller, DW; McLendon, C; Strand, C; Leonard, AJ; Abou-Sleiman, PM; Healy, DG; Ariga, H; Wood, NW; de Silva, R; Revesz, T; Hardy, JA; Lees, AJ. The expression of DJ-1 (PARK7) in normal human CNS and idiopathic Parkinson's disease. *Brain : a journal of neurology*, 2004, 127 (Pt 2), 420-30.
- [106] Besong Agbo, D; Klafki, H; Poschmann, G; Seyfarth, K; Genius, J; Janssen, C; Stuhler, K; Wurst, W; Meyer, HE; Klingenspor, M; Wiltfang, J. Development of a capillary isoelectric focusing immunoassay to measure DJ-1 isoforms in biological samples. *Analytical biochemistry*, 2013, 443 (2), 197-204.
- [107] Kumaran, R; Kingsbury, A; Coulter, I; Lashley, T; Williams, D; de Silva, R; Mann, D; Revesz, T; Lees, A; Bandopadhyay, R. DJ-1 (PARK7) is associated with 3R and 4R tau neuronal and glial inclusions in neurodegenerative disorders. *Neurobiol Dis*, 2007, 28 (1), 122-32.
- [108] Kumaran, R; Vandrovcova, J; Luk, C; Sharma, S; Renton, A; Wood, NW; Hardy, JA; Lees, AJ; Bandopadhyay, R. Differential DJ-1 gene expression in Parkinson's disease. *Neurobiol Dis*, 2009, 36 (2), 393-400.
- [109] Lin, X; Cook, TJ; Zabetian, CP; Leverenz, JB; Peskind, ER; Hu, SC; Cain, KC; Pan, C; Edgar, JS; Goodlett, DR; Racette, BA; Checkoway, H; Montine, TJ; Shi, M; Zhang, J. DJ-1 isoforms in whole blood as potential biomarkers of Parkinson disease. *Scientific reports*, 2012, 2, 954.
- [110] Zhong, N; Kim, CY; Rizzu, P; Geula, C; Porter, DR; Pothos, EN; Squitieri, F; Heutink, P; Xu, J. DJ-1 transcriptionally up-regulates the human tyrosine hydroxylase by inhibiting the sumoylation of pyrimidine tract-binding protein-associated splicing factor. *The Journal of biological chemistry*, 2006, 281 (30), 20940-8.
- [111] Vilarino-Guell, C; Soto, AI; Lincoln, SJ; Ben Yahmed, S; Kefi, M; Heckman, MG; Hulihan, MM; Chai, H; Diehl, NN; Amouri, R; Rajput, A; Mash, DC; Dickson, DW; Middleton, LT; Gibson, RA; Hentati, F; Farrer, MJ. ATP13A2 variability in Parkinson disease. *Human mutation*, 2009, 30 (3), 406-10.
- [112] Ugolino, J; Fang, S; Kubisch, C; Monteiro, MJ. Mutant Atp13a2 proteins involved in parkinsonism are degraded by ER-associated degradation and sensitize cells to ER-stress induced cell death. *Human molecular genetics*, 2011, 20 (18), 3565-77.
- [113] Murphy, KE; Cottle, L; Gysbers, AM; Cooper, AA; Halliday, GM. ATP13A2 (PARK9) protein levels are reduced in brain tissue of cases with Lewy bodies. *Acta neuropathologica communications*, 2013, 1 (1), 11.
- [114] Lu, CS; Lai, SC; Wu, RM; Weng, YH; Huang, CL; Chen, RS; Chang, HC; Wu-Chou, YH; Yeh, TH. PLA2G6 mutations in PARK14-linked young-onset parkinsonism and sporadic Parkinson's disease. *American journal of medical genetics. Part B, Neuropsychiatric genetics : the official publication of the International Society of Psychiatric Genetics*, 2012, 159B (2), 183-91.

- [115] Di Fonzo, A; Dekker, MC; Montagna, P; Baruzzi, A; Yonova, EH; Correia Guedes, L; Szczerbinska, A; Zhao, T; Dubbel-Hulsman, LO; Wouters, CH; de Graaff, E; Oyen, WJ; Simons, EJ; Breedveld, GJ; Oostra, BA; Horstink, MW; Bonifati, V. FBXO7 mutations cause autosomal recessive, early-onset parkinsonian-pyramidal syndrome. *Neurology*, 2009, 72 (3), 240-5.
- [116] Deng, H; Liang, H; Jankovic, J. F-box only protein 7 gene in parkinsonian-pyramidal disease. *JAMA neurology*, 2013, 70 (1), 20-4.
- [117] Gomez-Garre, P; Jesus, S; Carrillo, F; Caceres-Redondo, MT; Huertas-Fernandez, I; Bernal-Bernal, I; Bonilla-Toribio, M; Vargas-Gonzalez, L; Carballo, M; Mir, P. Systematic mutational analysis of FBXO7 in a Parkinson's disease population from southern Spain. *Neurobiology of aging*, 2014, 35 (3), 727 e5-7.
- [118] Ilyin, GP; Rialland, M; Pigeon, C; Guguen-Guillouzo, C. cDNA cloning and expression analysis of new members of the mammalian F-box protein family. *Genomics*, 2000, 67 (1), 40-7.
- [119] Zhao, T; De Graaff, E; Breedveld, GJ; Loda, A; Severijnen, LA; Wouters, CH; Verheijen, FW; Dekker, MC; Montagna, P; Willemsen, R; Oostra, BA; Bonifati, V. Loss of nuclear activity of the FBXO7 protein in patients with parkinsonian-pyramidal syndrome (PARK15). *PloS one*, 2011, 6 (2), e16983.
- [120] Myhre, R; Klungland, H; Farrer, MJ; Aasly, JO. Genetic association study of synphilin-1 in idiopathic Parkinson's disease. *BMC medical genetics*, 2008, 9, 19.
- [121] Eyal, A; Szargel, R; Avraham, E; Liani, E; Haskin, J; Rott, R; Engelender, S. Synphilin-1A: an aggregation-prone isoform of synphilin-1 that causes neuronal death and is present in aggregates from alpha-synucleinopathy patients. *Proceedings of the National Academy of Sciences of the United States of America*, 2006, 103 (15), 5917-22.
- [122] Eyal, A; Engelender, S. Synphilin isoforms and the search for a cellular model of lewy body formation in Parkinson's disease. *Cell cycle*, 2006, 5 (18), 2082-6.
- [123] Szargel, R; Rott, R; Engelender, S. Synphilin-1 isoforms in Parkinson's disease: regulation by phosphorylation and ubiquitylation. *Cellular and molecular life sciences : CMLS*, 2008, 65 (1), 80-8.
- [124] Lei, P; Ayton, S; Finkelstein, DI; Adlard, PA; Masters, CL; Bush, AI. Tau protein: relevance to Parkinson's disease. *The international journal of biochemistry & cell biology*, 2010, 42 (11), 1775-8.
- [125] Jakubauskiene, E; Janaviciute, V; Peciuliene, I; Soderkvist, P; Kanopka, A. G/A polymorphism in intronic sequence affects the processing of MAO-B gene in patients with Parkinson disease. *FEBS letters*, 2012, 586 (20), 3698-704.
- [126] Ho, SL; Kapadi, AL; Ramsden, DB; Williams, AC. An allelic association study of monoamine oxidase B in Parkinson's disease. *Annals of neurology*, 1995, 37 (3), 403-5.
- [127] Kurth, JH; Kurth, MC; Poduslo, SE; Schwankhaus, JD. Association of a monoamine oxidase B allele with Parkinson's disease. *Annals of neurology*, 1993, 33 (4), 368-72.
- [128] Sobell, JL; Lind, TJ; Hebrink, DD; Heston, LL; Sommer, SS. Screening the monoamine oxidase B gene in 100 male patients with schizophrenia: a cluster of polymorphisms in African-Americans but lack of functionally significant sequence changes. *American journal of medical genetics*, 1997, 74 (1), 44-9.
- [129] Shehadeh, LA; Yu, K; Wang, L; Guevara, A; Singer, C; Vance, J; Papapetropoulos, S. SRRM2, a potential blood biomarker revealing high alternative splicing in Parkinson's disease. *PloS one*, 2010, 5 (2), e9104.

- [130] Stamper, C; Siegel, A; Liang, WS; Pearson, JV; Stephan, DA; Shill, H; Connor, D; Caviness, JN; Sabbagh, M; Beach, TG; Adler, CH; Dunckley, T. Neuronal gene expression correlates of Parkinson's disease with dementia. *Movement disorders : official journal of the Movement Disorder Society*, 2008, 23 (11), 1588-95.
- [131] Soreq, L; Bergman, H; Israel, Z; Soreq, H. Exon arrays reveal alternative splicing aberrations in Parkinson's disease leukocytes. *Neuro-degenerative diseases*, 2012, 10 (1-4), 203-6.
- [132] Soreq, L; Bergman, H; Israel, Z; Soreq, H. Deep brain stimulation modulates nonsense-mediated RNA decay in Parkinson's patients leukocytes. *BMC genomics*, 2013, 14, 478.
- [133] Potashkin, JA; Santiago, JA; Ravina, BM; Watts, A; Leontovich, AA. Biosignatures for Parkinson's disease and atypical parkinsonian disorders patients. *PloS one*, 2012, 7 (8), e43595.
- [134] Santiago, JA; Scherzer, CR; Harvard Biomarker, S; Potashkin, JA. Specific splice variants are associated with Parkinson's disease. *Movement disorders : official journal of the Movement Disorder Society*, 2013, 28 (12), 1724-7.
- [135] Soreq, L; Guffanti, A; Salomonis, N; Simchovitz, A; Israel, Z; Bergman, H; Soreq, H. Long non-coding RNA and alternative splicing modulations in Parkinson's leukocytes identified by RNA sequencing. *PLoS computational biology*, 2014, 10 (3), e1003517.
- [136] Feng, Y; Jankovic, J; Wu, YC. Epigenetic mechanisms in Parkinson's disease. *Journal of the neurological sciences*, 2015, 349 (1-2), 3-9.
- [137] Soreq, L; Salomonis, N; Bronstein, M; Greenberg, DS; Israel, Z; Bergman, H; Soreq, H. Small RNA sequencing-microarray analyses in Parkinson leukocytes reveal deep brain stimulation-induced splicing changes that classify brain region transcriptomes. *Frontiers in molecular neuroscience*, 2013, 6, 10.

Chapter 16

TAU ALTERNATIVE SPLICING IN ALZHEIMER'S DISEASE

***Gonzalo Emiliano Aranda-Abreu* ,
María Elena Hernández Aguilar, Fausto Rojas Durán,
Sonia Lilia Mestizo Gutiérrez and Jorge Manzo Denes***

Universidad Veracruzana/Centro de Investigaciones Cerebrales,
Xalapa Veracruz, Mexico

ABSTRACT

The microtubule-associated protein (MAP) tau is essential for the development of neuronal cell polarity. Tau protein is preferentially localized in the axon, whereas MAP2, another neuron-specific microtubule-associated protein, is localized in the somatodendritic domain. Previous studies have demonstrated that the localization of these proteins depends, at least in part, on mRNA subcellular localization - of tau mRNA into the axon and MAP2 mRNA into the dendrite.

Tau protein plays a pivotal role in the pathophysiology of Alzheimer's disease, in which its hyperphosphorylation promotes aggregation and microtubule destabilization. Tau undergoes alternative splicing, which generates six isoforms in the human brain, due to the inclusion/exclusion of exons 2, 3 and 10. Dysregulation of the splicing process of tau exon 10 is sufficient to cause tauopathy, and has been shown to be influenced by β -amyloid peptides, while there has been less research conducted on the splicing of other exons.

This study found that the effects of β -amyloid (42) on the alternative splicing of tau exon 2/3 and 6 caused formed cell processes to retract in differentiated cells and altered the expression of exon 2/3 in cell culture. Expression of exon 6 was repressed under β -amyloid treatment. Although the molecular mechanism for this amyloid-tau interaction remains to be determined, it may have potential implications for the understanding of the underlying neuropathological processes in Alzheimer's disease.

Keywords: alternative splicing, RNA, Alzheimer's disease

* Corresponding Author's Email: garanda@uv.mx.

RNA SPLICING

Eukaryotic cells eliminate introns through RNA splicing. As sequences within RNA that do not code for protein, introns are removed by a structure rich in ribonucleoproteins, known as spliceosome complex. They then join the exonic sequences that produce mature transcript, which migrates from the cell nucleus to the cytoplasm and is then translated into a protein.

The splicing mechanism must be very accurate, since at least 50% of human genetic diseases are associated with mutations that occur in the consensus sequences of the splicing sites. These sequences consist of a 5'-GU and an AG sequence in the 3'-end intron. A region rich in pyrimidines is found toward the end of the 5' intron (CU).

The spliceosome complex must be assembled for splicing to occur. Consensus sequences located in the exon-intron frontier are essential for the 5 ribonucleoproteins (snRNP) U1, U2, U4, U5 and U6 to join these sequences and thus form the spliceosome. This spliceosome consists of several protein complexes. Complex E is created by the binding of U1 to the GU sequence at the 5' site of an intron, the binding of SF1 to the intron branch point, the binding of U2AF1 to the 3' splice site, and the binding of U2F2 to a sequence of polypyrimidines.

Complex A is created when U2 moves to SF1 and binds to the branch point sequence, while Complex B is created when U5, U4 and U6 form a trimer and, together with U5, bind with U2, releasing U1, wherein U5 binds to the intron and exon and U6 binds to the 5' splice site. Complex C is created when U4 is released and U6/U2 catalyze a transesterification process, binding the 5'-end of the intron to the complex to form an intron-lariat structure, with the U5 joining the exon at the 3' splice site and the site is off. U2, U5 and U6 then remain attached to the lariat structure, with the 3' lariat site then cut and the exons linked through ATP hydrolysis. The lariat forms are degraded and the snRNPs are recycled (figure 1).

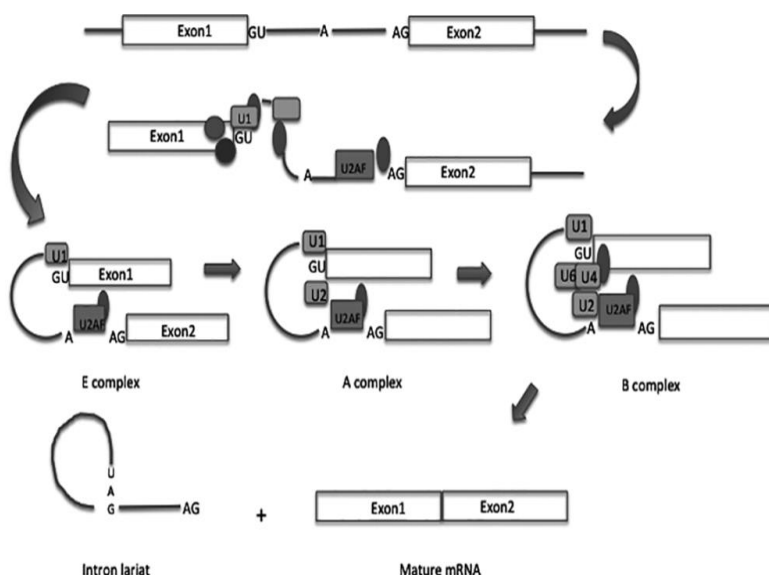


Figure 1. RNA splicing. Exon 1 flanked on its 3' end by the GU sequence and exon 2 on its 5' end by AG, with both target sites for the ribonucleoproteins and the assembled spliceosome complex. The spliceosome will cut the intron in the consensus sequences and will enable the joining of the exons, generating a mature RNA.

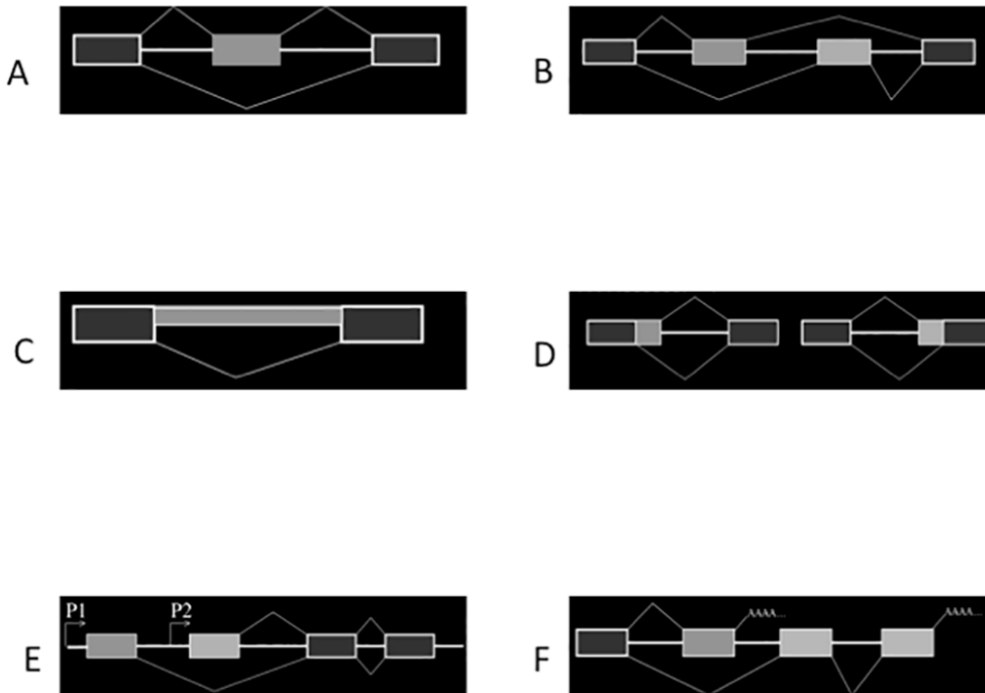


Figure 2. Forms of alternative splicing. A) exclusion or inclusion of exons, B) selection of one or more exons, C and D) competencies by the site of splicing in a particular exon in the region 5' or 3', E) retention of an intron, F) multiple promoters, G) multiple sites poly-A.

In general, through alternative splicing, genes are able to generate a variety of mRNAs (messenger RNA) with the significant biological function of encoding a protein. It has been estimated that at least 90% of expressed genes present alternative splicing.

There are at least 6 circumstances in which alternative splicing is generated: a) the exclusion or inclusion of exons; b) the selection of one or more exons, c and d) competencies by the splice site in a particular exon in the 5' or 3' region, e) the retention of an intron, f) multiple promoters, g) multiple poly-A [1] sites (figure 2). Both exonic and intronic sequences can modulate the splice site through either enhancers or silencer sequences.

ALZHEIMER'S DISEASE

Alzheimer's disease is one of the main genetic diseases associated with alternative splicing.

The most common neurodegenerative disorder occurring in the over 65s [2], Alzheimer's disease (AD) is slow and progressive in nature and is associated with memory loss and behavioral disorders [3].

Symptoms are imperceptible at the onset of the disease, usually for the first 2 or 3 years. There are a few cases of the transmission of chromosomal alterations through autosomal dominant inheritance.

This condition is the most common form of dementia, especially among older people, where it is estimated that the number of cases will reach 135 million people by 2050. When a patient is diagnosed with AD, the pathology will have already progressed for a number of years,

and the disease is, by this point, in its final stages. Although changes occur at least 20 or 30 years before symptoms first appear, it is difficult to detect the disease in advance, as there are no diagnostic tests to confirm these changes.

The pathological processes often linked with AD are premature ageing, amyloid protein deposits, neurofibrillar degeneration, synapse loss, inflammation, metabolic changes, loss of vascular integrity, injury and neuronal loss / neuronal injury and loss.

The degenerative process is neurofibrillary in type and generates nerve cell dysfunction, causing the death of specific neurons in the hippocampus, amygdala, entorhinal cortex and the nucleus basalis of Meynert [4].

Histologically, there are two types of proteinaceous aggregates, the Paired Helical Filaments (PHF) located inside the neuron, and the senile plaques, located in the extracellular space. The PHF include a compact filamentous network formed by aggregates of hyperphosphorylated tau [5], [6], while the senile plaques are formed by peptide β -amyloid deposits [6, 7].

Tau is a protein of the cytoskeleton involved in neuronal morphology and polarity and has the particularity of binding to the microtubules to provide stability and to maintain the neuronal phenotype at the level of the axon [8].

It has been determined that Tau is located in the axon hillock and axon growth cone, given that its mRNA is transported by a protein complex that uses kinesin-3 as a conveyor. It uses the HUD protein as an mRNA stabilizer for the site to which it will be translated [9-11], due to the fact that the mRNA of tau has uridine-rich sequences in its 3'-UTR region of a location in axonal [9,12].

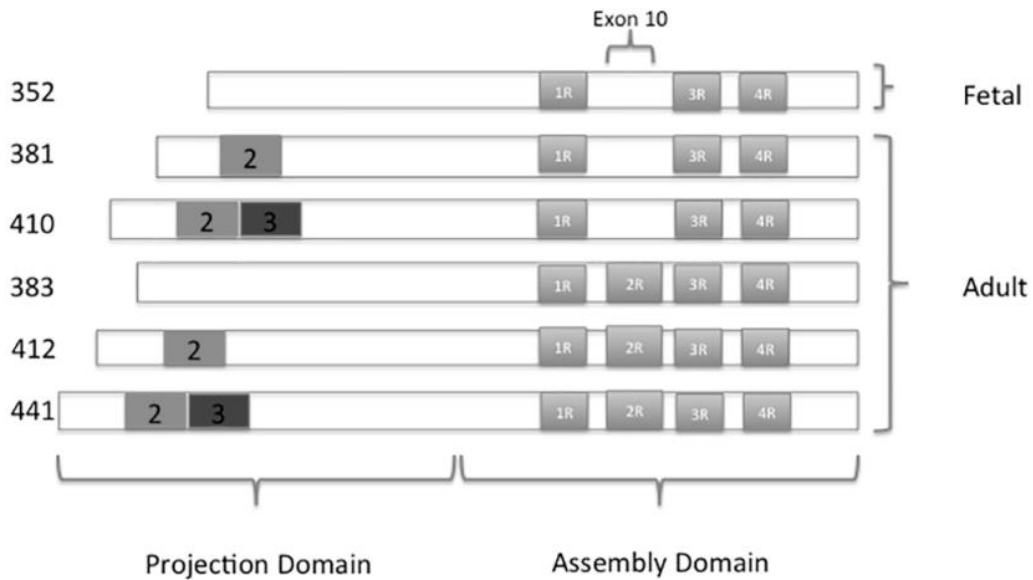
The Tau protein is mainly formed by two domains, the N-terminal, whose function is to interact with the plasma membrane [13], and a C-terminal microtubule-binding domain [14].

The human tau gene is located on chromosome 17 [15]. It is composed of 16 exons and its promoter region that provides neuronal specificity [16].

This gene transcribes three RNAs of 2, 6 and 9 kb that are differentially expressed in the central nervous system depending on their stage of maturity and neuronal type [13]. Six tau mRNA isoforms have been identified through alternative splicing. Five isoforms in the adult central nervous system and one fetal-type isoform. These RNA messengers encode 6 proteins ranging from 352 to 441 amino acids(aa). The fetal isoform (352 aa) does not contain exons 2, 3 and 10. The adult form of 383 aa lacks exons 2 and 3, but does, however, include exon 10. The 381 aa isoform includes exon 2 but not exon 10. The 412 aa isoform includes exon 2 and exon 10, while the 410 aa isoform includes exons 2 and 3, but not exon 10. The 441 aa isoform includes exons 2, 3 and 10 [17] (figure 3).

The alternative splicing of tau occurs in exons 2, 3 and 10, whose form is the type (a) which would correspond to the exclusion or inclusion of exons.

The alternative splicing of tau has been the subject of wide research, most of which has focused on exon 10, which presents a splicing pattern of inclusion and an absence of fetal neurons, and is influenced by exon 9, which promotes inclusion [13]. Exon 10 encodes the second repeat region of Tau (R) (KXGS). Alternative splicing generates isoforms of Tau, 3 or 4 of which are repeated and bind to microtubules. In the mature brain, the levels of repeated 3R and 4R are equal. The disruption of exon 10 is sufficient to cause neurodegeneration or tauopathies [18].



R: Repeat (KXGS)

Figure 3. Tau isoforms, showing the different Tau proteins from the alternative splicing of exons 2, 3 and 10.

Exon 10 is flanked by a long intron of 13.6 kb and a short one of 3.8 kb. Its weak 5' splice site, similar to the 3' site, enables the generation of proteins either with exon 10 or without it [19, 20].

Among the various tauopathies associated with exon 10 (table) that have been studied is Multiple System Tauopathy Disease (MSTD). A neurodegenerative disease with abundant protein tau filaments, this disease belongs to the group of frontotemporal dementia alongside Parkinsonism. MSTD has been reported to be caused by a G to A mutation in the 5' region of exon 10 [21]. This mutation in the splice site is likely to destabilize a stem-loop structure [22] which is involved in the regulation of exon 10 splicing, which in turn might cause an imbalance in the transcripts that include exon 10 and thus affect the amount of repeated microtubule bindings [23].

Another type of dementia related to exon 10 is Frontotemporal Dementia and Parkinsonism (FTDP), which is the most common cause of neurodegenerative disease after Alzheimer's disease. One study found exon 10 mutations in six French families resulting in a Pro→Leu change at amino acid 301. This substitution produces a drastic change in the conformation of the second repeats (Pro-Gly - Gly-Gly) located in the C-terminal of the MAPT protein highly conserved in species [24-26].

Since the late 1990s, research has been conducted on conditions such as: Pick's Disease, which does not include exon 10 or which might include mutations in exon 10; corticobasal degeneration; and, progressive supranuclear palsy, in which the region that encodes exon 10 is hyperphosphorylated (thus preventing Tau from binding to the microtubules) and in which the protein is assembled into filaments [27]. Studies have reported two mutations, Asn279Lys and

Pro301Leu, which are involved in Pallido-Ponto-Nigral (PPND) Disease, by causing neuronal loss in the subcortical nucleus and frontal cortex [28].

Studies conducted in murine models have shown that the deletion of exon 10 leads to an age-associated sensory-motor pathology [29].

While less research has been conducted on alternative exon 2 splicing, studies conducted in our laboratory show that when cell cultures of PC12 (rat pheochromocytoma) are exposed to the β -amyloid peptide 1 $\rightarrow$ 42, they affect the alternative splicing of exons 2 and 3, since immature forms of tau mRNA begin to be transcribed into mature PC12 cells (differentiated neuronal phenotype). Researchers from our laboratory have observed that the processes of these cells begin to retract. Although the mechanism is not yet known, the effect produced in the cells indicates that the immature forms of Tau protein enable the stabilization of microtubules in the processes of these cells [30].

The inclusion of exons 2 and 3 promotes the exchange of immature forms of Tau for mature forms that stabilize microtubules.

Disease	Exon	Mutation	Reference
Multiple system tauopathy.	10	3'GT splice donor	[21]
FTDP-17	10	G272V, P301L and R406W, 5' splice site of exon 10	[23]
Familial frontotemporal dementia and parkinsonism.	10	Pro-->Leu change at amino acid 301	[24]
Neurofibrillary degeneration processes.	10	hyperphosphorylated exon 10-tau.	[27]
Pallido-ponto-nigral degeneration	10	Asn279(Lys) in the PPND	[28]
FTDP-17 mutations.	10	N279K and S305N, increase splicing of exon 10	[25]
Sensorimotor affectation	10	Lack exon 10	[29]
Immature forms of tau mRNA.	2/3	Lack exons 2 and 3	[30]
DM1	2	Lack exon 2	[32]

It has been demonstrated that immature forms of Tau are similar to those found in the FHP [31], which suggests that the exclusion of exons 2 and 3 induced by the amyloid peptide in the AD could destabilize the neurites and thus cause cells to lose their polarity.

A pathology associated with the splicing of exon 2 of tau occurs in human subjects with type 1 Myotonic Dystrophy, which affects two splicing factors, MBLN1 and MBLN2 (muscleblind-like), by preventing the inclusion of exon 2 of tau [32].

Currently the regulation of splicing has been studied by focusing on the microRNAs (miRNAs), which are regulators of gene expression [33]. The miRNAs are short RNA molecules that bind to the transcript in order to suppress and regulate expression. This study found a deregulation of miRNA in the hippocampus and prefrontal cortex, determining that miR-132-3-p was the most altered as the disease progressed. The downregulation of Mir-132-

3-p in neurons was inversely proportional to the emergence of hyperphosphorylated tau [34] and is associated with the splicing of exon 10 [35].

The inclusion of exon 10 is inhibited by such factors as ASF/SF2, SRp55, SRP75 and SWAP [36].

The regulation of exon 2 has been determined by the inclusion and exclusion of exons 2 and 3, thus showing that exon 3 never appears independently of exon 2 [37].

CONCLUSION

Alternative splicing is an important process at a cellular level, since it enables the generation of various RNA messengers from the same gene, producing various proteins.

The regulation of alternative splicing should be performed under strict temporal and spatial conditions, depending on the type of cell and environment to which the cell is exposed.

Around 50% of the diseases have their origin in mutations in the consensus sequences of alternative RNA splicing.

Alzheimer's disease will claim the lives of 135 million people by the year 2050, meaning that this disease is fast becoming a public health problem.

While first world countries are already taking action against this disease, action is not undertaken to this level in third world countries, where this ailment is still not awarded the seriousness it is due. This may be due to the fact that the priorities of third-world populations lie in the development of other areas such as infrastructure, energy, and the economy. In these countries too, however, Alzheimer's disease will also become a serious problem in the future.

REFERENCES

- [1] Matlin, AJ; Clark, F; Smith, CW. Understanding alternative splicing: towards a cellular code. *Nat Rev Mol Cell Biol.*, 2005, 6, 386-398.
- [2] Aranda-Abreu, GE; Hernandez-Aguilar, ME; Manzo Denes, J; Garcia Hernandez, LI; Herrera Rivero, M. Rehabilitating a brain with Alzheimer's: a proposal. *Clin Interv Aging.*, 2011, 6, 53-59.
- [3] Herrera-Rivero, M; Hernandez-Aguilar, ME; Manzo, J; Aranda-Abreu, GE. [Alzheimer's disease: immunity and diagnosis]. *Rev Neurol.*, 2010, 51, 153-164.
- [4] West, MJ; Coleman, PD; Flood, DG; Troncoso, JC. Differences in the pattern of hippocampal neuronal loss in normal ageing and Alzheimer's disease. *Lancet.*, 1994, 344, 769-772.
- [5] Maccioni, RB; Cambiazo, V. Role of microtubule-associated proteins in the control of microtubule assembly. *Physiol Rev.*, 1995, 75, 835-864.
- [6] Selkoe, DJ. Translating cell biology into therapeutic advances in Alzheimer's disease. *Nature.*, 1999, 399, A23-A31.
- [7] Gandy, S; Petanceska, S. Regulation of Alzheimer beta-amyloid precursor trafficking and metabolism. *Biochim Biophys Acta.*, 2000, 1502, 44-52.

- [8] Littaer, UZ; Giveon, D; Thierauf, M; Ginzburg, I; Ponstingl, H. Common and distinct tubulin binding sites for microtubule-associated proteins. *Proc Natl Acad Sci U S A.*, 1986, 83, 7162-7166.
- [9] Aranda-Abreu, GE; Behar, L; Chung, S; Furneaux, H; Ginzburg, I. Embryonic lethal abnormal vision-like RNA-binding proteins regulate neurite outgrowth and tau expression in PC12 cells. *J Neurosci.*, 1999, 19, 6907-6917.
- [10] Aronov, S; Aranda, G; Behar, L; Ginzburg, I. Axonal tau mRNA localization coincides with tau protein in living neuronal cells and depends on axonal targeting signal. *J Neurosci.*, 2001, 21, 6577-6587.
- [11] Aronov, S; Aranda, G; Behar, L; Ginzburg, I. Visualization of translated tau protein in the axons of neuronal P19 cells and characterization of tau RNP granules. *J Cell Sci.*, 2002, 115, 3817-3827.
- [12] Aranda-Abreu, GE; Hernandez, ME; Soto, A; Manzo, J. Possible Cis-acting signal that could be involved in the localization of different mRNAs in neuronal axons. *Theor Biol Med Model.*, 2005, 2, 33.
- [13] Andreadis, A. Misregulation of Tau Alternative Splicing in Neurodegeneration and Dementia. In: Jeanteur P, editor. *Alternative Splicing and Disease*. France: Springer, 2006, 89-107.
- [14] Himmler, A; Drechsel, D; Kirschner, MW; Martin, DWJ. Tau consists of a set of proteins with repeated C-terminal microtubule-binding domains and variable N-terminal domains. *Mol Cell Biol.*, 1989, 9, 1381-1388.
- [15] Neve, RL; Harris, P; Kosik, KS; Kurnit, DM; Donlon, TA. Identification of cDNA clones for the human microtubule-associated protein tau and chromosomal localization of the genes for tau and microtubule-associated protein 2. *Brain Res.*, 1986, 387, 271-280.
- [16] Heicklen-Klein, A; Ginzburg, I. Tau promoter confers neuronal specificity and binds Sp1 and AP-2. *J Neurochem.*, 2000, 75, 1408-1418.
- [17] Buee, L; Bussiere, T; Buee-Scherrer, V; Delacourte, A; Hof, PR. Tau protein isoforms, phosphorylation and role in neurodegenerative disorders. *Brain Res Brain Res Rev.*, 2000, 33, 95-130.
- [18] Adams, SJ; DeTure, MA; McBride, M; Dickson, DW; Petrucelli, L. Three repeat isoforms of tau inhibit assembly of four repeat tau filaments. *PLoS One.*, 2010, 5, e10810.
- [19] D'Souza, I; Schellenberg, GD. Determinants of 4-repeat tau expression. Coordination between enhancing and inhibitory splicing sequences for exon 10 inclusion. *J Biol Chem.*, 2000, 275, 17700-17709.
- [20] Qian, W; Liu, F. Regulation of alternative splicing of tau exon 10. *Neurosci Bull.*, 2014, 30, 367-377.
- [21] Spillantini, MG; Murrell, JR; Goedert, M; Farlow, MR; Klug, A; Ghetti, B. Mutation in the tau gene in familial multiple system tauopathy with presenile dementia. *Proc Natl Acad Sci U S A.*, 1998, 95, 7737-7741.
- [22] Donahue, CP; Ni, J; Rozners, E; Glicksman, MA; Wolfe, MS. Identification of tau stem loop RNA stabilizers. *J Biomol Screen.*, 2007, 12, 789-799.
- [23] Hutton, M; Lendon, CL; Rizzu, P; et al. Association of missense and 5'-splice-site mutations in tau with the inherited dementia FTDP-17. *Nature.*, 1998, 393, 702-705.
- [24] Dumanchin, C; Camuzat, A; Campion, D; et al. Segregation of a missense mutation in the microtubule-associated protein tau gene with familial frontotemporal dementia and parkinsonism. *Hum Mol Genet.*, 1998, 7, 1825-1829.

- [25] Hasegawa, M; Smith, MJ; Iijima, M; Tabira, T; Goedert, M. FTDP-17 mutations N279K and S305N in tau produce increased splicing of exon 10. *FEBS Lett.*, 1999, 443, 93-96.
- [26] Morris, HR; Perez-Tur, J; Janssen, JC; et al. Mutation in the tau exon 10 splice site region in familial frontotemporal dementia. *Ann Neurol.*, 1999, 45, 270-271.
- [27] Mailliot, C; Sergeant, N; Bussiere, T; Caillet-Boudin, ML; Delacourte, A; Buee, L. Phosphorylation of specific sets of tau isoforms reflects different neurofibrillary degeneration processes. *FEBS Lett.*, 1998, 433, 201-204.
- [28] Clark, LN; Poorkaj, P; Wszolek, Z; et al. Pathogenic implications of mutations in the tau gene in pallido-ponto-nigral degeneration and related neurodegenerative disorders linked to chromosome 17. *Proc Natl Acad Sci U S A.*, 1998, 95, 13103-13107.
- [29] Gumucio, A; Lannfelt, L; Nilsson, LN. Lack of exon 10 in the murine tau gene results in mild sensorimotor defects with aging. *BMC Neurosci.*, 2013, 14, 148.
- [30] Lagunes, T; Herrera-Rivero, M; Hernandez-Aguilar, ME; Aranda-Abreu, GE. Abeta (1-42) induces abnormal alternative splicing of tau exons 2/3 in NGF-induced PC12 cells. *An Acad Bras Cienc.*, 2014, 86, 1927-1934.
- [31] Smith, CJ; Anderton, BH; Davis, DR; Gallo, JM. Tau isoform expression and phosphorylation state during differentiation of cultured neuronal cells. *FEBS Lett.*, 1995, 375, 243-248.
- [32] Carpentier, C; Ghanem, D; Fernandez-Gomez, FJ; et al. Tau exon 2 responsive elements deregulated in myotonic dystrophy type I are proximal to exon 2 and synergistically regulated by MBNL1 and MBNL2. *Biochim Biophys Acta.*, 2014, 1842, 654-664.
- [33] Roth, W; Hecker, D; Fava, E. Systems Biology Approaches to the Study of Biological Networks Underlying Alzheimer's Disease: Role of miRNAs. *Methods Mol Biol.*, 2016, 1303, 349-377.
- [34] Lau, P; Bossers, K; Janky, R; et al. Alteration of the microRNA network during the progression of Alzheimer's disease. *EMBO Mol Med.*, 2013, 5, 1613-1634.
- [35] Smith, PY; Delay, C; Girard, J; et al. MicroRNA-132 loss is associated with tau exon 10 inclusion in progressive supranuclear palsy. *Hum Mol Genet.*, 2011, 20, 4016-4024.
- [36] Wang, J; Gao, QS; Wang, Y; Lafyatis, R; Stamm, S; Andreadis, A. Tau exon 10, whose missplicing causes frontotemporal dementia, is regulated by an intricate interplay of cis elements and trans factors. *J Neurochem.*, 2004, 88, 1078-1090.
- [37] Li, K; Arikan, MC; Andreadis, A. Modulation of the membrane-binding domain of tau protein: splicing regulation of exon 2. *Brain Res Mol Brain Res.*, 2003, 116, 94-105.

Chapter 17

THE ROLE OF SPLICING FACTORS IN CANCER PROGNOSIS AND TREATMENT

Rebeca Martínez-Contreras and Nancy Martínez-Montiel

Molecular and Microbial Ecology Lab, Center of Research
on Microbiological Sciences, Science Institute,
Benemérita Universidad Autónoma de Puebla, Puebla, Mexico

ABSTRACT

Alternative splicing is a co-transcriptional mechanism that regulates eukaryotic gene expression that affects almost 90% of the human genes. In this mechanism, different combinations of exons and introns can be identified and removed from the pre-mRNA, allowing multiple mRNA configurations of joined sequences to arise from a single gene, increasing the coding potential of the genome. Alternative splicing events are catalyzed by a large complex known as the spliceosome, which is conformed by more than 300 proteins and ribonucleoproteins. At the catalytic core of the spliceosome, the small nuclear ribonucleoproteins (snRNPs) U1, U2, U4, U5 and U6 are found. The auxiliary factors responsible for the fine regulation of this mechanism include two major groups: the SR proteins and the hnRNP family.

Malfunctions of alternative splicing events can affect the natural expression of a large number of transcripts, including several factors involved in apoptosis or cell survival, molecular processes intimately associated with cancer evolution. In many cases, specific splicing factors or mutated components of the splicing machinery are linked to an anomalous event. Moreover, a switch in specific splicing factors occurs in particular types of cancer where the concomitant outcome is the production of non-functional proteins with added, deleted, or altered domains affecting tumorigenesis.

With all this evidence, several strategies have been developed to regulate alternative splicing in which central or auxiliary splicing factors are the target of modulatory molecules. Given the combination of elements needed to regulate alternative splicing, the mechanisms underlying the functional and physiological implications of these tools are also diverse. Collectively, these strategies are intended to improve cancer prognosis, therapeutic and treatment.

INTRODUCTION

mRNA processing is a key maturation mechanism that includes mRNA splicing, polyadenylation and capping. One of the key steps in the maturation process is alternative splicing, a nuclear co-transcriptional process that regulates eukaryotic gene expression. Initially, the extent of splicing was underestimated but it is currently established that almost 90% of mammalian genes undergo some splicing event, explaining in part the enormous coding potential of the human genome [1]. Malfunctions in this mechanism can generate different diseases, including rare disorders, neural conditions and cancer [2]. Unfortunately, it has been difficult to establish a precise correlation between one mutation and the correspondent alternative splicing event due to the transitive nature of gene expression. Moreover, a particular splicing event can be regulated differently according to the cell context, developmental stage or specific requirements [3].

General Splicing Mechanism

The general mechanism that regulates alternative splicing is presented in figure 1. During splicing, exonic or coding sequences are included while intronic or non-coding elements are excluded from the mature messenger RNA. This decision is determined by a large protein complex called the spliceosome, conformed by more than 300 proteins and ribonucleoproteins [4]. The catalytic center of the spliceosome is composed by the small nuclear ribonucleoproteins (snRNP) U1, U2, U4, U5 and U6.

Splicing regulation involves the recognition of *cis* elements coded in the pre-mRNA molecule. The main regulatory elements include the 5'ss and the 3'ss, which are located at the borders of the intron. The 5'ss is recognized by the snRNP U1. Close to the 3'ss, a pyrimidine-tract and a Branch-point sequence (BPS) can be also recognized by the snRNP U2 and some auxiliary factors. After the initial recognition, conformational changes and two consecutive trans-esterification reactions culminate with exon ligation catalyzed by snRNP U5 originating a mature product where introns have been removed. For human genes, conserved sequences have been detected for the 5'ss, the 3'ss, the polypyrimidine tract and the BPS. Depending on the location of these sequences along the pre-mRNA and how conserved they are, their relative strength can determine if these splicing signals are recognized during a precise splicing event. In many cases, a single-nucleotide mutation can alter one of these splice signals, activating maybe some other splicing signals and producing an aberrant mRNA or a non-functional protein.

In addition to the *cis* elements described above, other regulatory elements can improve or modulate the recognition of some splicing signals. These supplementary sequences can improve the recognition of a specific splice site and in this case we call them enhancers; they are called silencers when its activity block some splicing event. These sequences are either located at the exonic regions or coded inside the introns (Figure 1). Depending on their function and location, we can recognize Exonic Splicing Enhancers (ESE), Exonic Splicing Silencers (ESS), Intronic Splicing Enhancers (ISE) or Intronic Splicing Silencers (ISS). These regulatory elements are recognized by additional splicing factors, which will be further presented.

Mutations can also occur along these regulatory sequences with the same final outcome of aberrant mRNA or protein generation.

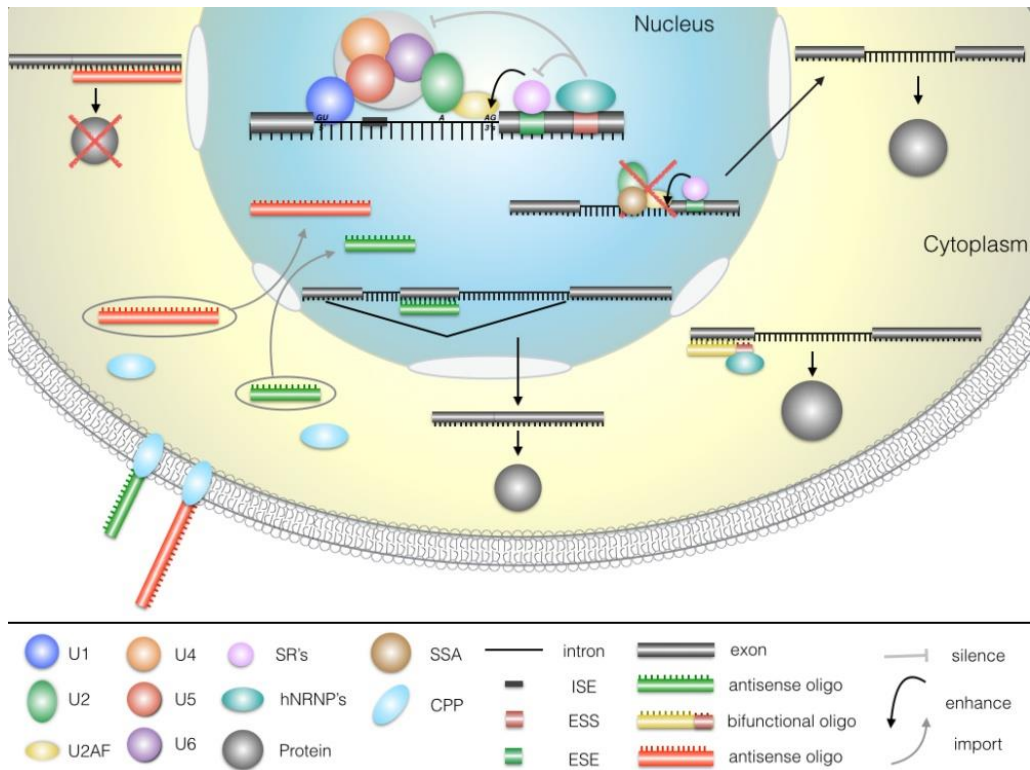


Figure 1. Alternative splicing mechanism and modulation. Alternative splicing occurs in the nucleus, where the core splicing factors (snRNPs) and regulatory proteins (SR and hnRNP families) bind consensus sequences in the pre-mRNA in order to produce mature mRNAs that will be later exported and translated in the cytoplasm. SR proteins will generally recognize exonic splicing enhancers (ESE) and hnRNPs could bind intronic splicing silencers (ISS). These events could be modulated by different molecules, like spliceostatin (SSA) that would change the final outcome of the splicing event. SSA interacts with SF3b complex of snRNP U2. Different types of oligonucleotides can also modulate splicing. These oligos can use carriers like CPP in order to enter the cell.

Splicing Regulatory Factors

The auxiliary factors responsible for the fine regulation of this mechanism include two major groups: the SR proteins [5] and the hnRNP family [6].

The SR proteins belong to a family of highly conserved splicing factors that has been described in vertebrates, fungi and plants. These factors possess a protein-interacting domain called RS due to its high content of arginine and serine residues and an RNA-interacting domain [5]. Initially, SR proteins were identified as splicing activators given their ability to recognize ESEs and recruiting the spliceosome to the adjacent intron. The activity of SR proteins in a specific cell context depends on several conditions, including its particular expression and concentration in that precise cell type and its phosphorylation status, which alters protein

conformation, RNA-binding or protein-binding capabilities, resulting in changes on spliceosome assembly and catalysis.

On the other hand, more than 20 heterogeneous nuclear ribonucleoproteins or hnRNPs have been identified and they comprise a family of RNA-binding proteins that possess diverse biological functions including splicing regulation and have been named alphabetically depending on their apparent mass from hnRNP A through hnRNP U [6]. In general, hnRNP proteins have shown predominantly antagonistic functions from SR proteins. In this regard, the ability of hnRNPs to prevent the recognition of the pre-mRNA by the snRNPs or SR proteins has been documented several cases. Since the initial evidence from *in vitro* studies where hnRNP A1 and SRSF1 showed the opposite effect over splice site selection [7] several investigations supported this observation [8, 9]. This antagonistic behavior has been observed also in cancer tissues [10]. However, further evidence supports the notion that hnRNP A1 and hnRNP F/H can facilitate the splicing of a long intron [11] arguing the common assumption that hnRNP proteins are splicing inhibitors.

Despite their specific function concerning alternative splicing decisions, splicing factors share several features that are relevant for their function, including their ability to bind RNA, to shuttle between the nucleus and the cytoplasm, to produce different protein variants due to alternative splicing and to interact with other proteins. Moreover, splicing factors have been involved in several co- and post-transcriptional events responsible for gene expression, including nuclear export, non-sense-mediated decay and translation [12]. This global relationship between splicing and further events can explain the relevant role of splicing decisions in cancer events and in cellular fate [13].

Summarizing, splicing regulation involves the cis-acting elements that correspond to sequences coded in the pre-mRNA, which can be recognized by trans acting factors where the core proteins are the snRNPs. The relative position of the cis elements can alter splicing selection. In addition to this combination, auxiliary factors that belong to the SR and hnRNP families can contribute to splicing decisions. Finally, the availability of central and auxiliary splicing factors, according to their expression, concentration and phosphorylation status may alter the splicing choice. The complex nature of splicing regulation correlates well with the precise nature of gene expression control.

Alternative Splicing and Cancer

It was initially believed that splicing was a rather exceptional mechanism. However, a growing amount of evidence supported the notion that alternative splicing governs the expression of the majority of human genes [14]. After these observations, the important contribution of alternative splicing in human disease was acknowledged and more recently its determinant role in cancer was widely recognized [15, 16]. For example, it has been shown that hereditary mutations linked to cancer predisposition in the BRCA1/2 genes are related to alternative splicing regulation [17, 18].

In general, alternative splicing events that have been associated to cancer are active during tumor progression and reinforce metastases, which is the cause of 90% of all human cancer casualties [19]. Normally, the expression of a specific splicing isoform that is linked to tumor progression can be detected in normal tissues as well, but once that cell homeostasis is lost, alternative splicing can provide an advantageous background for tumor progression. In this

regard, it is well documented that the splicing of various genes is altered during oncogenic progression and it can be related to the development of cancer features, like an increase in proliferation, vascularization and invasion [19, 20].

In several cases, changes in splicing profiles in tumors are related to the altered activity, expression level or even mutations in regulatory splicing factors. Like many regulatory factors in the cell, splicing regulators undergo a wide range of post-translational modifications. These modifications have been studied mainly for SR proteins and include phosphorylation [21,22], methylation [23, 24] and sumoylation [25]. These post-translational modifications have an impact not only on splicing regulation, but also on different aspects of cellular physiology. Phosphorylation of splicing factors SRSF1 and SRSF5 affect splicing of fibronectin 1 (FN1) [26], caspase 9 [27] and PKC β II [28]; these splicing decisions can affect tumor progression. These phosphorylation effects are not general or redundant and they vary according to the cancer type and the splicing factor involved [29].

Mutations in the Core Splicing Machinery Related to Cancer

The real impact of somatic mutations in genes coding components of the splicing machinery has not been thoroughly analyzed. However, several examples have been deposited in the International Cancer Genome Projects Database (CGPD). As it can be observed in table 1, the record deposited in this database indicates that there are 102,603 mutations in genes related to splicing, affecting a total of 5,949 donors. Unfortunately, the exact correlation between a precise mutation and its functional implication remains to be elucidated. According to the results obtained from 51 projects deposited in the CGPD, 361 genes coding for splicing factors are mutated in all types of cancer. The most frequently mutated genes include hnRNP proteins (NOVA1, hnRNP M, hnRNP C, hnRNP A2/B1, hnRNP F, RALY), SR proteins (SRSF4, RBM39, Tra2 α , Tra2 β) and SR-protein kinases (SRPK1 and SRPK2), RBM proteins (RBM4, RBM5) and auxiliary splicing factors.

Interestingly the only snRNP components affected seem to be some U2-specific factors (SF3A1, SF3A2, SF3A3, SF3B1, SF3B2, SF3B3 and SF3B4) and the U1-specific protein U170K (SNRNP70). Amongst these genes, the most frequently occurring mutations lie at the SF3B1, U2AF1, RALY hnRNP and SNRNP70 genes (Table 2). On the other hand, splicing mutations have been detected in patients from all types of cancer, but individuals with the highest number of mutations in the greatest number of genes are related to liver, pancreas, skin and breast cancer (Table 3). All these data strongly support the relevant function of alternative splicing in cancer, but the precise mechanism underlying each gene and factor regulation remains to be elucidated.

Concerning particular cases where precise alternative splicing events have been further studied, there are some well-documented cases. For example, it has been reported that recurrent mutations are found in genes that code for splicing factors SF3B1 and U2AF1 [30] in myelodysplastic syndromes (MDS). This group includes myeloid neoplasms with altered myeloid blood cell production and disposition to progress into acute myeloid leukemia. SF3B1 is also frequently mutated in chronic lymphocytic leukemia (CLL) [31, 32, 33], as well as in uveal melanoma [34, 35]. All SF3B1 mutations are heterozygous, and none are nonsense mutations or introduce a frameshift.

Table 1. Distribution of cancer types related to splicing

Project	Site	Tumour Type	Tumour Subtype	# Donors affected	# Mutations	# Genes
PRAD-CA	Prostate	Prostate cancer	Adenocarcinoma	124 / 124 (100.00%)	1,692	290 / 361 (80.33%)
RECA-EU	Kidney	Renal cancer	Renal cell carcinoma (Focus on but not limited to clear cell subtype)	95 / 95 (100.00%)	2,936	335 / 361 (92.80%)
MALY-DE	Blood	Malignant Lymphoma	Germinal center B-cell derived lymphomas	44 / 44 (100.00%)	1,483	288 / 361 (79.78%)
OV-AU	Ovary	Ovarian cancer	Serous cystadenocarcinoma	93 / 93 (100.00%)	5,173	353 / 361 (97.78%)
SKCA-BR	Skin	Melanoma		66 / 66 (100.00%)	12,678	356 / 361 (98.61%)
LUSC-KR	Lung	Lung cancer	Adenocarcinoma, Squamous cell carcinoma	36 / 36 (100.00%)	734	239 / 361 (66.20%)
PRAD-UK	Prostate	Prostate cancer	Adenocarcinoma	31 / 31 (100.00%)	1,358	277 / 361 (76.73%)
THCA-SA	Head and neck	Thyroid cancer	Papillary thyroid carcinoma	15 / 15 (100.00%)	1,035	275 / 361 (76.18%)
EOPC-DE	Prostate	Prostate cancer	Early Onset	11 / 11 (100.00%)	170	78 / 361 (21.61%)
LIRI-JP	Liver	Liver cancer	Hepatocellular carcinoma (Virus associated)	259 / 260 (99.62%)	15,985	351 / 361 (97.23%)
ESAD-UK	Esophagus	Esophageal cancer	Esophageal adenocarcinoma	118 / 119 (99.16%)	13,478	336 / 361 (93.07%)
CMDI-UK	Blood	Chronic Myeloid Disorders	Myelodysplastic Syndromes, Myeloproliferative Neoplasms & Other Chronic Myeloid Malignancies	127 / 129 (98.45%)	29	2 / 361 (0.55%)
PACA-IT	Pancreas	Rare Pancreatic Tumors	Enteropancreatic endocrine tumors and rare pancreatic exocrine tumors	36 / 37 (97.30%)	607	216 / 361 (59.83%)
LINC-JP	Liver	Liver cancer	Hepatocellular carcinoma (Virus associated)	236 / 244 (96.72%)	3,935	346 / 361 (95.84%)
LUSC-US	Lung	Lung cancer	Squamous cell carcinoma	171 / 178 (96.07%)	1,048	297 / 361 (82.27%)
PACA-CA	Pancreas	Pancreatic cancer	Ductal adenocarcinoma	193 / 204 (94.61%)	8,374	349 / 361 (96.68%)
BLCA-US	Bladder	Bladder cancer	Invasive Urothelial Bladder cancer	122 / 130 (93.85%)	869	273 / 361 (75.62%)
SKCM-US	Skin	Skin cancer	Cutaneous melanoma	306 / 335 (91.34%)	2,977	340 / 361 (94.18%)
READ-US	Colorectal	Rectal cancer	Adenocarcinoma	73 / 80 (91.25%)	439	223 / 361 (61.77%)
PAEN-AU	Pancreas	Pancreatic cancer	Endocrine neoplasms	47 / 52 (90.38%)	864	243 / 361 (67.31%)
UCEC-US	Uterus	Endometrial cancer	Uterine corpus endometrial carcinoma	218 / 246 (88.62%)	3,915	346 / 361 (95.84%)
STAD-US	Stomach	Gastric cancer	Adenocarcinoma	255 / 289 (88.24%)	3,015	333 / 361 (92.24%)
LIHC-US	Liver	Liver cancer	Hepatocellular carcinoma	160 / 188 (85.11%)	528	245 / 361 (67.87%)
CESC-US	Cervix	Cervical cancer	Cervical squamous cell carcinoma	165 / 194 (85.05%)	1,048	305 / 361 (84.49%)
COAD-US	Colorectal	Colon cancer	Adenocarcinoma	181 / 216 (83.80%)	2,006	326 / 361 (90.30%)

Project	Site	Tumour Type	Tumour Subtype	# Donors affected	# Mutations	# Genes
BLCA-CN	Bladder	Bladder cancer	Urothelial carcinoma	86 / 103 (83.50%)	381	199 / 361 (55.12%)
KIRP-US	Kidney	Renal cancer	Papillary carcinoma	131 / 159 (82.39%)	340	191 / 361 (52.91%)
ORCA-IN	Head and neck	Oral cancer	Gingivobuccal	87 / 106 (82.08%)	283	158 / 361 (43.77%)
BOCA-FR	Bone	Bone cancer	Ewing sarcoma	79 / 98 (80.61%)	234	107 / 361 (29.64%)
GACA-CN	Stomach	Gastric cancer	Intestinal- and diffuse-type	7 / 9 (77.78%)	22	20 / 361 (5.54%)
KIRC-US	Kidney	Renal cancer	Clear cell carcinoma	313 / 404 (77.48%)	630	248 / 361 (68.70%)
LICA-FR	Liver	Liver cancer	Hepatocellular carcinoma (Secondary to alcohol and adiposity)	178 / 234 (76.07%)	2,836	339 / 361 (93.91%)
LIHM-FR	Liver	Liver cancer	Hepatocellular carcinoma (Secondary to alcohol and adiposity)	3 / 4 (75.00%)	4	4 / 361 (1.11%)
PACA-AU	Pancreas	Pancreatic cancer	Ductal adenocarcinoma	278 / 392 (70.92%)	7,488	352 / 361 (97.51%)
COCA-CN	Colorectal	Colorectal cancer	Adenocarcinoma, non-Western	35 / 50 (70.00%)	124	98 / 361 (27.15%)
OV-US	Ovary	Ovarian cancer	Serous cystadenocarcinoma	61 / 88 (69.32%)	120	95 / 361 (26.32%)
GBM-US	Brain	Brain cancer	Glioblastoma multiforme	184 / 268 (68.66%)	339	169 / 361 (46.81%)
BRCA-UK	Breast	Breast cancer	Triple Negative/lobular/other	78 / 117 (66.67%)	1,391	293 / 361 (81.16%)
ESCA-CN	Esophagus	Esophageal cancer	Squamous carcinoma	58 / 88 (65.91%)	125	107 / 361 (29.64%)
PRAD-US	Prostate	Prostate cancer	Adenocarcinoma	168 / 256 (65.63%)	305	181 / 361 (50.14%)
BRCA-US	Breast	Breast cancer	Ductal & lobular	579 / 955 (60.63%)	1,883	339 / 361 (93.91%)
LGG-US	Brain	Brain cancer	Lower grade glioma	161 / 283 (56.89%)	268	161 / 361 (44.60%)
RECA-CN	Kidney	Renal cancer	Clear cell renal cell carcinoma	4 / 10 (40.00%)	4	4 / 361 (1.11%)
CLLE-ES	Blood	Chronic Lymphocytic Leukemia	CLL with mutated and unmutated IgVH	41 / 109 (37.61%)	63	49 / 361 (13.57%)
LIAD-FR	Liver	Liver cancer	Hepatocellular adenoma	11 / 30 (36.67%)	13	13 / 361 (3.60%)
ALL-US	Blood	Blood cancer	Acute lymphoblastic leukemia	9 / 27 (33.33%)	23	22 / 361 (6.09%)
PBCA-DE	Brain	Pediatric Brain Tumors	Medulloblastoma & Pediatric Pilocytic Astrocytoma	81 / 248 (32.66%)	X	235 / 361 (65.10%)
BOCA-UK	Bone	Bone cancer	Osteosarcoma / chondrosarcoma / rare subtypes	21 / 66 (31.82%)	30	29 / 361 (8.03%)
THCA-US	Head and neck	Head and Neck cancer	Thyroid carcinoma	109 / 400 (27.25%)	138	100 / 361 (27.70%)
LUSC-CN	Lung	Lung cancer	Squamous cell carcinoma	1 / 10 (10.00%)	2	2 / 361 (0.55%)
NBL-US	Nervous System	Pediatric solid tumor	Neuroblastoma	4 / 108 (3.70%)	4	4 / 361 (1.11%)

Table 2. Splicing genes most frequently mutated in cancer patients

Symbol	Name	Location	Type	# Donors affected	# Mutations
RBFOX1	RNA binding protein, fox-1 homolog (C. elegans) 1	chr16:6069095-7763340	protein_coding	1,478 / 5,949 (24.84%)	18,078
AFF2	AF4/FMR2 family, member 2	chrX:147582139-148082193	protein_coding	965 / 5,949 (16.22%)	2,388
RBFOX3	RNA binding protein, fox-1 homolog (C. elegans) 3	chr17:77085427-77613550	protein_coding	844 / 5,949 (14.19%)	2,176
RSRC1	arginine/serine-rich coiled-coil 1	chr3:157823644-158263519	protein_coding	796 / 5,949 (13.38%)	1,695
CELF4	CUGBP, Elav-like family member 4	chr18:34823010-35146000	protein_coding	785 / 5,949 (13.20%)	1,773
SNRPN	small nuclear ribonucleoprotein polypeptide N	chr15:25068794-25223870	protein_coding	743 / 5,949 (12.49%)	1,767
SRRM4	serine/arginine repetitive matrix 4	chr12:119419300-119600856	protein_coding	713 / 5,949 (11.99%)	1,535
NOVA1	neuro-oncological ventral antigen 1	chr14:26912299-27066960	protein_coding	688 / 5,949 (11.56%)	1,628
SRPK2	SRSF protein kinase 2	chr7:104751151-105039755	protein_coding	659 / 5,949 (11.08%)	1,033
AHNAK	AHNAK nucleoprotein	chr11:62201016-62323707	protein_coding	623 / 5,949 (10.47%)	861

SF3B and U2AF1 participate in splice site selection and mutations in these genes could affect 3'ss recognition. It has also been observed that in some cases, these mutations can correlate with intron retention, leading to speculate that this defect on splicing could be related to nonsense-mediated mRNA decay (NMD). However, the final outcome may depend on the gene in which alternative splicing is regulated and on the specific cellular context. For example, it has been reported that in HeLa cells, while the expression or mutant SF3B induces intron retention, mutant U2AF1 activates cell death [30]. The relevance of cell context for splicing decisions is also supported by additional evidence, given the observation that patients with MDS carrying SF3B1 mutations have usually a favorable prognosis, whereas SF3B1 mutations in CLL correlate with poor survival and resistance to chemotherapy [32, 33, 36, 37, 38]. Likewise, mutations on the spliceosomal core have also shown a different effect in pediatric MDS and juvenile myelomonocytic leukemia [39]. This cellular-context dependency of effects provides an opportunity for developing antitumor drugs that could elicit a response only in tumor but not in normal cells. Moreover, this specificity could also partially explain why some chemotherapeutic agents are more effective in specific cancer types.

Cancer-Related Changes in Splicing Factors Expression

As it was mentioned earlier, SR proteins are splicing regulators that usually bind ESEs and facilitate splicing. The first SR protein discovered and one of the best-studied splicing factors is SRSF1. It has been observed that this factor is overexpressed in several human cancer cell

lines. In addition, several genes related to cell cycle progression and oncogenic transformation, are known targets for SRSF1. Further details concerning the particular mechanism by which SRSF1 regulates cancer progression will be presented later. SRSF3 is another overexpressed SR protein for several human cancers and knocking down this gene can lead to apoptosis in different cancer cell lines [40, 41].

Phosphorylation of SR proteins is an important determinant of their localization and activity [42] and this regulation of SR protein activity by phosphorylation has now been implicated in a cancer-associated AS event.

Table 3. Most Frequent Mutations in splicing genes reported for individuals with cancer

ID	DNA change	Type	Consequences	# Donors affected
MU9178	chr2:g.198266834 T>C	single base substitution	Missense: SF3B1 K700E	94 / 8,038 (1.17%)
			Upstream: SF3B1 - SNORA4	
			Exon: SF3B1	
MU639899	chr1:g.92446133G >T	single base substitution	Downstream: BRDT	15 / 8,038 (0.19%)
			Intron: BRDT	
MU820754	chr21:g.44524456 G>A	single base substitution	Missense: U2AF1 S34F	14 / 8,038 (0.17%)
			5 UTR: U2AF1	
			Upstream: U2AF1	
			Exon: U2AF1	
MU117612	chr8:g.145617535 GGGGGTGCAA GGTGA>-	deletion of <=200bp	Frameshift: ADCK5 LGVQD419	14 / 8,038 (0.17%)
			Splice Donor: ADCK5	
			Downstream: CPSF1 - ADCK5 - MIR939	
MU2185470	chr20:g.32664864 ->CAG	insertion of <=200bp	Disruptive Inframe Insertion: RALY A230AA, A214AA, A164AA	
			Upstream: RALY - RP1-64K7.4	13 / 8,038 (0.16%)
			Exon: RALY	
			Downstream: RALY	
MU93311	chr19:g.49611319 C>T	single base substitution	Synonymous: SNRNP70 G302G, G311G	12 / 8,038 (0.15%)
			3 UTR: SNRNP70	
			Exon: SNRNP70	
MU4174958	chr18:g.34960790 T>C	single base substitution	Intron: CELF4	12 / 8,038 (0.15%)
MU4547272	chr4:g.190878654 C>G	single base substitution	Missense: FRG1 I115M, I178M, I50M	12 / 8,038 (0.15%)
			Upstream: FRG1	
			Exon: FRG1	
			Downstream: FRG1	
MU1792328	chr5:g.67588951C >T	single base substitution	Stop Gained: PIK3R1 R348*, R48*, R78*, R21*	11 / 8,038 (0.14%)
			Start Gained: PIK3R1	
			3 UTR: PIK3R1	
			Exon: PIK3R1	
MU18487	chr17:g.7386217T >C	single base substitution	Missense: SLC35G6 V305A	10 / 8,038 (0.12%)
			Upstream: ZBTB4 - POLR2A	
			Intron: ZBTB4	

Table 4. Individuals with the highest number of mutations in splicing genes

ID	Site	Gender	Age	Stage	Survival (days)	# Mutations	# Genes
DO45299	Liver	female	70	1	1,440	1,308	292
DO35442	Pancreas	female	54	IIA	1,331	783	272
DO33091	Pancreas	male	56		142	905	264
DO35083	Pancreas	female	69	IIB	1,264	776	257
DO219117	Esophagus	female	80	T3NxM0	175	963	257
DO48052	Head and neck	male	54	III		599	240
DO219878	Skin	male	52	X	1,229	1,140	224
DO219881	Skin	male	85	3		477	214
DO218874	Skin	male	54	4		764	203
DO1076	Breast	female				521	192

hnRNP proteins were initially described as negative regulators of splicing, and the original findings suggested that these proteins are splicing inhibitors. hnRNP A1 and hnRNP A2 have been long related to cancer regulation, mainly according to its ability to recognize and protect telomeric sequences [43]. These factors are also overexpressed in a wide variety of cancers and the silencing of these genes can also induce apoptosis. This effect seems to be cancer-related, given that it has not been observed in normal cells [44, 45, 46]. In terms of splicing, it has been found that more than 2000 alternative splicing events could be regulated either by hnRNP A1 or A2 proteins [47] and some of the shared targets are related to metabolic abnormalities relevant for tumor cell promotion [48].

hnRNP I (also known as PTB) was initially related to the exclusion of an alternative exon through the recognition of an ISS element [49]. A neural variant of this splicing factor was identified [50] and a global analysis showed that this factor, currently known as NOVA, regulates several neural splicing events, where the inclusion or exclusion of the alternative exon would depend on the position of the binding site for this factor [51]. This position-dependent regulation has been also demonstrated for other hnRNP proteins [11]. In agreement with these observations, genome-wide studies in HeLa cells revealed that PTB can either repress or activate exon inclusion, depending on the location of its binding site [49, 52], suggesting that the differential regulation can also occur in cancer cell lines and that it may have a role in cancer progression. According to its relevant function in the brain, high levels of hnRNP I have been found in gliomas [44, 53] and it might facilitate progression of astrocytic tumors [54]. hnRNP H is also up-regulated in gliomas and the silencing of this gene produces cell death in glioma (U373) and HeLa cells [55].

RBM family of proteins are also involved in splicing regulation. Currently, fifty proteins are classified as RBM (RNA-binding proteins) factors in the HUGO Gene Nomenclature Committee Database. The most widely studied of these RBM proteins, including RBM4, RBM5, RBM9 and RBM17, are involved in RNA processing, specifically in pre-mRNA splicing regulation [56-58]. RBM factors are components of the spliceosome that modulate alternative splicing of a number of genes by modulating splice site pairing after recruitment of the U1 and U2 snRNPs to the 5' and 3' splice sites of the intron. Particularly, RBM5, RBM6 and RBM10 are highly similar RNA-binding proteins that in some cases can show redundant functions [56]. RBM5 has shown the ability to positively and negatively modulate apoptosis by regulating the alternative splicing of several genes involved in this process, including FAS and CASP2/caspase-2. In the case of FAS, RBM5 promotes the exclusion of exon 6, which in turn produces a soluble form of FAS that inhibits apoptosis. In the case of CASP2/caspase-2,

RBM5 promotes the exclusion of exon 9 and generates a catalytically active form of CASP2/Caspase-2 that promotes apoptosis [59]. RBM10 overexpression correlated with elevated levels of the soluble isoform of TNF- α (sTNF- α) and with an increase in apoptosis [58], supporting its function as an apoptosis modulator and cytokine expression regulator.

The functionality of RBM proteins varies in different cancer types and a better understanding of the role played by these paralogs in the regulation of the expression levels for different genes may contribute to the production of more effective therapies for a wide collection of diseases, including cancer.

Splicing Regulators That Have Been Considered Oncogenes or Tumor Suppressors

Over the last few years, several splicing factors attracted special attention due to its oncogenic properties, like SRSF1 [60], SRSF9 [61], hnRNP A1/A2 [62] and hnRNP H [63]. The strongest evidence supporting the oncogenic effect of a splicing factor has been provided for SRSF1, given that its over-expression induces sarcomas in nude mice [59]. The oncogenic activity of this particular splicing factor could be related to the following molecular events related to the multi-factorial regulation of SRSF1 expression. It has been shown that the well-studied oncogene Myc activates the transcription of SRSF1 through two E-boxes and this activation correlates well with an increase of several splicing variants that are well-known targets of SRSF1 and that are consistent with an oncogenic phenotype [64]. On the other hand, SRSF1 controls the alternative splicing of the tumor suppressor BIN1, which in turn blocks the activity of Myc [65]. In this same sense, knocking-down SRSF1 reduces Myc's oncogenic activity [64] while its up-regulation has been observed in different types of tumors, including lung and breast. Overexpression of the two proteins SRSF1 and Myc results in breast epithelial cell-transformation and the accompanying effect is an increase in the activity of the translation factor eIF4E [66].

The tumor promoting activity of SRSF1 has been linked to its first RNA-recognition motif (RRM1), which controls the alternative splicing of several cancer-related genes, like BIN1, MNK2, S6K1, the apoptotic gene BIM [60] and it increases the expression of B-Raf, which in turn activates the MEK/ERK signaling pathway. In accordance with this, MEK1 inhibition blocks the transformation mediated by SRSF1 [67]. Furthermore, SRPK1 is a protein kinase responsible for SRSF1 translocation to the nucleus, whose transcription can be repressed by the tumor suppressor WT1 [68], supporting the link between SRSF1 location and its activity in cancer.

SRSF1 expression can be also regulated through alternative splicing and by micro-RNAs. At least six variants of SRSF1 can be generated through alternative splicing out of which one generates the full-length factor and the other five generate aberrant products. These aberrant products could be targets for NMD due to the generation of premature termination codons. Moreover, SRSF1 has the ability to auto-regulate the splicing of its own pre-mRNA and it can also down-regulate its own expression [69], indicating the high and multifactorial control of its expression. A direct link between SRSF1 expression and cancer derives from the observation that the Leukemia/lymphoma-related factor (LRF) is an oncogenic transcription factor that represses miR-28 and miR-505 [70] and these two miRNAs bind the 3'UTR of SRSF1 mRNA. In accordance to this, a reduction in LRF generates a decrease in SRSF1 expression [71].

Considering all these, the 3'UTR of SRSF1 is a good target for fine regulation either by miRNAs or by NMD and it is not rare that its regulation can have an important effect on further cellular processes, like those implicated in cancer and other diseases.

Regulation of Proto-Oncogenes by Splicing Factors

The activation of proto-oncogenes is one of the key features of the initial stages of cancer. There are some examples where alternative splicing events show a strong effect on the activity of some proto-oncogenes. We will further discuss some of the best-documented examples corresponding to Cyclin D1 and H-ras.

Cyclin D1 regulates cell cycle progression through its association with CDK4/6 [72]. It is currently known that cyclin D1 pre-mRNA possess two alternative polyadenylation sites and each isoform produced has different oncogenic activity. The more common product is the full-length peptide Cyclin D1a, where the 5 exons of the pre-mRNA are included. The alternative isoform is Cyclin D1b, which utilizes a polyadenylation site in intron 4 [73]. Both isoforms have the ability to interact with CDK4 and to regulate its activity in a similar fashion. Cyclin D1b is up-regulated in breast and prostate cancer [74, 75]. One of the main differences between the two isoforms consists on its cell localization: while Cyclin D1b resides in the nucleus, D1a shuttles between the nucleus and the cytoplasm according to the cell cycle progression. This difference could be related to the loss of a phosphorylation site at the C-terminus of Cyclin D1, which is target for the glycogen synthase kinase β [76, 77]. This site has been mutated and the result is the constitutive nuclear localization and a more oncogenic activity [78], in accordance to the functional effect observed for the alternative isoform Cyclin D1b. Regarding splicing regulation of this pre-mRNA, a common polymorphism located near the 5'ss of exon 4 seems to be responsible for this alternative decision [73, 79]. The G870A polymorphism lies at the last nucleotide of exon 4 and may affect exon 4 recognition by the spliceosome, favoring polyadenylation at an intronic position and producing isoform D1b. Binding sites for the splicing factor Sam68 have been identified in intron 4 and a correlation between Sam68 expression and D1b production has been reported [80]. A similar correlation was observed for SRSF1 [81]. Moreover, the G879 allele is associated with an increased risk for different types of cancer, supporting the relevance of D1b production in tumorigenesis [72].

The H-Ras proto-oncogene is another example of alternative splicing regulation affecting proliferation that occurs when an intronic mutation generates an increase in H-ras expression, augmenting the transforming potential of this gene [82]. This mutation disrupts the 5'ss of the alternative exon IDX [83]. H-Ras pre-mRNA can be alternatively spliced in the IDX and 4A terminal exons, yielding the p19 and p21 proteins, respectively. IDX inclusion generates an in-frame stop codon that could be a target for Non-sense Mediated mRNA Decay (NMD). As mentioned before, alternative products can show different cell localization: H-Ras p19 localizes in the nucleus and binds p53, but it can also bind p73 and promote its transcription [84]. In agreement with this, knocking down H-Ras p19 increases proliferation in different cell lines [85, 86]. p19 also induces FOXO1 and a delay in G1/S phase, which correlates with the hypophosphorylation of both Akt and p70SK6, leading to the maintenance of a reversible cellular quiescence state and preventing apoptosis [87]. hnRNP A1 and the RNA helicase p68 inhibit IDX inclusion through the interaction with an ISS located downstream of the alternative exon [88]. On the other hand, hnRNP H, SRSF2 and SRSF5 promote IDX inclusion [89].

Some splicing factors have been linked to the modulation of some apoptosis-related transcripts and could be considered as apoptosis regulators. This is the case of TIA-1/TIAR factors. TIA-1 and TIAR are highly conserved mammalian proteins [90] that show similarity with U1 snRNP-associated protein Nam8 from *Saccharomyces cerevisiae* and its ability to regulate alternative splicing events has been demonstrated [91, 92, 93]. TIA-1/TIAR favors the production of the proapoptotic form of Fas [94] and increases proliferation of HeLa cells [95]. Besides, TIA-1 and TIAR regulate alternative splicing of exon IIIb in the fibroblast growth factor receptor 2 (FGFR2) pre-mRNA and they regulate the inflammatory response under stress conditions by silencing the translation of TNF- α and COX-2 [96, 97]. These observations strongly suggest the important role of TIA-1/TIAR in growth and survival of cancer cells and its role in cancer progression.

The Role of Alternative Splicing in Cancer Progression

As discussed earlier, alternative splicing regulates the expression of diverse genes that show important roles in tumor progression and invasion. The determinant role of splicing factors has been already summarized and it results evident how splicing genes and factors could correlate with cancer progression. Moreover, there are several genes that are associated with metastasis and invasion, which are targets for alternative splicing. Diverse cell processes involve alternative splicing regulation of key genes, modulating almost every aspect of cell fate (Figure 2). We will further present different examples of determinant alternative splicing events that could be related to cell invasion and metastasis.

CD44 is a transmembrane protein with a rather complicated pattern of alternative splicing, which involves 10 adjacent alternative exons that can be included solo or in combination [98]. CD44 was one of the first genes for which AS was shown to be altered by signaling pathways associated with growth, involving Ras signaling and Ras–Raf–MEK–ERK pathway [99, 100], where ERF phosphorylated the splicing factor Sam68, which in turn activates the inclusion of exon v5. This finding was relevant because for the first time the connection between a mitogenic signaling pathway and splicing control was demonstrated [101].

FGFR genes encode four closely related tyrosine kinases receptors and FGFR1-3 show several mutually exclusive exons that can produce different proteins according to the cellular context and development through alternative splicing [102]. hnRNP proteins, mainly hnRNP A1, F/H and PTB are involved in the regulation of these splicing events [103-105].

Rac1 transcripts can also generate alternative splicing isoforms with important repercussions in cancer cells. Rac1 pre-mRNA is alternatively spliced to produce Rac1 and Rac1b. While Rac1 activates transcription by NF κ B and AKT kinase [106-107], Rac1b includes the internal 57-nt exon 3b and shows less GTPase activity but high GDP/GTP exchange [108]. Rac1b has exhibited a tumor-specific expression in colorectal and breast tumors and is up-regulated in metastatic tissues [109, 110]. When splicing regulation of this event was studied, it was found that SRSF1 overexpression resulted in increased exon 3b inclusion, while SRSF3 and SRSF7 had the opposite effect [111] and Wnt/ β catenin pathway appeared to be involved [112].

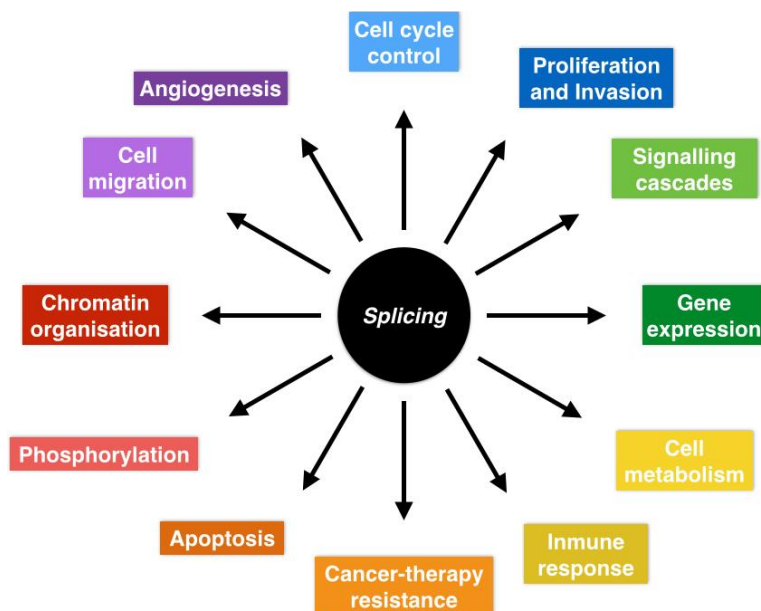


Figure 2. Alternative splicing and cell biology. Almost any kind of gene can be a target for alternative splicing and this mechanism can influence practically every cellular mechanism that could be related to cancer progression and tumorigenesis.

Alternative splicing of Ron transcripts has also been related to the invasive phenotype in cancer cell. Under normal conditions, Ron stimulation activates signaling pathways that enhance invasive growth [113, 114]. An alternative variant lacking exon 11 (Δ -Ron) and several isoforms have been found in epithelial cancers, including colorectal and breast cancer; the expression of these isoforms correlates with metastasis [113, 115]. The production of Δ -Ron is regulated by SRSF1 [113] and it has been proposed that Ron splicing is a key event responsible for the effect of SRSF1 in promoting metastasis [116, 117].

Closely related to metastasis, angiogenesis is a key hallmark of cancer and VEGFA, which promote the formation on new blood vessels in tumors, presents several transcripts that are alternatively spliced quite frequently [118]. In this case, SRSF1, SRSF5 and SRSF6 contribute to the production of the pro-angiogenic form of VEGFA [119].

Therapeutic Tools Developed to Regulate Splicing Factors Activity and Cancer

Considering the high impact of alternative splicing on gene expression, different molecular tools have been developed in order to correct or redirect aberrant splicing events [120]. The initial strategies applied to regulate alternative splicing focused on modifying this process using nucleic acids or nucleic acids analogs such as short oligonucleotides, which could be designed either to silence or to enhance the expression of a particular gene (Figure 1). These nucleic acids analogs could be single stranded antisense oligonucleotides with a complementary sequence for a specific region in certain gene. This molecule has the ability to target a specific mRNA and regulate its expression. Usually, the confirmation of this regulation is tested both

in vitro and *in vivo*. In some cases, these oligonucleotides include a sequence that is complementary for a regulatory element contained in the mRNA. These elements could be targets for auxiliary proteins, like SR and hnRNPs or they could be mutated sequences and restoring the wild-type sequence could alleviate the normal splicing regulation. For example, an oligonucleotide directed to regions located at or close to a splice site can mask normal or aberrant splicing events leading either to exon exclusion or inclusion and an artificial oligo could correct this aberrant event. In order to increase the half-life of these molecules and their resistance to nucleases, modulatory oligonucleotides have been modified being more effective due to a longer life into the cell. One modification involves their association with peptidic adapters, which are called peptide-nucleic acids (PNAs). Another chemical modifications include phosphorodiamidate morpholino oligos (PMOs), 2'-OMe (2'-ortho-methyl) or splice-switching oligonucleotides (SSOs). All these molecules are effective *in vitro* and some of them have been tested in several clinical trials. The most common diseases in which the treatment involves splicing modulatory molecules correspond to different types of atrophies. In order to increase availability, it has been necessary to develop some strategies to introduce the nucleic acid or analog into the cell, such as cell-penetrating peptides (CPPs), which are based on antimicrobial peptides and have been successfully applied (Figure 1). CPPs like Penetratin and Transportan are efficient delivery vectors that have been tested both *in vivo* and *in vitro* to carry several splicing-regulatory oligonucleotides with high efficiency and low toxicity [121].

Microbial derivatives like Spliceostatin (SSA) are newly discovered molecules that modulate alternative splicing and have shown effective anti-proliferative and anti-cancer activities [122]. Spliceostatin was identified in *Pseudomonas* sp. No. 2663, but since this initial discovery, similar molecules have been isolated from different bacterial strains [123, 124]. Spliceostatin mechanism of action involves the interaction with an essential component of the spliceosome, SF3b (Figure 1). Herboxidiene is another molecule that targets the spliceosome and inhibits proliferation. In HeLa cells treated with herboxidiene, translation of the intron 1–containing pre-mRNA leads to production of a C-terminal truncated protein isoform p27*, which is resistant to proteasomal degradation and inhibits CDK2 kinase activity, thereby inhibiting cell growth [125]. SSA treatment also leads to intron retention in VEGF and results in reduction of VEGF levels (possibly by NMD), inhibiting cancer cell angiogenesis [126]. Although the exact mechanism of selective tumor cytotoxicity remains to be fully explored, one explanation is that growth of cancer cells often relies on oncogenic protein isoforms (arising from alternative splicing), which are lacking in normal cells.

CONCLUSION

Alternative splicing is already a sophisticated mechanism of precise regulation. The key elements of this process are the central and accessory splicing factors. The role of alternative splicing in cancer becomes more and more studied and there is increasing evidence linking gene regulation to cancer diagnosis, progression and treatment. The implication of splicing factors in cancer still is under development. Further insights in the genetic origin of cancer, the splicing factors interacting with particular genes and its relationship to eukaryotic gene expression regulation could provide additional information that may be applied to several aspects of cancer. Moreover, splicing factors could be considered as new drug targets for many

diseases, including cancer. According to the importance of this mechanism and the general repercussions of this process in cell fate, it becomes logical that the development of different molecules and compounds to modulate alternative splicing should occupy immediate research. Technical improvements and additional basic studies could provide the appropriate tools to fight this prevalent disease.

REFERENCES

- [1] Lander ES, Linton LM, Birren B, Nusbaum C, et al. 2001. Initial sequencing and analysis of the human genome. *Nature* 409: 860–921.
- [2] Tazi J, Bakkour N, Stamm S. 2009. Alternative splicing and disease. *Biochim. Biophys. Acta* 1792: 14–26.
- [3] Martínez-Contreras R, Galindo JM, Aguilar-Rojas A, Valdés J. Two exonic elements in the flanking constitutive exons control the alternative splicing of the alpha exon of the ZO-1 pre-mRNA. *Biochim. Biophys. Acta* 2003; 1630: 71-83.
- [4] Jurica MS, Moore MJ. 2003. Pre-mRNA splicing: awash in a sea of proteins. *Mol. Cell* 12: 5-14.
- [5] Long JC, Caceres JF. 2009. The SR protein family of splicing factors: master regulators of gene expression. *Biochem. J.* 417: 15–27.
- [6] Martinez-Contreras R, Cloutier P, Shkreta L, Fisette JF, et al. 2007. hnRNP proteins and splicing control. *Adv. Exp. Med. Biol.* 623: 123-47.
- [7] Mayeda A, Helfman DM, Krainer AR. 1993. Modulation of exon skipping and inclusion by heterogeneous nuclear ribonucleoprotein A1 and pre-mRNA splicing factor SF2/ASF. *Mol. Cell Biol.* 13: 2993-3001.
- [8] Bai Y, Lee D, Yu T, Chasin LA. 1999. Control of 3' splice site choice in vivo by ASF/SF2 and hnRNP A1. *Nucleic Acids Res.* 27: 1126-34.
- [9] Expert-Bezançon A, Sureau A, Durosay P, Salesse R, et al. 2004. hnRNP A1 and the SR proteins ASF/SF2 and SC35 have antagonistic functions in splicing of beta-tropomyosin exon 6B. *J. Biol. Chem.* 279: 38249-59.
- [10] Guo R, Li Y, Ning J, Sun D, et al. 2013. HnRNP A1/A2 and SF2/ASF regulate alternative splicing of interferon regulatory factor-3 and affect immunomodulatory functions in human non-small cell lung cancer cells. *PLoS One* 8: e62729.
- [11] Martinez-Contreras R, Fisette JF, Nasim FU, Madden R, et al. 2006. Intronic binding sites for hnRNP A/B and hnRNP F/H proteins stimulate pre-mRNA splicing. *PLoS Biol.* 4: e21.
- [12] Huang Y, Steitz JA. 2005. SRprises along a messenger's journey. *Mol. Cell* 17: 613-5.
- [13] Chabot B. 2015. My road to alternative splicing control: from simple paths to loops and interconnections. *Biochem. Cell Biol.* 93: 171-9.
- [14] Johnson JM, Castle J, Garrett-Engle P, Kan Z, et al. 2003. Genome wide survey of human alternative pre-mRNA splicing with exon junction microarrays. *Science* 302: 2141–4.
- [15] Bonomi S, Gallo S, Catillo M, Pignataro D, et al. 2013. Oncogenic Alternative Splicing Switches: Role in Cancer Progression and Prospects for Therapy. *International Journal*

- of *Cell Biology*. Hindawi Publishing Corporation. Volume 2013, Article ID 962038, 17 pages.
- [16] Ladomery M. 2013. Aberrant Alternative Splicing Is Another Hallmark of Cancer. *International Journal of Cell Biology*. Hindawi Publishing Corporation. Volume 2013, Article ID 463786, 6 pages.
 - [17] Mucaki EJ, Ainsworth P, Rogan PK. 2011. Comprehensive prediction of mRNA splicing effects of BRCA1 and BRCA2 variants. *Hum. Mutat.* 32: 735–42.
 - [18] Whiley PJ, Guidugli L, Walker LC, Healey S, et al. 2011. Splicing and multifactorial analysis of intronic BRCA1 and BRCA2 sequence variants identifies clinically significant splicing aberrations up to 12 nucleotides from the intron/exon boundary. *Hum. Mutat.* 32: 678–87.
 - [19] Hagen RM, Ladomery MR. 2012. Role of splice variants in the metastatic progression of prostate cancer. *Biochem. Soc. Trans* 40: 870–4.
 - [20] Oltean S, Bates DO. 2013. Hallmarks of alternative splicing in cancer. *Oncogene* 33: 5311–8.
 - [21] Zhou Z, Fu X-D. 2013. Regulation of splicing by SR proteins and SR protein-specific kinases. *Chromosoma* 122: 191–207.
 - [22] Leva V, Giuliano S, Bardoni A, Camerini S, et al. 2012. Phosphorylation of SRSF1 is modulated by replicational stress. *Nucleic Acids Res.* 40: 1106–17.
 - [23] Kaneb HM, Dion PA, Rouleau GA. 2012. The FUS about arginine methylation in ALS and FTL. *EMBO J.* 31: 4249–51.
 - [24] Sinha R, Allemand E, Zhang Z, Karni R, et al. 2010. Arginine methylation controls the subcellular localization and functions of the oncoprotein splicing factor SF2/ASF. *Mol. Cell Biol.* 30: 2762–74.
 - [25] Rouvière JO, Geoffroy M-C, Palancade B. 2013. Multiple crosstalks between mRNA biogenesis and SUMO. *Chromosoma* 122: 387–99.
 - [26] White ES, Sagana RL, Booth AJ, Yan M, et al. 2010. Control of fibroblast fibronectin expression and alternative splicing via the PI3K/Akt/mTOR pathway. *Exp. Cell Res.* 316: 2644–53.
 - [27] Shultz JC, Goehe RW, Wijesinghe DS, Murudkar C, et al. 2010. Alternative splicing of caspase 9 is modulated by the phospho-inositide 3-kinase/Akt pathway via phosphorylation of SRp30a. *Cancer Res.* 70: 9185–96.
 - [28] Patel NA, Kaneko S, Apostolatos HS, Bae SS, et al. 2005. Molecular and genetic studies imply Akt-mediated signaling promotes protein kinase C β II alternative splicing via phosphorylation of serine/arginine-rich splicing factor SRp40. *J. Biol. Chem.* 280: 14302–9.
 - [29] Druillennec S, Dorard C, Eychène A. 2012. Alternative splicing in oncogenic kinases: from physiological functions to cancer. *J. Nucleic Acids* 2012: 639062.
 - [30] Yoshida K, Sanada M, Shiraishi Y, Nowak D, et al. 2011. Frequent pathway mutations of splicing machinery in myelodysplasia. *Nature* 478: 64–9.
 - [31] Wang L, Lawrence MS, Wan Y, Stojanov P, et al. 2011. SF3B1 and other novel cancer genes in chronic lymphocytic leukemia. *N. Engl. J. Med.* 365: 2497–506.
 - [32] Rossi D, Brusca A, Spina V, Rasi S, et al. 2011. Mutations of the SF3B1 splicing factor in chronic lymphocytic leukemia: association with progression and fludarabine-refractoriness. *Blood* 118: 6904–8.

- [33] Quesada V, Conde L, Villamor N, Ordonez GR, et al. 2012. Exome sequencing identifies recurrent mutations of the splicing factor SF3B1 gene in chronic lymphocytic leukemia. *Nat. Genet.* 44: 47 – 52.
- [34] Harbour JW, Roberson ED, Anbunathan H, Onken MD, et al. 2013. Recurrent mutations at codon 625 of the splicing factor SF3B1 in uveal melanoma. *Nat. Genet.* 45: 133– 5.
- [35] Martin M, Masshofer L, Temming P, Rahmann S, et al. 2013. Exome sequencing identifies recurrent somatic mutations in EIF1AX and SF3B1 in uveal melanoma with disomy 3. *Nat. Genet.* 45: 933 – 6.
- [36] Malcovati L, Papaemmanuil E, Bowen DT, Boulton J, et al. 2011. Clinical significance of SF3B1 mutations in myelodysplastic syndromes and myelodysplastic/myeloproliferative neoplasms. *Blood* 118: 6239 – 46.
- [37] Patnaik MM, Lasho TL, Hodnefield JM, Knudson RA, et al. 2012. SF3B1 mutations are prevalent in myelodysplastic syndromes with ring sideroblasts but do not hold independent prognostic value. *Blood* 119: 569 – 72.
- [38] Makishima H, Visconte V, Sakaguchi H, Jankowska AM, et al. 2012. Mutations in the spliceosome machinery, a novel and ubiquitous pathway in leukemogenesis. *Blood* 119: 3203 – 10.
- [39] Hirabayashi S, Flotho C, Moetter J, Heuser M, et al. 2012. Spliceosomal gene aberrations are rare, coexist with oncogenic mutations, and are unlikely to exert a driver effect in childhood MDS and JMML. *Blood* 119: e96 – 9.
- [40] Jia R, Li C, McCoy JP, Deng CX, et al. 2010. SRp20 is a protooncogene critical for cell proliferation and tumor induction and maintenance. *Int. J. Biol. Sci.* 6: 806 – 26.
- [41] Kurokawa K, Akaike Y, Masuda K, Kuwano Y, et al. 2013. Downregulation of serine/arginine-rich splicing factor 3 induces G1 cell cycle arrest and apoptosis in colon cancer cells. *Oncogene* 33: 1407-17.
- [42] Shepard PJ, Hertel KJ. 2009. The SR protein family. *Genome Biol.* 10: 242.
- [43] Fiset S, Chabot B. 2001. hnRNP A1 may interact simultaneously with telomeric DNA and the human telomerase RNA in vitro. *Nucleic Acids Res.* 29: 2268-75.
- [44] David CJ, Chen M, Assanah M, Canoll P, et al. 2010. HnRNP proteins controlled by c-Myc deregulate pyruvate kinase mRNA splicing in cancer. *Nature* 463: 364 – 8.
- [45] Patry C, Bouchard L, Labrecque P, Gendron D, et al. 2003. Small interfering RNA-mediated reduction in heterogeneous nuclear ribonucleoparticle A1/A2 proteins induces apoptosis in human cancer cells but not in normal mortal cell lines. *Cancer Res.* 63: 7679 – 88.
- [46] Golan-Gerstl R, Cohen M, Shilo A, Suh SS, et al. 2011. Splicing factor hnRNP A2/B1 regulates tumor suppressor gene splicing and is an oncogenic driver in glioblastoma. *Cancer Res.* 71: 4464 – 72.
- [47] Huelga SC, Vu AQ, Arnold JD, Liang TY, et al. 2012. Integrative genome-wide analysis reveals cooperative regulation of alternative splicing by hnRNP proteins. *Cell Rep.* 1: 167– 78.
- [48] Christofk HR, Vander Heiden MG, Harris MH, Ramanathan A, et al. 2008. The M2 splice isoform of pyruvate kinase is important for cancer metabolism and tumour growth. *Nature* 452: 230 – 3.
- [49] Xue Y, Zhou Y, Wu T, Zhu T, et al. 2009. Genome-wide analysis of PTB-RNA interactions reveals a strategy used by the general splicing repressor to modulate exon inclusion or skipping. *Mol. Cell* 36: 996 – 1006.

- [50] Polydorides AD, Okano HJ, Yang YY, Stefani G, et al. 2000. A brain-enriched polypyrimidine tract-binding protein antagonizes the ability of Nova to regulate neuron-specific alternative splicing. *Proc. Natl. Acad. Sci. USA*. 97: 6350-5.
- [51] Ule J, Stefani G, Mele A, Ruggiu M, et al. 2006. An RNA map predicting Nova-dependent splicing regulation. *Nature* 444: 580-6.
- [52] Llorian M, Schwartz S, Clark TA, Hollander D, et al. 2010. Position-dependent alternative splicing activity revealed by global profiling of alternative splicing events regulated by PTB. *Nat. Struct. Mol. Biol.* 17: 1114–23.
- [53] Izaguirre DI, Zhu W, Hai T, Cheung HC, et al. 2012. PTBP1- dependent regulation of USP5 alternative RNA splicing plays a role in glioblastoma tumorigenesis. *Mol. Carcinog.* 51: 895 – 906.
- [54] Yamaguchi F, Saya H, Bruner JM, Morrison RS. 1994. Differential expression of two fibroblast growth factor-receptor genes is associated with malignant progression in human astrocytomas. *Proc. Natl. Acad. Sci. USA* 91: 484 – 8.
- [55] Mulherkar N, Prasad KV, Prabhakar BS. 2007. MADD/DENN splice variant of the IG20 gene is a negative regulator of caspase-8 activation. Knockdown enhances TRAIL-induced apoptosis of cancer cells. *J. Biol. Chem.* 282: 11715 – 21.
- [56] Bonnal S, Martinez C, Forch P, Bachi A, et al. 2008. RBM5/ Luca-15/H37 regulates Fas alternative splice site pairing after exon definition. *Mol. Cell.* 32: 81–95.
- [57] Seth P, Yeowell HN. 2010. Fox-2 protein regulates the alternative splicing of scleroderma- associated lysyl hydroxylase 2 messenger RNA. *Arthritis Rheum.* 62: 1167–75.
- [58] Wang K, Bacon ML, Tessier JJ, Rintala-Maki ND, et al. 2012. RBM10 Modulates Apoptosis and Influences TNF- α Gene Expression. *J. Cell Death* 5: 1–19.
- [59] David CJ, Manley JL. 2010. Alternative pre-mRNA splicing regulation in cancer: pathways and programs unhinged. *Genes Dev.* 24: 2343–64.
- [60] Karni R, de Stanchina E, Lowe SW, Sinha R, et al. 2007. The gene encoding the splicing factor SF2/ASF is a proto-oncogene. *Nat. Struct. Mol. Biol.* 14: 185–93.
- [61] Fu Y, Huang B, Shi Z, Han J, et al. 2013. SRSF1 and SRSF9 RNA binding proteins promote Wnt signalling-mediated tumorigenesis by enhancing-catenin biosynthesis. *EMBO Mol. Med.* 5: 737–50.
- [62] Golan-Gerstl R, Cohen M, Shilo A, Suh S-S, et al. 2011. Splicing factor hnRNP A2/B1 regulates tumor suppressor gene splicing and is an oncogenic driver in glioblastoma. *Cancer Res.* 71: 4464–72.
- [63] Lefave CV, Squatrito M, Vorlova S, Rocco GL, et al. 2011. Splicing factor hnRNP H drives an oncogenic splicing switch in gliomas. *EMBO J.* 30: 4084–97.
- [64] Das S, Anczuków O, Akerman M, Krainer AR. 2012. Oncogenic splicing factor SRSF1 is a critical transcriptional target of MYC. *Cell Rep.* 1: 110–7.
- [65] Ge K, DuHadaway J, Du W, Herlyn M, et al. 1999. Mechanism for elimination of a tumor suppressor: aberrant splicing of a brain-specific exon causes loss of function of Bin1 in melanoma. *Proc. Natl. Acad. Sci.* 96: 9689–94.
- [66] Anczuków O, Rosenberg AZ, Akerman M, Das S, et al. 2012. The splicing factor SRSF1 regulates apoptosis and proliferation to promote mammary epithelial cell transformation. *Nat. Struct. Mol Biol.* 19: 220–8.

- [67] Shimoni-Sebag A, Lebenthal-Loinger I, Zender L, Karni R. 2013. RRM1 domain of the splicing oncoprotein SRSF1 is required for MEK1-MAPK-ERK activation and cellular transformation. *Carcinogenesis* 34: 2498–504.
- [68] Amin EM, Oltean S, Hua J, Gammons MVR, et al. 2011. WT1 mutants reveal SRPK1 to be a downstream angiogenesis target by altering VEGF splicing. *Cancer Cell* 20: 768–80.
- [69] Sun S, Zhang Z, Sinha R, Karni R, et al. 2010. SF2/ASF autoregulation involves multiple layers of posttranscriptional and translational control. *Nat. Struct. Mol. Biol.* 17: 306–12.
- [70] Verduci L, Simili M, Rizzo M, Mercatanti A, et al. 2010. MicroRNA (miRNA)-mediated interaction between leukemia/lymphoma-related factor (LRF) and alternative splicing factor/splicing factor 2 (ASF/SF2) affects mouse embryonic fibroblast senescence and apoptosis. *J. Biol. Chem.* 285: 39551–63.
- [71] Akaike Y, Kurokawa K, Kajita K, Kuwano Y, et al. 2011. Skipping of an alternative intron in the srsf1 3' untranslated region increases transcript stability. *J. Med. Invest.* 58: 180–7.
- [72] Knudsen KE, Diehl JA, Haiman CA, Knudsen ES. 2006. Cyclin D1: Polymorphism, aberrant splicing and cancer risk. *Oncogene* 25: 1620–8.
- [73] Betticher DC, Thatcher N, Altermatt HJ, Hoban P, et al. 1995. Alternate splicing produces a novel cyclin D1 transcript. *Oncogene* 11: 1005–11.
- [74] Burd CJ, Petre CE, Morey LM, Wang Y, et al. 2006. Cyclin D1b variant influences prostate cancer growth through aberrant androgen receptor regulation. *Proc. Natl. Acad. Sci.* 103: 2190–5.
- [75] Wang Y, Dean JL, Millar EK, Tran TH, et al. 2008. Cyclin D1b is aberrantly regulated in response to therapeutic challenge and promotes resistance to estrogen antagonists. *Cancer Res.* 68: 5628–38.
- [76] Lu F, Gladden AB, Diehl JA. 2003. An alternatively spliced cyclin D1 isoform, cyclin D1b, is a nuclear oncogene. *Cancer Res.* 63: 7056–61.
- [77] Solomon DA, Wang Y, Fox SR, Lambeck TC, et al. 2003. Cyclin D1 splice variants. Differential effects on localization, RB phosphorylation, and cellular transformation. *J. Biol. Chem.* 278: 30339–47.
- [78] Alt JR, Cleveland JL, Hannink M, Diehl JA. 2000. Phosphorylation-dependent regulation of cyclin D1 nuclear export and cyclin D1-dependent cellular transformation. *Genes Dev.* 14: 3102–14.
- [79] Comstock CE, Augello MA, Benito RP, Karch J, et al. 2009. Cyclin D1 splice variants: Polymorphism, risk, and isoformspecific regulation in prostate cancer. *Clin. Cancer Res.* 15: 5338–49.
- [80] Paronetto MP, Cappellari M, Busa R, Pedrotti S, et al. 2010. Alternative splicing of the cyclin D1 proto-oncogene is regulated by the RNA-binding protein Sam68. *Cancer Res.* 70: 229–39.
- [81] Olshavsky NA, Comstock CE, Schiewer MJ, Augello MA, 2010. Identification of ASF/SF2 as a critical, allele-specific effector of the cyclin D1b oncogene. *Cancer Res.* 70: 3975–84.
- [82] Cohen JB, Levinson AD. 1988. A point mutation in the last intron responsible for increased expression and transforming activity of the c-Ha-ras oncogene. *Nature* 334: 119–24.

- [83] Cohen JB, Broz SD, Levinson AD. 1989. Expression of the H-ras proto-oncogene is controlled by alternative splicing. *Cell* 58: 461–72.
- [84] Jeong MH, Bae J, Kim WH, Yoo SM, et al. 2006. p19ras interacts with and activates p73 by involving the MDM2 protein. *J. Biol. Chem.* 281: 8707–15.
- [85] Guil S, de La Iglesia N, Fernandez-Larrea J, Cifuentes D, et al. 2003. Alternative splicing of the human proto-oncogene c-H-ras renders a new Ras family protein that trafficks to cytoplasm and nucleus. *Cancer Res.* 63: 5178–87.
- [86] Jang SM, Kim JW, Kim CH, Kim D, et al. 2010. p19 ras represses proliferation of non-small cell lung cancer possibly through interaction with neuron-specific enolase (NSE). *Cancer Lett.* 289: 91–8.
- [87] Camats M1, Kokolo M, Heesom KJ, Lodomery M, et al. 2009. P19 H-ras induces G1/S phase delay maintaining cells in a reversible quiescence state. *PLoS One* 4: e8513.
- [88] Camats M, Guil S, Kokolo M, Bach-Elias M. 2008. P68 RNA helicase (DDX5) alters activity of cis- and trans-acting factors of the alternative splicing of H-Ras. *PLoS ONE* 3: e2926.
- [89] Guil S, Gattoni R, Carrascal M, Abian J, et al. 2003. Roles of hnRNP A1, SR proteins, and p68 helicase in c-H-ras alternative splicing regulation. *Mol. Cell Biol.* 23: 2927–41.
- [90] Beck AR, Medley QG, O'Brien S, Anderson P, et al. 1996. Structure, tissue distribution and genomic organization of the murine RRM-type RNA binding proteins TIA-1 and TIAR. *Nucleic Acids Res.* 24: 3829–3835.
- [91] Gottschalk A, Tang J, Puig O, Salgado J, et al. 1998. A comprehensive biochemical and genetic analysis of the yeast U1 snRNP reveals five novel proteins. *RNA* 4: 374–393.
- [92] Del Gatto-Konczak F, Bourgeois CF, Le Guiner C, Kister L, et al. 2000. The RNA binding protein TIA-1 is a novel mammalian splicing regulator acting through intron sequences adjacent to a 59 splice site. *Mol. Cell Biol.* 20: 6287–99.
- [93] Forch P, Puig O, Kedersha N, Martinez C, et al. 2000. The apoptosis promoting factor TIA-1 is a regulator of alternative pre-mRNA splicing. *Mol Cell* 6: 1089–98.
- [94] Izquierdo JM, Valcarcel J. 2007. Fas-activated serine/threonine kinase (FAST K) synergizes with TIA-1/TIAR proteins to regulate Fas alternative splicing. *J. Biol. Chem.* 282: 1539–43.
- [95] Reyes R, Alcalde J, Izquierdo JM. 2009. Depletion of T-cell intracellular antigen proteins promotes cell proliferation. *Genome Biol.* 10: R87.
- [96] Piecyk M, Wax S, Beck AR, Kedersha N, et al. 2000. TIA-1 is a translational silencer that selectively regulates the expression of TNF- α . *EMBO J* 19: 4154–63.
- [97] Lopez de Silanes I, Galban S, Martindale JL, Yang X, et al. 2005. Identification and functional outcome of mRNAs associated with RNA-binding protein TIA-1. *Mol. Cell Biol.* 25: 9520–31.
- [98] Ponta H, Sherman L, Herrlich PA. 2003. CD44: From adhesion molecules to signalling regulators. *Nat. Rev. Mol. Cell Biol.* 4: 33–45.
- [99] Konig H, Ponta H, Herrlich P. 1998. Coupling of signal transduction to alternative pre-mRNA splicing by a composite splice regulator. *EMBO J.* 17: 2904–13.
- [100] Cheng C, Yaffe MB, Sharp PA. 2006. A positive feedback loop couples Ras activation and CD44 alternative splicing. *Genes Dev.* 20: 1715–20.
- [101] Matter N, Herrlich P, Konig H. 2002. Signal-dependent regulation of splicing via phosphorylation of Sam68. *Nature* 420: 691–5.

- [102] Carstens RP, Eaton JV, Krigman HR, Walther PJ, et al. 1997. Alternative splicing of fibroblast growth factor receptor 2 (FGF-R2) in human prostate cancer. *Oncogene* 15: 3059–65.
- [103] Del Gatto-Konczak F, Olive M, Gesnel MC, Breathnach R. 1999. hnRNP A1 recruited to an exon in vivo can function as an exon splicing silencer. *Mol. Cell Biol.* 19: 251–60.
- [104] Carstens RP, Wagner EJ, Garcia-Blanco MA. 2000. An intronic splicing silencer causes skipping of the IIIb exon of fibroblast growth factor receptor 2 through involvement of polypyrimidine tract binding protein. *Mol. Cell Biol.* 20: 7388–400.
- [105] Mauger DM, Lin C, Garcia-Blanco MA. 2008. hnRNP H and hnRNP F complex with Fox2 to silence fibroblast growth factor receptor 2 exon IIIc. *Mol. Cell Biol.* 28: 5403–19.
- [106] Keely PJ, Westwick JK, Whitehead IP, Der CJ, et al. 1997. Cdc42 and Rac1 induce integrin-mediated cell motility and invasiveness through PI(3)K. *Nature* 390: 632–6.
- [107] Perona R, Montaner S, Saniger L, Sanchez-Perez I, et al. 1997. Activation of the nuclear factor- κ B by Rho, CDC42, and Rac-1 proteins. *Genes Dev.* 11: 463–75.
- [108] Fiegen D, Haeusler LC, Blumenstein L, Herbrand U, et al. 2004. Alternative splicing of Rac1 generates Rac1b, a self-activating GTPase. *J. Biol. Chem.* 279: 4743–9.
- [109] Jordan P, Brazao R, Boavida MG, Gespach C, et al. 1999. Cloning of a novel human Rac1b splice variant with increased expression in colorectal tumors. *Oncogene* 18: 6835–6839.
- [110] Schnelzer A, Prectel D, Knaus U, Dehne K, et al. 2000. Rac1 in human breast cancer: Overexpression, mutation analysis, and characterization of a new isoform, Rac1b. *Oncogene* 19: 3013–20.
- [111] Goncalves V, Matos P, Jordan P. 2009. Antagonistic SR proteins regulate alternative splicing of tumor-related Rac1b downstream of the PI3-kinase and Wnt pathways. *Hum. Mol. Genet.* 18: 3696–707.
- [112] Goncalves V, Matos P, Jordan P. 2008. The b-catenin/TCF4 pathway modifies alternative splicing through modulation of SRp20 expression. *RNA* 14: 2538–49.
- [113] Ghigna C, Giordano S, Shen H, Benvenuto F, et al. 2005. Cell motility is controlled by SF2/ASF through alternative splicing of the Ron protooncogene. *Mol. Cell* 20: 881–90.
- [114] Wagh PK, Peace BE, Waltz SE. 2008. Met-related receptor tyrosine kinase Ron in tumor growth and metastasis. *Adv. Cancer Res.* 100: 1–33.
- [115] Zhou YQ, He C, Chen YQ, Wang D, et al. 2003. Altered expression of the RON receptor tyrosine kinase in primary human colorectal adenocarcinomas: Generation of different splicing RON variants and their oncogenic potential. *Oncogene* 22: 186–97.
- [116] Bardella C, Costa B, Maggiora P, Patane S, et al. 2004. Truncated RON tyrosine kinase drives tumor cell progression and abrogates cell-cell adhesion through E-cadherin transcriptional repression. *Cancer Res* 64: 5154–61.
- [117] Wang D, Shen Q, Chen YQ, Wang MH. 2004. Collaborative activities of macrophage-stimulating protein and transforming growth factor- β 1 in induction of epithelial to mesenchymal transition: Roles of the RON receptor tyrosine kinase. *Oncogene* 23: 1668–80.
- [118] Harper SJ, Bates DO. 2008. VEGF-A splicing: The key to antiangiogenic therapeutics? *Nat. Rev. Cancer* 8: 880–7.

- [119] Nowak DG, Woolard J, Amin EM, Konopatskaya O, et al. 2008. Expression of pro- and anti-angiogenic isoforms of VEGF is differentially regulated by splicing and growth factors. *J. Cell Sci.* 121: 3487–95.
- [120] Arechavala-Gomez V, Khoo B, Aartsma-Rus A. 2014. Splicing modulation therapy in the treatment of genetic diseases. *Appl. Clin. Genet.* 7: 245-52.
- [121] Martínez-Montiel N, Rosas-Murrieta N, Martínez-Contreras R. 2015. Peptidic tools applied to redirect alternative splicing events. *Peptides* 67: 1-11.
- [122] Kaida D, Motoyoshi H, Tashiro E, Nojima T, et al. 2007. Spliceostatin A targets SF3b and inhibits both splicing and nuclear retention of pre-mRNA. *Nat. Chem. Biol.* 2007; 3: 576-83.
- [123] Liu X, Biswas S, Tang GL, Cheng YQ. 2013. Isolation and characterization of spliceostatin B, a new analogue of FR901464, from *Pseudomonas* sp. No. 2663. *J. Antibiot.* (Tokyo). 66: 555-8.
- [124] He H, Ratnayake AS, Janso JE, He M, et al. 2014. Cytotoxic Spliceostatins from *Burkholderia* sp. and Their Semisynthetic Analogues. *J. Nat. Prod.* 77: 1864-70.
- [125] Hasegawa M1, Miura T, Kuzuya K, Inoue A, et al. 2011. Identification of SAP155 as the target of GEX1A (Herboxidiene), an antitumor natural product. *ACS Chem. Biol.* 6: 229-33.
- [126] Furumai R, Uchida K, Komi Y, Yoneyama M, et al. 2010. Spliceostatin A blocks angiogenesis by inhibiting global gene expression including VEGF. *Cancer Sci.* 101: 2483 – 9.

Chapter 18

NONSENSE-MEDIATED DECAY AND HUMAN DISEASE

Nancy Martínez-Montiel and Rebeca Martinez-Contreras

Molecular and Microbial Ecology Lab, Center of Research
on Microbiological Sciences, Science Institute,
Benemérita Universidad Autónoma de Puebla, Puebla, México

ABSTRACT

The precise control of protein production is essential for the appropriate cell physiology and survival. However, single mutations or accidentally introduced errors can occur during the flow of genetic information. In eukaryotic cells, some messenger RNA (mRNA) molecules leave the nucleus after splicing but further mechanisms are involved in the ultimate outcome of the correspondent protein. One of such systems corresponds to the mRNA surveillance network that includes nonsense-mediated mRNA decay (NMD), an important quality control system that ensures the accuracy of transcripts, to maintain a healthy cellular homeostasis. NMD eliminates anomalous mRNAs harboring premature termination codons (PTCs) to prevent the production of potentially harmful truncated proteins, but it can also regulate the steady state of many physiological mRNAs. Targets for NMD are sometimes linked to mutations or introduced errors but the vast majority can be generated as a result of alternative splicing.

The key components required for NMD include the UPF and SMG proteins. These factors interact with a set of proteins (the exon junction complex or EJC) that are deposited just upstream of exon-exon junctions after mRNA splicing and orchestrate a regulated mechanism in order to identify PTCs and either sequester these mRNAs or target them for degradation. Some of these factors are linked to particular disorders and could be modulated in order to correct the defect.

The NMD pathway is physiologically and medically important, because an escape from NMD can result in severe clinical phenotypes. Consistent with this, it is estimated that more than 60% of human genes have alternatively spliced products that generate at least one PTC isoform and that approximately 30% of inherited genetic disorders are caused by nonsense mutations or by frameshifts that generate nonsense codons. Initially associated as a genetic cause for beta-thalassemia, the NMD process has expanded from the haematopoietic system and it has reached different kind of human disorders, including

cystic fibrosis, Duchenne muscular dystrophy, Hurler syndrome, recessive spinal muscular atrophy and polycystic kidney disease. Finally, it is predicted that some cancers are aided by NMD and the repression of this mechanism may be a potential target for the treatment of certain cancers. In this regard, pharmacological agents have been developed for the treatment of diseases caused by premature stop mutations, including aminoglycosides, but the whole therapeutic potential of NMD targets to correct genetic disorders remains to be exploited.

INTRODUCTION

During the flow of gene expression, there are several co- and post-transcriptional events that ensure fidelity of the information and the functionality of the generated products. In eukaryotes, transcription is not only responsible for the synthesis of the mRNA, but it also includes the co-transcriptional maturation of the pre-mRNA corresponding to the 5' capping, splicing and 3' end formation. Altogether, these three processes occur while a pre-mRNA is still associated with a transcribing RNA polymerase II leading into an export, translation and decay pathway. Once in the cytoplasm, eukaryotic cells possess several quality control mechanisms to ensure the fidelity of gene expression, including the Nonsense-mediated mRNA decay (NMD) pathway.

Initially, NMD was described as a surveillance mechanism that detects and degrades mRNA containing premature termination codons (PTCs). This event was first reported for humans and *Saccharomyces cerevisiae* in 1979 [1, 2] and later it was also found in other analyzed eukaryotes, such as *Caenorhabditis elegans* [3], *Drosophila melanogaster* [4] and plants [5]. After these initial observations, it was also demonstrated that NMD not only targets PTC-containing mRNAs, but it can also sequester inappropriately spliced transcripts [6]. Moreover, it is now established that there are several types of mRNA molecules that can be targets for NMD, including some noncoding RNAs, mRNAs that contain an upstream open reading frame or a 3'UTR that contains an intron, selenocysteine codons during selenium depletion and mRNAs with an unusually long 3'UTR [6]. In this regard, several bioinformatic analyses suggest that more than 60% of human genes have alternatively spliced products that generate at least one PTC isoform [7]. Accordingly, the production of some of these PTC-containing isoforms and its degradation by NMD was experimentally validated [8, 9]. However, the physiological and pathological significance of NMD in terms of human health remains largely obscure.

Initially, PTC-generating mutations were linked to beta-thalassemia, but after this initial discovery, the number of mutations linked to NMD has been growing. The official report from the National Organization for Rare Disorders (NORD), considers nearly 7000 known rare genetic disorders and it is believed that 30% of these disorders occurs due to a nonsense or frameshift mutation that creates a PTC [10, 11].

NMD MECHANISM

In all eukaryotes studied, NMD requires a set of conserved regulatory factors [12], where the core components of the NMD machinery are the UPF proteins: UPF1, UPF2 and UPF3 [13

– 15] which are highly conserved. In multicellular organisms, NMD requires four additional NMD factors: SMG1, SMG5, SMG6 and SMG7, which regulate the phosphorylation state of UPF1 [6], controlling its functionality during NMD. When these four genes were mutated in *C. elegans*, morphological defects were observed in the male bursa or in the hermaphrodite vulva, giving rise to their name: Smg or suppressor of morphological defects on genitalia [16]. Besides these central proteins, a growing list of trans acting NMD factors has been described, which can vary among species. In mammals, the NMD pathway includes the components of the exon-junction complex (EJC) and SURF complexes, some Cap (CBP20 and CBP80) and PolyA binding proteins, splicing factors and components of the mRNA degradation pathway [17]. Despite the great conservation of some of these factors in different organisms, their expression and distribution fluctuates across species and even when various mechanistic models have been proposed [18, 19], we will focus here in the model established for human (Figure 1).

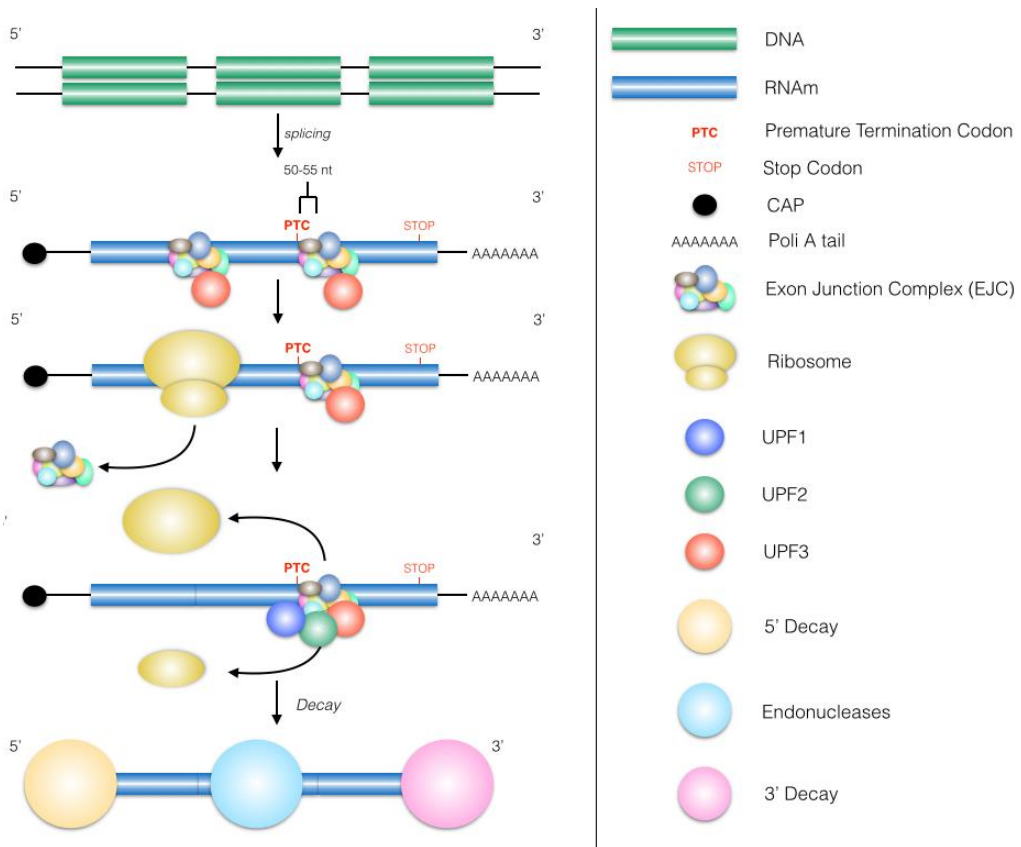


Figure 1. NMD mechanism. After splicing, EJC complex are deposited along the mRNA molecule. If a PTC is detected, the NMD machinery stops translation and arrests the mRNA. After this recognition, UPF1 and UPF2 are recruited to the complex and the ribosome is disassembled from the mRNA. Finally, degradation of the aberrant mRNA involves different decay processes and nucleolitic activities.

In mammals, the UPF2 and UPF3 proteins associate with the EJC, comprised by the Y14, MAGOH, eIF4AIII and MLN51 proteins. This EJC is normally deposited 20-24 nucleotides upstream of most exon-exon junctions during pre-mRNA splicing [20-22]. The classic NMD

route is usually triggered during the so-called pioneer round of translation when a nonsense codon lies at 50-55 nucleotides upstream of an EJC that has been deposited by the splicing machinery. The relative position of the stop codon determines how sensitive the mRNA could be to NMD or even if it can escape from degradation in many mammalian mRNAs. Moreover, this differential sensitivity for NMD could correlate with the severity of the disease [23]. PTC recognition occurs due to the formation of the SURF (SMG-1/SMG-8/SMG-9, Upf1 and eRF) complex at sites containing a PTC-recognizing ribosome and a downstream EJC. The transiently formed SURF complex detects the downstream EJC and forms a DECID complex, activating UPF1 through phosphorylation by SMG1. The SMG-5/SMG-7 complex uncouples UPF1 from the ribosome and release factor and finally SMG-6 promotes UPF1 dissociation from the mRNA (Figure 1). At the end, mRNA degradation is initiated by deadenylation in mammals and when approximately 110 nt have been deadenylated, mRNA is degraded from both ends with the additional involvement of the cleavage by the endonuclease SMG-6 [6].

NMD FACTORS RELATED TO PARTICULAR DISEASES

The most relevant evidence supporting the role of NMD in human genetic diseases comes from patients with developmental neural disorders that carry mutations in the UPF3B gene. Genetic studies in families with multiple affected individuals have demonstrated that mutations in UPF3B are related to the intellectual disability of these individuals and in many cases some patients present concomitant psychiatric disorders, including schizophrenia or autism [24-27].

Besides the UPF3 mutations discussed above, UPF2 has been also associated with neural and developmental disorders [23]. Particularly, heterozygous deletions of genomic regions that include UPF2 have been linked to clinical phenotypes and a de novo missense mutation was detected in a patient with schizophrenia [28]. An interesting study of copy number variations including 18 known NMD and EJC genes associated these genes with neurological disorders, where UPF2 or RBM8A were associated with various neural disorders, along with congenital abnormalities in some cases [29]. Genomic irregularities in UPF3A, SMG6, EIF4A3 or RNPS1 have also been associated with disorders related to neurological or psychiatric disorders.

Another NMD factor that has been strongly associated to several diseases is Y14, which is a core component of the EJC system that functions in NMD, translational enhancement and mRNA metabolism. In the EJC core, Y14-Magoh directly interacts with eIF4AIII and inhibits its ATPase activity to elicit the interaction between the EJC and the mRNA [30] and this association remains after the export to the cytoplasm [31]. The critical role of Y14 in NMD was confirmed by the observations that depletion of this factor impairs the degradation of transcripts that contain PTCs [32]. Besides its role in NMD regulation, Y14 enhances translation [33, 34], augments PRMT5-mediated methylation of Sm proteins *in vitro* [35] and inhibits mRNA decapping [36]. Additional functions of Y14 include the production of anti-apoptotic forms for the alternatively spliced genes Bcl-x, Bim and Mcl1 [37] and cell cycle control [38]. Considering all the relevant functions of Y14 in mRNA biogenesis and in translation, it is reasonable to presume that it could be related to some pathological conditions. In this regard, Y14 has been implicated in the thrombocytopenia-absent radius (TAR) Syndrome, which occurs when one RBM8A allele that encodes Y14 is deleted and a deleterious regulatory SNP (single-nucleotide polymorphism) is found in the other allele [39, 40]. TAR

patients have low numbers of platelet precursors in bone marrow and mental retardation [41-43]. Thus, Y14 may regulate the expression of hematopoietic cells and in brain development [44].

Even when a clear correlation between an NMD factor and a particular disease has been established only in a few cases, it could be tempting to suggest that there is a more sophisticated interconnection between the NMD machinery and several cell processes related to human disorders.

GENETIC DISEASES LINKED TO NMD

The first evidence indicating that a PTC could be related to a decrease in mRNA translation for a mammalian gene came from studies of β -globin transcripts from patients with β -thalassemia [45, 46]. In normal conditions, hemoglobin is a tetramer integrated by two α -globin and two β -globin subunits. In the common recessive form of β -thalassemia, the human β -globin mRNA contains PTCs at the 5' portion in both copies of the β -globin mRNA, which are targets for NMD. In this situation, free α -globin is degraded and less amount of functional hemoglobin is formed, generating severe anemia in the patient. On the other hand, when PTCs are located in the last exon of the β -globin mRNA, these anomalous forms are able to escape from NMD leading to the production of truncated proteins that result in the symptomatic dominantly inherited anemia [47]. β -thalassemia is one of the few genetic diseases where the effect of NMD has been extensively studied; however it is possible to propose that a similar effect could occur in some other diseases, like susceptibility to mycobacterial infections caused by mutations in the IFNGR1 gene [48], brachydactyly type B and Robinson syndrome that are caused by mutations in the ROR2 gene [49], von Willebrand disease [50], factor X deficiency [51] and retinitis pigmentosa associated to mutations in the CRX gene [52].

Another interesting example of a genetic disease that could be related to NMD is osteogenesis imperfecta (OI). The main cause of OI is sequence variation in the genes COL1A1 and COL1A2, which code for the α 1- and α 2-chain of collagen type 1. Sequence variations at the exon-exon junction could result in exon skipping or to the insertion of PTCs, which could in turn lead to NMD. OI caused by PTCs resulting from single-base substitutions is confined to the COL1A1 gene where a total of 39 unique variants have been recorded, suggesting that COL1A2 PTCs may be generally recessive and under-reported. Mutations in exons 50 and 51 create PTCs through frameshifts in the last exon, which will not result in NMD, creating a dominant negative effect of the disease [53].

RELEVANCE OF NMD IN HUMAN CONDITION

Considering the increasing amount of evidence, it is now clear that NMD is a key regulator of gene expression. The relevance of NMD for cell viability is also supported by the observations concerning lethality of knocking-down key NMD factors in different species [12]. Moreover, mouse embryos that are inactive for NMD resorb shortly after implantation and NMD-inactive blastocysts undergo apoptosis in culture shortly after plated [54].

As mentioned before, mutations in the Upf3b human gene generate intellectual disabilities and neurological conditions, including autism, hyperactivity disorder and schizophrenia. Besides, a large number of diseases including cystic fibrosis, different types of dystrophy and several groups of cancer that are generated because of the presence of PTCs. It was estimated that 12% of all mutations reported in the Human Gene Mutation Database, generated a premature stop codon [55]. If all mutations are considered, including deletions, insertions and splicing mutations, the amount of events that can induce an aberrant product that could be target for NMD increases to almost 30% of all genetic diseases and cancers [11]. Overall, the estimated frequency of PTC inclusion in mRNA molecules due to different causes, strongly suggests that NMD seems to have a huge impact on human health.

Regulation of the activity of several NMD factors can originate differences in NMD accuracy depending on the specific transcript, cell type or tissue in which the regulatory event occurs. This context can in turn show a modulatory effect on the NMD outcome. Generally, NMD protects the cell from aberrant, dysfunctional and potentially toxic peptides that could generate from the mutations depicted above. However, in some cases the truncated proteins produced by PTC-containing mRNAs can still possess, at least in part, its normal function. In these cases, NMD degrades the truncated product that could partially rescue the related phenotype, as demonstrated for Ullrich's disease [56].

Table 1. Most frequently mutated NMD genes in cancer patients

Symbol	Name	Location	Type	# Donors affected	# Mutations
SMG1	SMG1 phosphatidylinositol 3-kinase-related kinase	chr16:18816175-18937776	protein_coding	468 / 3,239 (14.45%)	613
SMG6	SMG6 nonsense mediated mRNA decay factor	chr17:1963133-2207065	protein_coding	464 / 3,239 (14.33%)	787
SMG7	SMG7 nonsense mediated mRNA decay factor	chr1:183441351-183567381	protein_coding	456 / 3,239 (14.08%)	601
RPS9	ribosomal protein S9	chr19:54704610-54752862	protein_coding	434 / 3,239 (13.40%)	506
UPF2	UPF2 regulator of nonsense transcripts homolog (yeast)	chr10:11962021-12085169	protein_coding	382 / 3,239 (11.79%)	486
SMG5	SMG5 nonsense mediated mRNA decay factor	chr1:156219015-156252620	protein_coding	248 / 3,239 (7.66%)	290
RPL37A	ribosomal protein L37a	chr2:217362912-217443903	protein_coding	241 / 3,239 (7.44%)	330
PABPC1	poly(A) binding protein, cytoplasmic 1	chr8:101698044-101735037	protein_coding	233 / 3,239 (7.19%)	241
EIF4G1	eukaryotic translation initiation factor 4 gamma, 1	chr3:184032283-184053146	protein_coding	229 / 3,239 (7.07%)	261
PPP2R1A	protein phosphatase 2, regulatory subunit A, alpha	chr19:52693292-52730687	protein_coding	222 / 3,239 (6.85%)	238

Regarding cancer, it is believed that some carcinomas could be reinforced by NMD abnormalities. Cancer cells commonly show abnormalities in processes like DNA repair, DNA damages, cell cycle progression and apoptosis making reasonable to propose that point and frame-shift mutations can accumulate in tumors, contributing to proliferation and malignancy. This observation could be related to non-essential genes.

Table 2. Most frequent mutations occurring in NMD genes reported for different cancer patients

ID	DNA change	Type	Consequences	# Donors affected
MU80428	chr1:g.6257785T>-	deletion of ≤200bp	Frameshift: RPL22 K15	41 / 8,038 (0.51%)
			5 UTR: RPL22	
			Exon: RPL22	
MU4473544	chr1:g.153963239C>T	single base substitution	5 UTR: RPS27	25 / 8,038 (0.31%)
			Upstream: RAB13 - RP11-422P24.9	
			Exon: RPS27	
			Downstream: NUP210L	
MU612590	chr13:g.115047559C>T	single base substitution	Upstream: UPF3A	25 / 8,038 (0.31%)
			Synonymous: UPF3A L91L	
			Exon: UPF3A	
			Intron: UPF3A	
MU3892625	chr19:g.54754843A>G	single base substitution	Missense: LILRB5 S598P	10 / 8,038 (0.12%)
			Upstream: AC010492.4 - CTD-2337J16.1	
			Exon: LILRB5	
			Downstream: RPS9	
			Intron: LILRB5	
MU4307851	chr1:g.183496786G>-	deletion of ≤200bp	Exon: SMG7	10 / 8,038 (0.12%)
			Downstream: SMG7	
			Intron: SMG7	
MU35402822	chr8:g.101713496CAA>-	deletion of ≤200bp	Downstream: PABPC1	9 / 8,038 (0.11%)
			Intron: PABPC1	
MU1800908	chr19:g.52715971C>G	single base substitution	Missense: PPP2R1A P219R, P179R, P124R	8 / 8,038 (0.10%)
			5 UTR: PPP2R1A	
			Upstream: PPP2R1A	
			Exon: PPP2R1A	
			Downstream: PPP2R1A	
MU715959	chr16:g.18875513T>C	single base substitution	Upstream: SMG1	8 / 8,038 (0.10%)
			Downstream: SMG1	
			Intron: SMG1	
MU64881	chr10:g.11990443T>A	single base substitution	Missense: UPF2 E1033D	7 / 8,038 (0.09%)
MU4555399	chr19:g.54726833T>C	single base substitution	Missense: LILRB3 T6A	7 / 8,038 (0.09%)
			Upstream: LILRB3 - CTB-83J4.1	
			Exon: LILRB3	
			Intron: RPS9 - LILRB3 - LILRA6	

However, NMD could still contribute to cell homeostasis by degrading aberrant products produced from essential genes that would be otherwise toxic [57]. Other possibility is that NMD affect the expression of cancer-associated genes by modulating the quantity of splicing variants produced by these genes. Even when this observation needs to be further demonstrated, there is some evidence suggesting that PTC-containing splicing variants of some tumor-suppressor genes could be targets for NMD, like BRCA1, TP53 and WT1. According to the information deposited in the International Cancer Genome Projects Database (CGPD), the most frequently

mutated NMD genes correspond to different SMG proteins (Table 1). A total of 13,243 mutations have been reported for NMD genes in individuals with cancer and some of the genes with higher number of mutations are UPF3A, SMG7, SMG1 and UPF2 (Table 2).

THERAPIES FOR NMD-RELATED DISEASES

In order to treat diseases caused by premature stop mutations, one possibility is to reduce the efficiency of translation termination to restore the production of some full-length functional proteins. In eukaryotes, translation termination occurs when the stop codon enters the ribosomal A site. Stop codon recognition is mediated by the eukaryotic release factor 1 (eRF1) instead of the codon-anticodon interaction that occurs during translation. Upon this recognition, eRF1 triggers the release of the nascent polypeptide from the peptidyl-tRNA located in the P site of the ribosome [58]. Usually, eRF1 competes with a near-cognate aminoacyl-tRNA, which is able to bind the ribosomal A site even when it contains only two of the three nucleotides of the stop codon. Under different conditions, the incorporation of near-cognate aminoacyl-tRNAs can be stimulated in order to convert stop codons into sense codons [59]. This process is named “termination suppression” or “readthrough”. When this occurs at the PTC that is in frame, the synthesis of the full-length protein continues restoring the proper expression and the event is called “stop codon suppression” or “nonsense suppression”. This natural-occurring process may be exploited to provide a therapeutic tool for patients with genetic disease (Figure 2). Several molecules have been shown to induce “nonsense-suppression” both *in vitro* and *in vivo*, including aminoglycosides (gentamicin, amikacin, tobramycin), modified aminoglycosides (NB30, NB54, NB84), ataluren (PTC124) and RTC13. Demonstrated targets for these molecules comprehend cystic fibrosis (CF), Becker and Duchenne muscular dystrophies (BMD/DMD), spinal muscular atrophy, ataxia telangiectasia, Rett syndrome, Usher syndrome type I, Hurler syndrome, Maroteaux–Lamy syndrome, infantile neuronal ceroid lipofuscinosis, cystinosis, X-linked nephrogenic diabetes insipidus, carnitine palmitoyltransferase 1A, hemophilia, methylmalonic aciduria, neuronal ceroid lipofuscinosis, peroxisome biogenesis disorder (PBD), obesity, poor drug metabolism, and cancer [60]. Aminoglycosides have also shown an effect suppressing nonsense mutations in the tumor suppressor genes p53 and ATM [10]. Some disorders have moved to the stage of clinical trial using suppression therapy, like CF, BMD/DMD, factor VII deficiency, Hailey–Hailey disease, hemophilia A and hemophilia B, leukocyte adhesion deficiency 1 and McArdle disease [61, 62].

Unfortunately, many aminoglycosides used for nonsense suppression possess important risks, like reversible nephrotoxicity and permanent ototoxicity, which have been developed in 2-25% of patients treated with these molecules [62]. Additional efforts have been pursued in order to reduce aminoglycoside toxicity. For example, the concomitant administration of aminoglycosides with various antioxidants or polyanions has shown reduced toxicity [63, 64]. Another successful approach has been to encapsulate aminoglycosides in liposomes [65], but its application for suppression therapy will need more adjustments [66]. Finally, careful monitoring and dose regulation have been also implemented to reduce toxicity, without sacrificing the therapeutic effect [67].

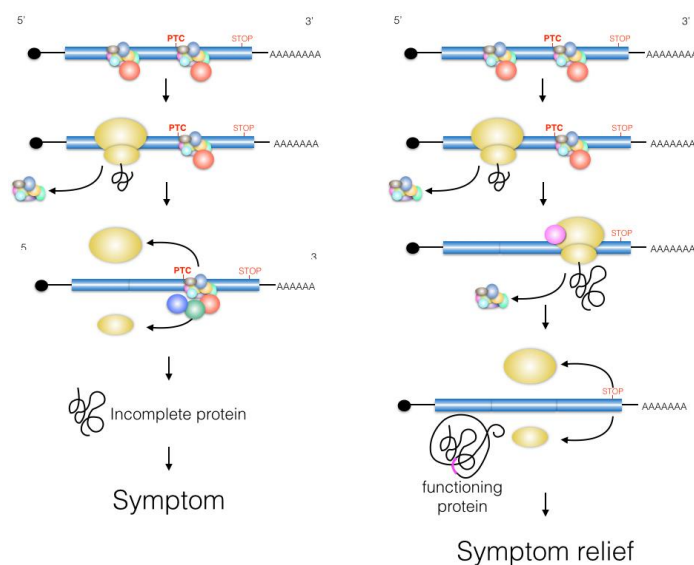


Figure 2. Read through process. Under normal conditions, when a PTC is detected, the NMD machinery is recruited to the mRNA to be degraded. In this case, a truncated version of the protein is produced and a symptom or phenotype can be observed. When an aminoglycoside or a similar molecule (pink circle) interacts with the ribosome, the stop codon can be overlooked and a functional protein can be synthesized, alleviating the symptom caused by the mutation that inserted the PTC.

Table 3. Clinical and preclinical studies evaluating the activity of ataluren

DISORDER OR DISEASE TYPE	DISORDER OR DISEASE NAME	Status	REFERENCE
Muscle Disorders	DMD	Licensed in EU and European Union	74-77
	Miyoshi myopathy	Preclinical study	78
Ion Channel Disease	Cystic Fibrosis	Licensed in EU and European Union	79-86
	Long QT syndrome	Preclinical study	87
Neurological Disorders	Infantile Neuronal Ceroid Lipofuscinoses (INCL)	Preclinical study	88, 89
	Late Infantile Ceroid Lipofuscinoses (LINCL)	Preclinical study	88
	Ataxia telangiectasia	Preclinical study	90
Skin Disease	Usher syndrome (USCH1C)	Preclinical study	91, 92
	Pseudoxanthoma elasticum	Preclinical study	93
	Xeroderma pigmentosum	Preclinical study	94
Eye Disorders	Aniridia	Preclinical study	95
	Retinitis Pigmentosa	Preclinical study	96
Pulmonary Disease	Heritable pulmonary arterial hypertension	Preclinical study	97
Metabolic Disorders	Carnitine palmitoyltransferase 1A Deficiency	Preclinical study	98
	Methylmalonic aciduria (MMA)	Preclinical study	99
	Propionic acidemia (PA)	Preclinical study	100
	Maroteaux-Lamy syndrome (MPS VI)	Preclinical study	101

Table 4. Genes affected by NMD showing important differences in clinical phenotype

Gene symbol	Gene name	PTC location	Phenotype	References
COL1A1	Collagen type I, alpha 1	5' PTC	Osteogenesis imperfecta type I	102, 103
		3' PTC	Osteogenesis imperfect type II-IV	
COL2A1	Collagen type II, alpha 1	5' PTC	Stickler syndrome	102,104
		3' PTC	Spondyloepiphyseal dysplasia	
DMD	Dystrophin	5' PTC	Duchenne muscular dystrophy	105, 106
		3' PTC	Becker muscular dystrophy	
FBN1	Fibrillin-1		Marfan syndrome and type 1 fibrillinopathies	107, 108
		5' PTC	More severe	
SALL1	Sal-like 1	5' PTC	Milder phenotype	109
		3' PTC	Townes–Brocks syndrome	
GHR	Growth hormone receptor	5' PTC	Growth hormone insensitivity syndrome	110
FGD4	Frabin		Charcot Marie Tooth disease type 4H	111, 112
		5' PTC	Less severe	
HEXA	Hexosaminidase A		Tay-Sachs disease	113
		5' PTC	Infantile (severe)	
RB1	Retinoblastoma		Retinoblastoma	114
		5' PTC	Early onset	
ATM	Ataxia-telangiectasia mutated	5' PTC	Mild	115
		3' PTC	Severe	
HBB	beta-Globin gene	5' PTC	Recessively inherited beta-Thalassemia major	116-119
		3' PTC	Dominantly inherited beta-Thalassemia intermedia	
IFNGR1	Interferon gamma receptor 1	5' PTC	Mycobacterial infection recessively inherited susceptibility	120, 48
		3' PTC	Mycobacterial infection dominantly inherited susceptibility	
ROR2	Receptor tyrosine kinase-like orphan receptor 2	5' PTC	Robinow syndrome (recessive)	49, 121
		3' PTC	Brachydactyly type B (dominant)	
VWF	von Willebrand factor	5' PTC	Recessively inherited type 3 von Willebrand disease	50
		3' PTC	Dominantly inherited type 2A von Willebrand disease	
F10	Coagulation factor X	5' PTC	Bleeding tendency associated to factor X deficiency (recessive)	51
		3' PTC	Bleeding tendency associated to factor X deficiency (dominant)	
RHO	Rhodopsin	5' PTC	Retinitis pigmentosa (recessive)	122, 123
		3' PTC	Retinitis pigmentosa (dominant)	
CLCN1	Chloride channel 1	5' PTC	Becker disease	124
		3' PTC	Thomsen disease	
CRX	Cone-rod homeobox-containing gene	5' PTC	No homozygotes to date, normal heterozygotes	52
		3' PTC	Leber congenital amaurosis (dominant)	
ABCC6	ATP-binding cassette, subfamily C member 6	5' PTC	Pseudoxanthoma elasticum (recessive)	125
		3' PTC	Pseudoxanthoma elasticum (dominant)	

In addition to developing alternative strategies to reduce aminoglycosides toxicity, the designing of novel molecules that are safer for nonsense suppression have demonstrated encouraging results. Modified molecules like Ataluren, RT13 and RT14 provide additional structural scaffolds to develop novel, safe and nontoxic drugs with increased efficacy. For example, Ataluren (PTC124) has shown low toxicity and has been approved by the Food and Drug Administration (FDA) and the European Medicines Agency (EMA) for the treatment of DMD and CF originated by nonsense mutations [68]. Ataluren has also been tested in several preclinical studies for different rare disorders [69-73], which are summarized in Table 3.

Even when nonsense suppression has shown promising results, its application as a more general therapy could be limited due to the specific genetic and cellular context. For example, it is possible that the introduction of the random amino acid can generate missense mutation that could still fail to rescue the activity of the full-length protein. In some cases, it has been reported that a full- or nearly full-length protein generated by nonsense suppression can activate a T-cell mediated immune response.

CONCLUSION

NMD relevance for human health is starting to be discovered. Several rare disorders have been linked to NMD (Table 4). Moreover, even when a PTC could be located in any part of the mRNA, it appears that a correlation exists between the PTC position and the strength of the disease. In many cases, so-called silent mutations correlated with a particular phenotype and after NMD discovery, it results clear that many of these point mutations inserted a PTC, generating in many cases a severe disease. With all the information available and summarized here, it results clear that NMD is a relevant mechanism for human health. However, an important amount of diseases that may be targets for NMD remain to be studied. Fortunately, some therapeutic tools have been developed to treat some of the NMD-related diseases. Even when the general strategy for modulating NMD has been applied with success, an enormous effort will be necessary in order to discover the molecular details underlying several genetic diseases and the therapeutic tools that could be applied to correct the defects.

REFERENCES

- [1] Chang JC, Kan YW. 1979. Beta-thalassemia, a nonsense mutation in man. *Proc. Natl. Acad. Sci. U. S. A.* 76: 2886–9.
- [2] Losson R, Lacroute F. 1979. Interference of nonsense mutations with eukaryotic messenger RNA stability. *Proc. Natl. Acad. Sci. U. S. A.* 76: 5134–7.
- [3] Pulak R, Anderson P. 1993. mRNA surveillance by the *Caenorhabditis elegans* smg genes. *Genes Dev.* 7: 1885–97.
- [4] Brogna S. 1999. Nonsense mutations in the alcohol dehydrogenase gene of *Drosophila melanogaster* correlate with an abnormal 3' end processing of the corresponding pre-mRNA, *RNA* 5: 562–73.
- [5] van Hoof A, Green PJ. 1996. Premature nonsense codons decrease the stability of phytohemagglutinin mRNA in a position-dependent manner, *Plant J.* 10: 415–24.

- [6] Yamashita A. 2013. Role of SMG-1-mediated Upf1 phosphorylation in mammalian nonsense-mediated mRNA decay. *Genes to Cells* 18: 161–175.
- [7] Green RE, Lewis BP, Hillman RT, Blanchette M, et al. 2003. Widespread predicted nonsense-mediated mRNA decay of alternatively-spliced transcripts of human normal and disease genes. *Bioinformatics* 19(S1): i118–i121.
- [8] Pan Q, Saltzman AL, Kim YK, Misquitta C, et al. 2006 Quantitative microarray profiling provides evidence against widespread coupling of alternative splicing with nonsense mediated mRNA decay to control gene expression. *Genes Dev.* 20, 153–8.
- [9] Weischenfeldt J, Waage J, Tian G, Zhao J, et al. 2012. Mammalian tissues defective in nonsense-mediated mRNA decay display highly aberrant splicing patterns. *Genome Biol.* 13: R35.
- [10] Keeling KM, Bedwell DM. 2011. Suppression of nonsense mutations as a therapeutic approach to treat genetic diseases, Wiley Interdiscip. Rev. *RNA* 2: 837–52.
- [11] Frischmeyer PA, Dietz HC. 1999. Nonsense-mediated mRNA decay in health and disease. *Hum. Mol. Genet.* 8: 1893–900.
- [12] Culbertson MR, Leeds PF. 2003. Looking at mRNA decay pathways through the window of molecular evolution. *Curr. Opin. Genet. Dev.* 13: 207–14.
- [13] Lelivelt M, Culbertson MR. 1999. Yeast Upf Proteins required for RNA surveillance affect global expression of the yeast transcriptome. *Mol. Cell Biol* 19: 6710–9.
- [14] Serin G, Gersappe A, Black JD, Aronoff R, et al. 2001. Identification and characterization of human orthologues to *Saccharomyces cerevisiae* Upf2 protein and Upf3 protein (*Caenorhabditis elegans* SMG-4). *Mol. Cell Biol.* 2001 21: 209–23.
- [15] Wang W, Czaplinski K, Rao Y, Peltz SW. 2001. The role of Upf proteins in modulating the translation read-through of nonsense-containing transcripts. *EMBO J.* 20: 880–90.
- [16] Hodgkin J, Papp A, Pulak R, Ambros V, et al. 1989. A new kind of informational suppression in the nematode *Caenorhabditis elegans*. *Genetics* 123: 301–13.
- [17] Nicholson P, Yepiskoposyan H, Metze S, Zamudio-Orozco R, et al. 2010. Nonsense-mediated mRNA decay in human cells: mechanistic insights, functions beyond quality control and the double-life of NMD factors. *Cell. Mol. Life Sci.* 67: 677–700.
- [18] Conti E, Izaurralde E. 2005. Nonsense-mediated mRNA decay: molecular insights and mechanistic variations across species. *Curr. Opin. Cell Biol.* 17: 316–25.
- [19] Lejeune F, Maquat LE. 2005. Mechanistic links between nonsense-mediated mRNA decay and pre-mRNA splicing in mammalian cells. *Curr. Opin. Cell Biol.* 17: 309–15.
- [20] Le Hir H, Izaurralde E, Maquat LE, Moore MJ. 2000. The spliceosome deposits multiple proteins 20–24 nucleotides upstream of mRNA exon–exon junctions. *EMBO J.* 19: 6860–9.
- [21] Lejeune F, Ishigaki Y, Li X, Maquat LE. 2002. The exon junction complex is detected on CBP80-bound but not eIF4E-bound mRNA in mammalian cells: dynamics of mRNP remodeling. *EMBO J.* 21: 3536–45.
- [22] Singh G, Kucukural A, Cenik C, Leszyk JD, et al. 2012. The cellular EJC interactome reveals higher-order mRNP structure and anEJC-SR protein nexus. *Cell* 151: 750–64.
- [23] Nguyen LS, Wilkinson MF, Gecz J. 2014. Nonsense-mediated mRNA decay: Inter-individual variability and human disease. *Neurosci. Biobehav. Rev.* 46: 175–86.
- [24] Addington AM, Gauthier J, Piton A, Hamdan, et al. 2011. A novel frameshift mutation in UPF3B identified in brothers affected with childhood onset schizophrenia and autism spectrum disorders. *Mol. Psychiatry* 16: 238–9.

- [25] Laumonnier F, Shoubbridge C, Antar C, Nguyen LS, et al. 2010. Mutations of the UPF3B gene, which encodes a protein widely expressed in neurons, are associated with nonspecific mental retardation with or without autism. *Mol. Psychiatry* 15: 767–76.
- [26] Lynch SA, Nguyen LS, Ng LY, Waldron M, et al. 2012. Broadening the phenotype associated with mutations in UPF3B: two further cases with renal dysplasia and variable developmental delay. *Eur. J. Med. Genet.* 55: 476–9.
- [27] Tarpey PS, Nguyen LS, Raymond FL, Rodriguez J, et al. 2007. Mutations in UPF3B, a member of the nonsense-mediated mRNA decay complex, cause syndromic and nonsyndromic mental retardation. *Nat. Genet.* 39:1127–1133.
- [28] Gulsuner S, Walsh T, Watts AC, Lee MK, et al. 2013. Spatial and temporal mapping of de novo mutations in Schizophrenia to a fetal prefrontal cortical network. *Cell* 154: 518–29.
- [29] Nguyen LS, Kim HG, Rosenfeld JA, Shen Y, et al. 2013. Contribution of copy number variants involving nonsense-mediated mRNA decay pathway genes to neurodevelopmental disorders. *Hum. Mol. Genet.* 22: 1816–1825.
- [30] Ballut L, Marchadier B, Baguet A, Tomasetto C, et al. 2005. The exon junction complex is locked onto RNA by inhibition of EIF4AIII ATPase activity. *Nat. Struct. Mol. Biol.* 12: 861–9.
- [31] Tange TO, Shibuya T, Jurica MS, Moore MJ. 2005. Biochemical analysis of the EJC reveals two new factors and a stable tetrameric protein core. *RNA* 11: 1869–75.
- [32] Gehring NH, Neu-Yilik G, Schell T, Hentze MW, et al. 2003. Y14 and hUpf3b form an NMD-activating complex. *Mol. Cell* 11: 939–49.
- [33] Diem MD, Chan CC, Younis I, Dreyfuss G. 2007. PYM binds the cytoplasmic exon-junction complex and ribosomes to enhance translation of spliced mRNAs. *Nat. Struct. Mol. Biol.* 14: 1173–9.
- [34] Lee HC, Choe J, Chi SG, Kim YK. 2009. Exon junction complex enhances translation of spliced mRNAs at multiple steps. *Biochem. Biophys. Res. Commun.* 384: 334–40.
- [35] Chuang TW, Peng PJ, Tarn WY. 2011. The exon junction complex component Y14 modulates the activity of the methylosome in biogenesis of spliceosomal small nuclear ribonucleoproteins. *J. Biol. Chem.* 286: 8722–8.
- [36] Chuang TW, Chang WL, Lee KM, Tarn WY. 2013. The RNA-binding protein Y14 inhibits mRNA decapping and modulates processing body formation. *Mol. Biol. Cell* 24: 1–13.
- [37] Michelle L, Cloutier A, Toutant J, Shkreta L, et al. 2012. Proteins associated with the exon junction complex also control the alternative splicing of apoptotic regulators. *Mol. Cell. Biol.* 32: 954–67.
- [38] Rehwinkel J, Letunic I, Raes J, Bork P, et al. 2005. Nonsense-mediated mRNA decay factors act in concert to regulate common mRNA targets. *RNA* 11: 1530–44.
- [39] Albers CA, Paul DS, Schulze H, Freson K, et al. 2012. Compound inheritance of a low-frequency regulatory snp and a rare null mutation in exon-junction complex subunit RBM8A causes tar syndrome. *Nat. Genet.* 44: 435–9.
- [40] Klopocki E, Schulze H, Strauss G, Ott CE, et al. 2007. Complex inheritance pattern resembling autosomal recessive inheritance involving a microdeletion in thrombocytopenia-absent radius syndrome. *Am. J. Hum. Genet.* 80: 232–40.

- [41] Silver DL, Watkins-Chow DE, Schreck KC, Pierfelice TJ, et al. 2010. The exon junction complex component Magoh controls brain size by regulating neural stem cell division. *Nat. Neurosci.* 13: 551–558.
- [42] Brunetti-Pierri N, Berg JS, Scaglia F, Belmont J, et al. 2008. Recurrent reciprocal 1q21.1 deletions and duplications associated with microcephaly or macrocephaly and developmental and behavioral abnormalities. *Nat. Genet.* 40: 1466–71.
- [43] Skorka A, Bielicka-Cymermann J, Gieruszczak-Bialek D, Korniszewski L. 2005. Thrombocytopenia-absent radius (TAR) syndrome: A case with agenesis of corpus callosum, hypoplasia of cerebellar vermis and horseshoe kidney. *Genet. Couns.* 16: 377–82.
- [44] Chuang TW, Lee KM, Tarn WY. 2015. Function and Pathological Implications of Exon Junction Complex Factor Y14. *Biomolecules* 5: 343-55.
- [45] Cui Y, Hagan KW, Zhang S, Peltz SW. 1995. Identification and characterization of genes that are required for the accelerated degradation of mRNAs containing a premature translational termination codon. *Genes Dev.* 9: 423-36.
- [46] He F, Brown AH, Jacobson A. 1997. Upf1p, Nmd2p, and Upf3p are interacting components of the yeast nonsense-mediated mRNA decay pathway. *Mol. Cell Biol.* 173: 1580-94.
- [47] Danckwardt S, Neu-Yilik G, Thermann R, Frede U, et al. 2002. Abnormally spliced beta-globin mRNAs: a single point mutation generates transcripts sensitive and insensitive to nonsense-mediated mRNA decay. *Blood* 99: 1811-6.
- [48] Jouanguy E, Lamhamedi-Cherradi S, Lammass D, Dorman SE, et al. 1999. A human IFNGR1 small deletion hotspot associated with dominant susceptibility to mycobacterial infection. *Nat. Genet.* 21: 370-8.
- [49] Schwabe GC, Tinschert S, Buschow C, Meinecke P, et al. 2000. Distinct mutations in the receptor tyrosine kinase gene ROR2 cause brachydactyly type B. *Am. J. Hum. Genet.* 67: 822-31.
- [50] Schneppenheim R, Budde U, Obser T, Brassard J, et al. 2001. Expression and characterization of von Willebrand factor dimerization defects in different types of von Willebrand disease. *Blood.* 97: 2059-66.
- [51] Millar DS, Kemball-Cook G, McVey JH, Tuddenham EG, et al. 2000. Molecular analysis of the genotype-phenotype relationship in factor VII deficiency. *Hum. Genet.* 107: 327-42.
- [52] Rivolta C, Berson EL, Dryja TP. 2001. Dominant Leber congenital amaurosis, cone-rod degeneration, and retinitis pigmentosa caused by mutant versions of the transcription factor CRX. *Hum. Mutat.* 18: 488-98.
- [53] Dalgleish R. 2014. *Osteogenesis Imperfecta Genotypes and Genotype-Phenotype Relationships*. Elsevier. DOI: <http://dx.doi.org/10.1016/B978-0-12-397165-4.00010-1>.
- [54] Medghalchi SM, Frischmeyer PA, Mendell JT, Kelly AG, et al. 2001. Rent1, a trans-effector of nonsense-mediated mRNA decay, is essential for mammalian embryonic viability. *Hum. Mol. Genet.* 10: 99-105.
- [55] Krawczak M, Ball EV, Fenton I, Stenson PD, et al. 2000. Human gene mutation database-a biomedical information and research resource. *Hum. Mutat.* 15: 45-51.
- [56] Usuki F, Yamashita A, Higuchi I, Ohnishi T, et al. 2004. Inhibition of nonsense-mediated mRNA decay rescues the phenotype in Ullrich's disease. *Ann. Neurol.* 55: 740-4.

- [57] Scholzová E, Malík R, Sevcík J, Kleibl Z. 2007. RNA regulation and cancer development. *Cancer Lett.* 246: 12-23.
- [58] Kisselev L, Ehrenberg M, Frolova L. 2003. Termination of translation: interplay of mRNA, rRNAs and release factors? *EMBO J.* 22: 175-82.
- [59] Keeling KM, Bedwell DM. 2011. Suppression of nonsense mutations as a therapeutic approach to treat genetic diseases, Wiley Interdiscip. Rev. *RNA* 2: 837-52.
- [60] Miller JN, Pearce DA. 2014. Nonsense-mediated decay in genetic disease: Friend or foe? *Mutation Research* 762: 52-64.
- [61] Lee HL, Dougherty JP. 2012. Pharmaceutical therapies to recode nonsense mutations in inherited diseases. *Pharmacol. Ther.* 136: 227-266.
- [62] Keeling KM, Wang D, Conard SE, Bedwel DM. 2012. Suppression of premature termination codons as a therapeutic approach. *Crit. Rev. Biochem. Mol. Biol.* 47: 444-63.
- [63] Reiter RJ, Tan DX, Korkmaz A, Fuentes-Broto L. 2011. Drug-mediated ototoxicity and tinnitus: Alleviation with melatonin. *J. Physiol. Pharmacol.* 62: 151-7.
- [64] Thibault N, Grenier L, Simard M, Bergeron MG, et al. 1995. Protection against gentamicin nephrotoxicity by daptomycin in nephrectomized rats. *Life Sci.* 56: 1877-87.
- [65] Meers P, Neville M, Malinin V, Scotto AW, et al. 2008. Biofilm penetration, triggered release and in vivo activity of inhaled liposomal amikacin in chronic *Pseudomonas aeruginosa* lung infections. *J. Antimicrob. Chemother* 61: 859-68.
- [66] Yukihiro M, Ito K, Tanoue O, Goto K, et al. 2011. Effective drug delivery system for duchenne muscular dystrophy using hybrid liposomes including gentamicin along with reduced toxicity. *Biol. Pharm. Bull* 34:712-6.
- [67] Smyth AR, Bhatt J. 2012. Once-daily versus multiple-daily dosing with intravenous aminoglycosides for cystic fibrosis. *Cochrane Database Syst. Rev* 2: CD002009.
- [68] Peltz SW, Morsy M, Welch EM, Jacobson A. 2013. Ataluren as an agent for therapeutic nonsense suppression. *Annu. Rev. Med.* 64: 407-25.
- [69] Wang B, Yang Z, Brisson BK, Feng H, et al. 1985. Membrane blebbing as an assessment of functional rescue of dysferlin-deficient human myotubes via nonsense suppression. *J. Appl. Physiol.* 109: 901-5.
- [70] Yu H, Liu X, Huang J, Zhang Y, et al. 2014. Comparison of read-through effects of aminoglycosides and PTC124 on rescuing nonsense mutations of HERG gene associated with long QT syndrome. *Int. J. Mol. Med.* 33: 729-35.
- [71] Miller JN, Chan CH, Pearce DA. 2013. The role of nonsense-mediated decay in neuronal ceroid lipofuscinosis. *Hum. Mol. Genet.* 22: 2723-34.
- [72] Du L, Jung ME, Damoiseaux R, Completo G, et al. 2013. A new series of small molecular weight compounds induce read through of all three types of nonsense mutations in the ATM gene. *Mol. Ther.* 21: 1653-60.
- [73] Goldmann T, Overlack N, Möller F, Belakhov V, et al. 2012. A comparative evaluation of NB30, NB54 and PTC124 in translational read-through efficacy for treatment of an USH1C nonsense mutation. *EMBO Mol. Med.* 4: 1186-99.
- [74] Li M, Andersson-Lendahl M, Sejersen T, Arner A. 2014. Muscle dysfunction and structural defects of dystrophin-null sapje mutant zebrafish larvae are rescued by ataluren treatment. *FASEB J.* 28: 1593-9.
- [75] Bushby K, Finkel R, Wong B, Barohn R, et al. 2014. Ataluren treatment of patients with nonsense mutation dystrophinopathy. *Muscle Nerve.* 50: 477-87.

- [76] Finkel RS, Flanigan KM, Wong B, Bönnemann C, et al. 2013. Phase 2a study of ataluren-mediated dystrophin production in patients with nonsense mutation Duchenne muscular dystrophy. *PLoS One* 8: e81302.
- [77] Kayali R, Ku JM, Khitrov G, Jung ME, et al. 2012. Read-through compound 13 restores dystrophin expression and improves muscle function in the mdx mouse model for Duchenne muscular dystrophy. *Hum. Mol. Genet.* 21: 4007-20.
- [78] Wang B, Yang Z, Brisson BK, Feng H, et al. 1985. Membrane blebbing as an assessment of functional rescue of dysferlin-deficient human myotubes via nonsense suppression. *J. Appl. Physiol.* 109: 901-5.
- [79] Kerem E, Konstan MW, De Boeck K, Accurso FJ, et al. (2014) Ataluren for the treatment of nonsense-mutation cystic fibrosis: a randomised, double-blind, placebo-controlled phase 3 trial. *Lancet Respir Med.* 2: 539-47.
- [80] Johansson J, Vezzalini M, Verzè G, Caldrier S, et al. 2014. Detection of CFTR protein in human leukocytes by flow cytometry. *Cytometry A.* 85: 611-20.
- [81] Lentini L, Melfi R, Di Leonardo A, Spinello A, et al. 2014. Toward a rationale for the PTC124 (Ataluren) promoted readthrough of premature stop codons: a computational approach and GFP-reporter cell-based assay. *Mol Pharm.* 11: 653-64.
- [82] Gonzalez-Hilarion S, Beghyn T, Jia J, Debreuck N, et al. 2012. Rescue of nonsense mutations by amlexanox in human cells. *Orphanet J. Rare Dis.* 7: 58.
- [83] Wilschanski M, Miller LL, Shoseyov D, Blau H, et al. 2011. Chronic ataluren (PTC124) treatment of nonsense mutation cystic fibrosis. *Eur Respir J.* 38: 59-69.
- [84] Sermet-Gaudelus I., Boeck KD, Casimir GJ, Vermeulen F, et al. 2010. Ataluren (PTC124) induces cystic fibrosis transmembrane conductance regulator protein expression and activity in children with nonsense mutation cystic fibrosis *Am. J. Respir Crit. Care Med.* 182: 1262-72.
- [85] Kerem E, Hirawat S, Armoni S, Yaakov Y, et al. 2008. Effectiveness of PTC124 treatment of cystic fibrosis caused by nonsense mutations: a prospective phase II trial. *The Lancet.* 372: 719-27.
- [86] Du M, Liu X, Welch EM, Hirawat S, et al. 2008. PTC124 is an orally bioavailable compound that promotes suppression of the human CFTR-G542X nonsense allele in a CF mouse model. *Proc. Natl. Acad. Sci. U. S. A.* 105: 2064-9.
- [87] Yu H, Liu X, Huang J, Zhang Y, et al. 2014. Comparison of read-through effects of aminoglycosides and PTC124 on rescuing nonsense mutations of HERG gene associated with long QT syndrome. *Int. J. Mol. Med.* 33: 729-35.
- [88] Miller JN, Chan CH, Pearce DA. 2013. The role of nonsense-mediated decay in neuronal ceroid lipofuscinosis. *Hum. Mol. Genet.* 22: 2723-34.
- [89] Sarkar C, Zhang Z, Mukherjee AB. 2011. Stop codon read-through with PTC124 induces palmitoyl-protein thioesterase-1 activity, reduces thioester load and suppresses apoptosis in cultured cells from INCL patients. *Mol. Genet. Metab.* 104: 338-45.
- [90] Du L, Jung ME, Damoiseaux R, Completo G, et al. 2013. A new series of small molecular weight compounds induce read through of all three types of nonsense mutations in the ATM gene. *Mol. Ther.* 21: 1653-60.
- [91] Goldmann T, Overlack N, Möller F, Belakhov V, et al. 2012. A comparative evaluation of NB30, NB54 and PTC124 in translational read-through efficacy for treatment of an USH1C nonsense mutation. *EMBO Mol. Med.* 4: 1186-99.

- [92] Goldmann T, Overlack N, Wolfrum U, Nagel-Wolfrum K. 2011. PTC124-mediated translational readthrough of a nonsense mutation causing Usher syndrome type 1C. *Hum. Gene Ther.* 22: 537-47.
- [93] Zhou Y, Jiang Q, Takahagi S, Shao C, et al. 2013. Premature termination codon read-through in the ABCC6 gene: potential treatment for pseudoxanthoma elasticum. *J. Invest Dermatol.* 133: 2672-7.
- [94] Kuschal C, DiGiovanna JJ, Khan SG, Gatti RA, et al. 2013. Repair of UV photolesions in xeroderma pigmentosum group C cells induced by translational readthrough of premature termination codons. *Proc. Natl. Acad. Sci. U. S. A.* 110: 19483-8.
- [95] Gregory-Evans CY, Wang X, Wasan KM, Zhao J, et al. 2014. Postnatal manipulation of Pax6 dosage reverses congenital tissue malformation defects. *J. Clin. Invest.* 124: 111-6.
- [96] Schwartz N. 2014. Translational read-through of the RP2 Arg120stop mutation in patient iPSC-derived retinal pigment epithelium cells. *Hum. Mol. Genet.* 24: 972-86.
- [97] Drake KM, Dunmore BJ, McNelly LN, Morrell NW, et al. 2013. Correction of nonsense BMPR2 and SMAD9 mutations by ataluren in pulmonary arterial hypertension. *Am. J. Respir. Cell Mol. Biol.* 49: 403-9.
- [98] Tan L, Narayan SB, Chen J, Meyers GD, et al. 2011. PTC124 improves readthrough and increases enzymatic activity of the CPT1A R160X nonsense mutation. *J. Inherit. Metab Dis.* 34: 443-7.
- [99] Buck NE, Wood LR, Hamilton NJ, Bennett MJ, et al. 2012. Treatment of a methylmalonyl-CoA mutase stop codon mutation. *Biochem. Biophys. Res. Commun.* 427: 753-7.
- [100] Sánchez-Alcudia R, Pérez B, Ugarte M, Desviat LR. 2012. Feasibility of nonsense mutation readthrough as a novel therapeutical approach in propionic acidemia. *Hum. Mutat.* 33: 973-80.
- [101] Bartolomeo R, Polishchuk EV, Volpi N, Polishchuk RS, et al. 2013. Pharmacological read-through of nonsense ARSB mutations as a potential therapeutic approach for mucopolysaccharidosis VI. *J. Inherit. Metab. Dis.* 36: 363-71
- [102] Körkkö J, Ala-Kokko L, De Paepe A, Nuytinck L, et al. 1998. Analysis of the COL1A1 and COL1A2 genes by PCR amplification and scanning by conformation-sensitive gel electrophoresis identifies only COL1A1 mutations in 15 patients with osteogenesis imperfecta type I: identification of common sequences of null-allele mutations. *Am. J. Hum. Genet.* 62 : 98–110.
- [103] Willing MC, Deschenes SP, Slation RL, Roberts EJ. 1996. Premature chain termination is a unifying mechanism for COL1A1 null alleles in osteogenesis imperfecta type I cell strains. *Am. J. Hum. Genet.* 59: 799–809.
- [104] Snead MP, Yates JR. 1999. Clinical and molecular genetics of Stickler syndrome. *J. Med. Genet.* 36: 353–9.
- [105] Kerr TP, Sewry CA, Robb SA, Roberts RG. 2001. Long mutant dystrophins and variable phenotypes: evasion of nonsense-mediated decay? *Hum. Genet.* 109: 402–7.
- [106] Pillers DA, Fitzgerald KM, Duncan NM, Rash SM, et al. (1999) Duchenne/Becker muscular dystrophy: correlation of phenotype by electroretinography with sites of dystrophin mutations. *Hum. Genet.* 105: 2–9.

- [107] Schrijver I, Liu W, Odom R, Brenn T, et al. 2002. Premature termination mutations in FBN1: distinct effects on differential allelic expression and on protein and clinical phenotypes. *Am. J. Hum. Genet.* 71: 223–37.
- [108] Faivre L, Collod-Beroud G, Loeys BL, Child A, et al. 2007. Effect of mutation type and location on clinical outcome in 1,013 probands with Marfan syndrome or related phenotypes and FBN1 mutations: an international study. *Am. J. Hum. Genet.* 81: 454–66.
- [109] Furniss D, Critchley P, Giele H, Wilkie AO. 2007. Nonsense-mediated decay and the molecular pathogenesis of mutations in SALL1 and GLI3. *Am. J. Med. Genet. A* 143A: 3150–60.
- [110] Gorbenko del Blanco D, de Graaf LC, Visser TJ, Hokken-Koelega AC. (2012) Growth hormone insensitivity syndrome caused by a heterozygous GHR mutation: phenotypic variability owing to moderation by nonsense-mediated decay. *Clin. Endocrinol. (Oxf)* 76: 706–12.
- [111] Houlden H, Hamans S, Katifi H, Reilly MM. 2009. A novel Frabin (FGD4) nonsense mutation p.R275X associated with phenotypic variability in CMT4H. *Neurology* 72: 617–20.
- [112] Stendel C, Roos A, Deconinck T, Pereira J, et al., 2007. Peripheral nerve demyelination caused by a mutant Rho GTPase guanine nucleotide exchange factor, frabin/FGD4. *Am. J. Hum. Genet.* 81:158–64.
- [113] R. Myerowitz, 1997. Tay-Sachs disease-causing mutations and neutral polymorphisms in the Hex A gene. *Hum. Mutat.* 9: 195–208.
- [114] Alonso J, Garcia-Miguel P, Abelairas J, Mendiola M, et al. 2001. Spectrum of germline RB1 gene mutations in Spanish retinoblastoma patients: phenotypic and molecular epidemiological implications. *Hum. Mutat.* 17: 412–22.
- [115] Li A, Swift M. 2000. Mutations at the ataxia-telangiectasia locus and clinical phenotypes of A-T patients. *Am. J. Med. Genet.* 92: 170–7.
- [116] Hall GW, S. Thein Nonsense codon mutations in the terminal exon of the beta-globin gene are not associated with a reduction in beta-mRNA accumulation: a mechanism for the phenotype of dominant beta-thalassemia, *Blood* 83 (1994) 2031–2037.
- [117] Thein SL, Hesketh C, Taylor P, Temperley IJ, et al. 1990. Molecular basis for dominantly inherited inclusion body beta- thalassemia, *Proc. Natl. Acad. Sci. U. S. A.* 87: 3924–8.
- [118] Baserga SJ, Benz EJ Jr. 1988. Nonsense mutations in the human beta-globin gene affect mRNA metabolism, *Proc. Natl. Acad. Sci. U. S. A.* 85: 2056–60.
- [119] Forget BG, Benz EJ Jr, Sloultchi A, Baglioni C, et al. 1974. Absence of messenger RNA for beta globin chain in beta(0) thalassaemia, *Nature* 247: 379–81.
- [120] Jouanguy E, Altare F, Lamhamedi S, Revy P, et al. (1996). Interferon-gamma-receptor deficiency in an infant with fatal bacille Calmette–Guerin infection, *N. Engl. J. Med.* 335:1956–1961.
- [121] Patton MA, Afzal AR. 2002. Robinow syndrome. *J. Med. Genet.* 39: 305–10.
- [122] Rosenfeld PJ, Cowley GS, McGee TL, Sandberg MA, et al. 1992. A null mutation in the rhodopsin gene causes rod photoreceptor dysfunction and autosomal recessive retinitis pigmentosa, *Nat. Genet.* 1: 209–13.

- [123] Sung CH, Davenport CM, Nathans J. 1991. Rhodopsin mutations in autosomal dominant retinitis pigmentosa, *Proc. Natl. Acad. Sci. U. S. A.* 88: 6481–5.
- [124] Pusch M. 2002. Myotonia caused by mutations in the muscle chloride channel gene CLCN1, *Hum. Mutat.* 19: 423–34.
- [125] Plomp AS, Hu X, de Jong PT, Bergen AA. 2004. Does autosomal dominant pseudoxanthoma elasticum exist? *Am. J. Med. Genet. A* 126A: 403–12.

Chapter 19

ALTERNATIVE SPLICING ALTERATIONS IN ALZHEIMER'S DISEASE

T. A. Ishunina^{1,*} and D. F. Swaab²

¹Department of Histology, Embryology, Cytology, Kursk State Medical University,
Kursk, Russia

²Netherlands Institute for Neuroscience, an Institute of the Royal Netherlands Academy
of Arts and Sciences, Amsterdam, The Netherlands

ABSTRACT

Alzheimer's disease (AD) is the most common type of dementia in the elderly population with a higher prevalence in women. Memory impairment and cognitive decline in AD are primarily linked to its neuropathological hallmarks, i.e., cholinergic neuronal atrophy and death, presence of intraneuronal neurofibrillary tangles and accumulation of amyloid β ($A\beta$) deposits within the extracellular senile plaques. Consequently, genes encoding $A\beta$, tau and proteins involved in $A\beta$ procession received the most careful scientific attention. However, a growing body of evidence suggests that AD pathogenesis is not limited to the changes of AD genes' expression, but also depends on their alternative splicing. Therefore, the present review focuses on data concerning alternative splicing changes in AD.

First of all, pivotal AD genes (APP (amyloid precursor protein), tau, presenilin 1 (PS-1) and 2 (PS-2) and apolipoprotein E (APOE)) have a number of splice variants with divergent functions that are differentially expressed in AD and normal brain tissues. Second, alternative splicing of genes involved in $A\beta$ processing and metabolism is also affected in AD. This group includes BACE-1 (β -site APP cleaving enzyme 1), nicastrin and APL-1 which are components of the γ -secretase complex, AIDA-1 (protein that binds to the intracellular domain of $A\beta$ PP following cleavage by γ -secretase; FE65 (binds to the cytoplasmic tail of β PP). Finally, splice variants of such candidate AD genes as acetylcholinesterase (cholinergic deficit is one of the central mechanisms in AD development), estrogen receptor α (ER α) (estrogen deficiency is one of the predisposing to

* Corresponding Author's Email: T.Ishunina@nin.knaw.nl (Ishunina T.A., MD, PhD, Associate Professor, Department of Histology, Embryology, Cytology, Kursk State Medical University, Karl Marx Street 3, 305041 Kursk, Russia).

AD factors), BDNF (brain derived neurotrophic factor that is crucial for neuronal survival and synaptic transmission) and its receptor TrkB, excitatory aminoacid transporter 2 (EAAT2) (colocalized with tau in dystrophic neurons), genes of the ion channels and neurotransmitter receptors, synapsin, RCAN-1 (regulator of calcineurin), neuronal GFAP, ubiquilin-1 (related to the proteasomal degradation of proteins and interaction with PS-1 and PS-2) and genes involved in the regulation of the cell cycle and apoptosis (CIZ1, DENN/MADD) should be considered as important elements of the AD pathogenesis.

Together, these data show that a lot of changes in alternative splicing are intimately linked to AD, which should be, thus, considered as a disorder with compromised and disregulated splicing events.

INTRODUCTION

Alzheimer's disease (AD) is the most common cause of dementia with a higher prevalence in women. Cognitive decline in AD was found to be correlated to hyperphosphorylated tau (neurofibrillary tangles, (NFT)) and A β deposits (senile plaques) (Duyckaerts et al., 2009). Therefore, genetic studies focused on genes encoding tau, A β and proteins involved in A β procession. A growing body of evidence suggests that alternative splicing of these genes is closely related to AD pathogenesis, and, in particular, to sporadic or spontaneous cases (Mills and Janitz, 2012). In the present review splice variants of following AD-related genes will be presented: 1) pivotal genes associated with inherited predisposition to AD (tau, APP (amyloid precursor protein), presenilin 1 (PS-1) and 2 (PS-2) and apolipoprotein E (APOE); 2) genes involved in A β processing and metabolism: BACE-1 (β -site APP cleaving enzyme 1), nicastrin, APH-1 (anterior pharynx defective), AIDA-1 (AID-associated protein 1), clusterin, FE65; 3) other candidate AD genes: acetylcholinesterase, tissue transglutaminase (tTG), estrogen receptor α (ER α), BDNF (brain derived neurotrophic factor), TrkB (tropomyosin receptor kinase B), excitatory aminoacid transporter 2 (EAAT2), genes of ion channels and neurotransmitter receptors, synapsin, synphilin-1, RCAN-1 (regulator of calcineurin), neuronal GFAP (glial fibrillar acidic protein), ubiquilin-1, CIZ1 (CDKN1A interacting zinc finger protein 1), DENN/MADD (differentially expressed in normal and neoplastic tissues MAPK-activating, death domain protein), DBI (diazepam-binding inhibitor).

Alternative Splicing of Pivotal AD Genes

Tau

Tau protein is crucial for microtubule assembly in neurons (Uversky, 2009). In AD tau is abnormally phosphorylated, N-truncated and accumulated in the neurons in the form of neurofibrillary tangles (Duyckaerts et al., 2009). The distribution of tau deposition in the brain progresses in a stepwise way from the entorhinal cortex and hippocampus to isocortex and the visual cortex as the last affected areas, providing the basis for Braak staging of AD (Braak and Braak, 1991). The tau gene contains 16 exons, from which exons 2, 3 and 10 are alternatively spliced. In the human brain 6 tau isoforms were identified that are different in the number of 29 amino acids inserts at the N-terminal part (encoded by exons 2 and 3) and three (3R tau, exon 10 missing) or four repeat (4R tau, exon 10 is present) regions at the C-terminus: Δ 2,3,10 (exons 2,3 and 10 deleted); Δ 3,10 (exons 3 and 10 deleted, exon 2 present); Δ 10 (exon 10 is

skipped, exons 2 and 3 present); $\Delta 2,3$ (exons 2 and 3 missing, exon 10 present); $\Delta 3$ (exon 3 is skipped, exons 2 and 10 present); isoform with exons 2,3 and 10 present) (Goedert et al., 1989; Buee et al., 2000; Wang and Liu, 2008; Wolfe, 2009; Uversky, 2009). Isoforms without exon 10 are known as 3R tau ($\Delta 2,3,10$; $\Delta 3,10$; $\Delta 10$), whereas those retaining this exon are called 4R tau. Exons 2 and 3 encode the N-terminal domains that are important for microtubule spacing in axons and for the interaction of tau with the plasma membrane. Exon 10 encodes an additional microtubule binding repeat (4R tau) that is involved in microtubule interactions (Buee et al., 2000). Tau variants lacking exons 2 and 3 appeared to be present at different stages of rat neuronal development and were suggested to contribute to plasticity in the adult rat brain (Collet et al., 1997). Alternative expression of exon 10 plays an important role in the pathogenesis of various tauopathies (i.e., frontotemporal lobe degeneration, progressive supranuclear palsy and corticobasal degeneration) in which the ratio of 4Rtau to 3R tau isoforms is altered. In control subjects these tau isoforms are equally distributed with a ratio 4R tau/3R tau being around 0.48 (Ingelsson et al., 2007). In AD brain both 4R tau and 3R tau are present with a notable abundance of 4R tau isoforms. Levels of exon 10 containing tau mRNAs (4R tau) were elevated in affected brain areas of sporadic AD cases (Glatz et al., 2006; Niblock and Gallo, 2012). Furthermore, exon 10 immunoreactivity (indicating the presence of 4R tau) was demonstrated in intracellular NFT (Ishizawa et al., 2000). Although there were no significant differences in 4R tau/3R tau mRNA ratio in CA1 hippocampal neurons and entorhinal cortex between AD and control groups in the study of Ingelsson et al. (2006), methodological issues should be taken into account (Conrad et al., 2007). Detailed analysis of exons 2, 3 and 10 showed an increased incidence of exon 10 inclusion and exon 2 exclusion in the temporal cortex of AD cases leading to the overexpression of $\Delta 2,3$ 4R tau (Conrad et al., 2007). Taken together, the data indicate that abnormal splicing of tau mRNA resulting in the overproduction of 4R tau isoforms is one of the mechanisms by which tau accumulates in the AD brain. These alterations were not related to tau gene mutations. Therefore, pharmacological or biological agents that are involved in the regulation of tau alternative splicing may be used to delay the progression of neuropathology in AD (Zhou et al., 2008; Niblock and Gallo, 2012).

More recently, 6D and 6P tau isoforms with unique 11 amino acid sequences resulting from alternative splice sites at the end of exon 6 were identified (Lapoint et al., 2009). They contain only the N-terminus and a part of the proline-rich region. 6D and 6P isoforms lack the MTBR region that is essential for tau-microtubule binding and tau aggregation. Both of these splice variants were detected in all brain areas, including the cerebral cortex and the hippocampus. Interestingly, 6D protein expression was particularly high in the human cerebellum that is not affected by tau neuropathology and was very low in the hippocampus and cerebral cortex that are known tangle-prone regions (Luo et al., 2004). Moreover, 6D did not co-localize with neurofibrillary tangles reinforcing the idea about its neuroprotective role. In a functional study, indeed, 6D and 6P isoforms were shown to inhibit polymerization of the full-length tau, and were suggested to be endogenous inhibitors of tau filament formation (Lapoint et al., 2009). In conclusion, a loss in coordination of tau alternative splicing may be an important contributing factor to the pathogenesis of AD (Glatz et al., 2006; Conrad et al., 2007; Niblock and Gallo, 2012) (figure 1).

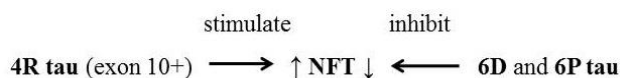


Figure 1. Influence of different tau splice variants on NFT accumulation.

APP (Amyloid Precursor Protein) Gene

In the amyloid plaques in the brain of AD patients APP-derived A β is a main constituent. The APP gene has 19 exons, from which exons 7 and 8 can be alternatively spliced in neurons resulting in the formation of three major isoforms: APP695, APP751 and APP770 (Beyreuther et al., 1993; Tanaka et al., 1992; Rockenstein et al., 1995; Thakur and Mani, 2005). In APP751 exon 8 is skipped, whereas APP695 lacks both exons 7 and 8. Exon 7 encodes the domain of Kunitz-type serine protease inhibitor (KPI), whereas exon 8 codes for the OX-2 domain which is homologous to the MRC OX-2 antigen present on the surface of neurons and immune cells. Consequently, APP770 possesses both of these domains. APP751 has the KPI but lacks the OX-2 domain, whereas APP695 is devoid of both domains (Tanaka et al., 1992; Rockenstein et al., 1995; Thakur and Mani, 2005). APP695 is produced by neurons and is the most abundant in the brain. Brain levels of APP770 are generally low. The reason for the higher expression of APP695 and low levels of APP770 in the brain was found in the embryonic development. APP695 was expressed in neuroectodermal derivatives, whereas APP770 was restricted to mesodermal and endodermal derivatives (Sarasa et al., 2000). Brain aging is associated with a decline in APP695. In old mice mRNA levels of this isoform were lower than in adult animals. Furthermore, gonadectomy decreased APP695 mRNA levels in adult females, whereas estradiol supplementation reversed this effect (Thakur and Mani, 2005). In humans APP695 is the dominant isoform in younger individuals, whereas in the older subjects its levels are significantly reduced (Konig et al., 1989). In AD brain APP695 mRNA levels were diminished, whereas APP751 and APP770 were increased (Thakur and Mani, 2005; Barrachina et al., 2005; Rockenstein et al., 1995; Tanaka et al., 1992). It should be noted here that an age-related rise in APP770 and APP751 was also reported in non-AD subjects, but the ratio of the KPI-harboring (APP770+APP751) to KPI-lacking (APP695) isoforms increased more dramatically and earlier in the AD cases (Tanaka et al., 1992). The increase in APP isoforms with KPI domain was most prominent in the brain areas that are most severely affected by AD process, i.e., in the hippocampus and the nucleus basalis of Meynert and was proposed to be a possible origin of A β deposition (Rockenstein et al., 1995; Barrachina et al., 2005). This observation in AD brain was reinforced by similar data obtained in PDGF-hAPP transgenic mice, expressing AD pathology. The animals showed high levels of APP770 and APP751 and lower levels of APP695 mRNA (Rockenstein et al., 1995). On the other hand, a specific increase in APP695 mRNA was noted in brain regions involved in amyloid plaque formation (Jacobsen et al., 1991). This observation was explained by the fact that the deposition of A β in senile plaques is derived from APP695 lacking KPI and OX-2 domains (Thakur and Mani, 2005). Therefore, it was suggested that an imbalance in the protease inhibitor plays an important role in AD (Jacobsen et al., 1991) and that the deregulation of splicing of the exon 7 in brain aging is a putative risk factor for AD (Beyreuther et al., 1993). In addition to exons 7 and 8, exon 15 of the APP gene is also subject to alternative splicing. Omission of the exon 15 results in L-APP isoform with a functional recognition sequence for xylosyltransferase-mediated addition of glycosaminoglycans and proteoglycan formation (Sandbrink et al., 1997). In the brain it can be found in activated astrocytes and microglia. The absence or low levels of L-APP isoforms in neurons might be important for the susceptibility of the brain towards AD (Beyreuther et al., 1993; Sandbrink et al., 1997).

A β which is the main constituent of senile amyloid plaques is cleaved from APP by two proteolytic enzymes, the β - and γ -secretases (Hardy and Selkoe, 2002). The β -secretase is a transmembrane aspartyl-protease that is recognized as the β -site amyloid precursor protein-

cleaving enzyme 1 (BACE1). BACE1 first cleaves the APP at the N-terminus of the A β . This results in the formation of a 100 kDa soluble fragment and a 12kDa membrane anchored protein C99 at the C-terminus. The C99 is then cleaved by the γ -secretase producing the A β 40 and A β 42 species. Alterations in the balance between the β - and γ -secretases and their substrate APP were proposed to influence the development of AD (Zohar et al., 2005).

Alternative Splicing of Genes involved in A β Processing

BACE 1 (β -site APP Cleaving Enzyme 1)

BACE 1 activity and protein levels are significantly elevated in the brain of AD patients (Fukumoto et al., 2002). BACE1 overexpression increases the accumulation of A β in the brain, whereas BACE1 inhibitors reduce the A β formation (review in Annies et al., 2009). Deletion of the BACE1 gene in mice improves memory deficits and recovers neurodegeneration (Ohno et al., 2007). The BACE1 gene has nine exons that are translated into the full-length 501 amino acid active protein BACE1 501. The use of cryptic donor and acceptor sites in exons 3 and 4 generates in-frame deletions within the catalytic domain of BACE1 (Tanahashi and Tabira, 2001). Deletion of 75 bp inside exon 4 due to the alternative 3' splice site produces isoform BACE1 476 (the length of the protein is 476 amino acids compared to 501 amino acids in the full length protein). Deletion of 132bp inside exon 3 due to the alternative 5' splice site forms the isoform BACE1 457 (457 amino acids). Isoform BACE1 432 (432 amino acids) results from the use of both mentioned above alternative 5' and 3' splice sites in exons 3 and 4 (Tanahashi and Tabira, 2001). In vitro studies showed that enzymatic activity of BACE1 isoforms is dramatically decreased (Tanahashi and Tabira, 2001, Mowrer and Wolfe, 2008). Moreover, following co-expression of BACE1 476 and BACE1 457 with APP695 in HEK 293 cells, significant reduction in the secretion of A β 40 and A β 42 was noted (Tanahashi and Tabira, 2001). Furthermore, the use of antisense RNA oligos targeted against normal 5' and 3' BACE1 splice sites also showed substantial decrease in A β 40 production in HEK-swe cells (Mowrer and Wolfe, 2008). Although all of the known BACE1 isoforms were identified in different human brain regions, no data are available concerning possible changes in BACE1 alternative splicing in AD patients. At the moment there is only one study of BACE1 splice variants in tg2576 mice with the Swedish mutation of APP (Zohar et al., 2005). These mice develop age-dependent A β plaques and learning deficits. The ratio of BACE1 splice variants to the wild type BACE1 501 was diminished in different brain areas of tg2576 mice. This decrease was more pronounced for the BACE1 476 and BACE1 432 isoforms. In this animal model of AD the full length BACE1 with the highest enzymatic activity showed a huge, 2 to 4-fold dominance over the splice variants. On the contrary, in wild type animals the expression of BACE1 isoforms was even comparable to that of the full length BACE1 (BACE1 501 ~ BACE1 476) (Zohar et al., 2005). Together, these studies demonstrated that the increase in the levels of BACE1 splice variants with diminished enzymatic activity reduces A β production and is protective against AD development (figure 2). Thus, targeting of BACE1 normal splice sites with antisense RNA oligos or affecting factors that regulate BACE1 splicing may be promising therapeutic strategies in AD patients (Mowrer and Wolfe, 2008).

The γ -secretase complex responsible for the C-terminal cleavage of the APP consists of at least four proteins: presenilin, nicastrin, APH-1 (anterior pharynx defective) and PEN-2 (presenilin enhancer 2). Presenilin and APH-1 exist in two forms: presenilin 1 and 2, APH-1a

and APH-1b. APH-1a and APH-1b incorporate into different presenilin complexes. Moreover, APH-1a has two major splice variants APH-1aS and APH-1aL (Lee et al., 2002; Saito et al., 2005). Consequently, at least six types of the γ -secretase complex may be present depending on the type of presenilin and APH-1 inclusion and all of them were shown to have proteolytic activity (Shirotani et al., 2004).

Presenilins

Presenilins (PSs) represent a group of multi-pass transmembrane proteins that function as catalytic subunits of the γ -secretase intramembrane protease complex. In particular, PSs are required for the cleavage of the C-terminal fragment of the APP within the extracellular domain. PS-1 containing complexes have higher γ -secretase activity than those with PS-2. PSs gene mutations lead to an increase in the production of A β 42 that is the predominant compound accumulated in senile plaques. Like for the APP gene, mutations in the PS-1 and PS-2 genes were linked to familial (early-onset) autosomal dominant AD (Manabe et al., 2007; Suzuki et al., 2009).

Alternative splicing at the 3' end of exon 3 of PS-1 mRNA results in the appearance of the long form of PS-1 mRNA (VRSQ+) that contains four amino acids (VRSQ) and the short form VRSQ- that is devoid of this sequence (Isoe-Wada et al., 1999). In the brain of subjects with sporadic AD VRSQ+ isoform mRNA levels were reduced as compared to the control group. This alteration in exon 3 alternative splicing was brain-specific and was proposed to play a role in AD pathogenesis (Isoe-Wada et al., 1999).

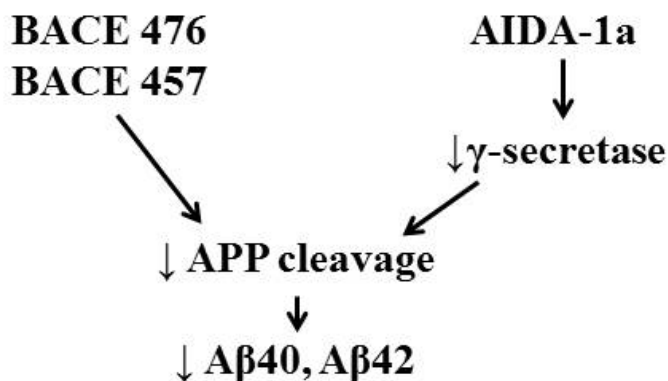


Figure 2. Alternatively spliced transcripts may decrease A β secretion.

A splice donor site mutation in intron 4 of PS-1 is present in early-onset AD cases and is inherited in an autosomal dominant manner (De Jonghe et al., 1999). It produces 2 deletion isoforms (Δ 4 and Δ 4 cryptic) that form C-truncated PS-1 proteins (7kD) and one variant with an insert (insTAC) that is translated into a full-length PS-1 with one extra amino acid threonine between codons 113 and 114 (PS-1 T113-114ins). Following expression in vitro, these isoforms appeared to increase A β 42 production. Furthermore, the presence of the PS-1 T113-114ins was linked to the AD pathophysiology in intron 4 mutation carriers (De Jonghe et al., 1999). It should be noted that PS-1 splice variants are not necessarily related to AD pathology. For example Δ 8 isoform (lacking exon 8) was found not to influence A β production (Moriyama et al., 2000).

Concerning PS-2 two major splice variants associated with AD are known, i.e., PS2V and PS2 β . PS2V lacks exon 5 and encodes the N-terminal domain of the PS-2. It is translated into 14kDa deleterious protein with a novel C-terminus created by alternative splicing (Smith et al., 2004; Manabe et al., 2007; Sato et al., 1999, 2001). Exclusion of exon 5 during PS-2 mRNA processing is induced by hypoxia-mediated oxidant stress (Sato et al., 1999). This isoform is much more frequently expressed in the brain of sporadic AD cases. It was observed in 70% of AD and only 17.6% of control subjects (Sato et al., 1999; Manabe et al., 2007). PS2V protein levels were increased 2-fold in the frontal cortex of AD patients (Smith et al., 2004). Moreover, this isoform was found in neuropathologically affected neurons of the CA1 hippocampal subfield and temporal cortex of AD subjects. Furthermore, PS2V overexpression was shown to increase the production of A β 40 and A β 42 (Sato et al., 2001). Taken together, these data indicate that PS2V isoform contributes to neurodegeneration in AD (Sato et al., 2001; Smith et al., 2004).

PS2 β was so far cloned only in the mouse. This isoform results from small extension of the exon 9 of the PS-2 gene leading to the translation of four additional out-of-frame residues at the loop domain (Suzuki et al., 2009). As a consequence of these splicing events, PS2 β appeared to have the N-terminal part and a hydrophilic loop domain, but it lacked the seventh transmembrane domain and downstream C-terminal regions. PS2 β blocked formation of the γ -secretase complex in HEK293 cells by disturbing the interaction between nicastrin and APH-1 (anterior pharynx-defective phenotype 1). In this way it inhibited γ -secretase activity and reduced A β levels (Suzuki et al., 2009). Since the splicing sites of the PS2 β were conserved between mouse and human, it is plausible that this variant may be involved in the A β production in AD patients (figure 3). However, its expression in the AD brain remains to be elucidated.

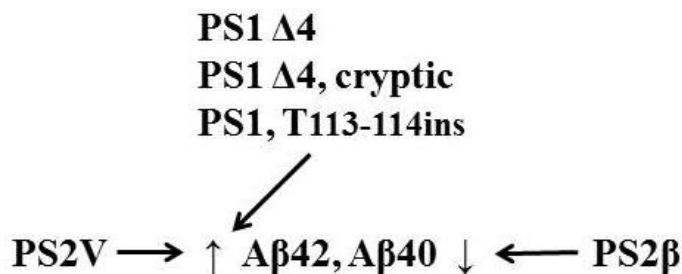


Figure 3. Differential effect of PS-1 and PS-2 splice variants on A β 42 and A β 40 generation.

Nicastrin

Nicastrin maintains the stability of the γ -secretase complex and is important for its activation. Down-regulation of nicastrin expression by RNA-i leads to the decrease in the levels of other γ -secretase components, i.e., PS-1, APH-1 L and PEN-2 and, subsequently, to the reduction in A β secretion (Capell et al., 2003). Deletion of exon 16 (NCTN- Δ E16) of the nicastrin gene was identified in the human brain. In the hippocampus of APOE- ϵ 4 carriers this splice variant was observed more often in AD cases than in the non-AD group. It was, therefore, concluded that the presence of the NCTN- Δ E16 isoform may confer additional risk for the development of AD (Mitsuda et al., 2006). However, the difference in that study was not statistically significant, and it should also be pointed out that the comparison of AD subjects in this study was made with the non-AD group that included cases with Parkinson disease and

diffuse Lewy body disease. It is plausible that the exclusion of cases with neurological disorders would clarify the AD-related specificity of the NCTN- Δ E16 splice variant.

APH-1

APH-1 associates with nicastrin during early stage of γ -secretase complex assembly. This subcomplex of APH-1 and nicastrin further combines with PS and PEN-2 to form the mature γ -secretase. Two APH-1 genes are known in humans: APH-1a and APH-1b (Lee et al., 2002). APH-1a is believed to be more important for the γ -secretase assembly and activity than APH-1b (Shirotani et al., 2004). Down regulation of APH-1a, but not of APH-1b, results in altered maturation of nicastrin and decreased expression of PS-1, PS-2 and PEN-2 proteins (Lee et al., 2002; Shirotani et al., 2004; Saito and Araki, 2005). APH-1a has two alternatively spliced isoforms APH-1aS (247 amino acids) and APH-1aL (265 amino acids) that differ in C-terminal sequences. APH-1aS has exons 1 to 5 with a stop codon at the beginning of exon 6 (exon 6a). APH-1aL contains exons 1 to 5 and exon 6a with a stop codon in the last two thirds of the exon 6 (designated as exon 6b) (Lee et al., 2002). In human tissues expression of APH-1aS is 1.5 to 3 times higher than that of APH-1aL (Saito and Araki, 2005). This means that γ -secretase complex might include the APH-1aS isoform more often and that its pathogenic activity is rather related to this particular APH-1 splice variant.

APH-1b has a splice variant APH-1b Δ 4 lacking exon 4 that encodes the fourth transmembrane domain. It is characterized by the in-frame deletion of 123 nucleotides and is devoid of the conserved motif GXXXG that is important for the assembly with presenilins. APH-1b Δ 4 mRNA is expressed at low levels and is translated into unstable protein. Despite these caveats, APH-1b Δ 4 interacts with nicastrin in transfected HeLa cells (Saito et al., 2005). However, the relevance of this splice variant for the AD pathogenesis remains to be elucidated.

AIDA-1

Upon processing APP several fragments other than A β are released. One of them is short intracellular (cytoplasmic) domain (AID) that is formed following the γ -secretase proteolysis. It plays a role in apoptosis, calcium homeostasis and transcriptional regulation (Ghersi et al., 2004). AID interacts with many proteins that modulate its functions or interfere with APP processing and/or internalization. AIDA-1 (AID-associated protein 1) is the novel AID-binding protein that modulates A β processing. Several alternatively spliced variants of AIDA-1 are known from which only AIDA-1a binds to APP and diminishes A β secretion through the inhibition of γ -secretase (Ghersi et al., 2004). AIDA-1a is a truncated isoform that has exons 15 to 19 and then exons 21 to 27. It is lacking exons 1 to 14 and exon 20 (Δ 1-14, 20). AIDA-1a was shown to be present in the normal human brain but was absent in AD brain tissues (Ghersi et al., 2004). It is therefore, conceivable that a loss of AIDA-1a which is down regulating APP processing is an important step in AD pathogenesis (figure 2).

Clusterin

Clusterin, known as apolipoprotein J, attaches to A β peptides and inhibits their fibrillization. This protein is also needed for the clearance of A β peptides and fibrils by glial cells. Clusterin suppresses complement protein activation in AD (review in Nuutinen et al., 2009). Its expression is increased in the brain of AD cases, in particular in the hippocampus and entorhinal cortex. Clusterin immunoreactivity was observed in neuritic plaques, neuropil

threads and cerebrovascular amyloid deposits. However, it was rarely seen in diffuse amyloid plaques or neurons with neurofibrillary tangles (Giannokopoulos et al., 1998). It was proposed that the enhanced expression of clusterin in AD brain may rather be neuroprotective (Nuutinen et al., 2009). Deletion of exon 2 results in the appearance of the alternatively spliced variant that is translated into uncleaved and non-glycosylated protein with two nuclear localizing sequences. This $\Delta 2$ clusterin isoform translocates to the nucleus where it prevents the DNA repair and induces apoptosis through the interaction with Ku70 protein (Leskov et al., 2003). Although the significance of DNA damage and alterations in DNA repair in AD pathogenesis is well understood (Cotman and Su, 1996), the precise role of the nuclear clusterin lacking exon 2 in AD remains to be clarified.

FE65 Family Proteins

The FE65 family includes adaptor proteins FE65, FE65L1, FE65L2 that bind to the cytoplasmic domain of APP for its subsequent internalization. The FE65-APP complex is important for the modulation of the metabolism and functions of APP (Hu et al., 2002). In particular, FE65 binding to APP increases the proteolytic processing of the latter resulting in the augmented secretion of A β 40 and A β 42 (Sabo et al., 1999; Tanahashi and Tabira, 2002). Increased expression of FE65 mRNA in the cerebellar cortex and enhanced FE65 immunoreactivity in the CA4 hippocampal area were found in patients with very late onset of AD (Hu et al., 2000; Delatour et al., 2001). It was furthermore reported that carriers of the minor allele 2 of FE65 were resistant to the very low onset AD. This protective role was associated with a polymorphism in intron 13 that causes the appearance of the cryptic acceptor splicing site within the last exon 14 resulting in the alternatively spliced isoform FE65a2 with 55 amino acid deletion at the C-terminus (Hu et al., 2002). FE65a2 showed reduced binding to APP (Hu et al., 2002). Importantly, this observation demonstrates that preferential expression of the alternatively spliced isoform with attenuated functional properties may lead to the decreased risk of the very low onset AD.

Modulation of alternative splicing of FE65L1 mRNA by the non-coding RNA 45A in vitro results in the preferential generation of exon 8-containing isoforms “a” and “b” that have reduced ability to interact with APP. This causes a decrease in the secretion of β -amyloid but increases the ratio of A β 42/A β 40 ratio providing favourable conditions for A β aggregation (Penna et al., 2013).

Alternative splicing of the FE65L2mRNA generates nuclear isoforms I-214 and I-245 lacking phosphotyrosine-interaction domain (PID) 2 that is necessary for the interaction with APP. Both of them also have alterations within PID1. I-214 isoform (214 amino acid length) is produced by the use of alternative donor splice site in exon 6 and alternative acceptor site in intron 6. These splicing events cause the deletion of 66 amino acids with the appearance of unique 19 amino acids at the C-terminus. Increased ratio of I-214 to canonical FE65L2 was associated with augmented apoptosis in vitro. Moreover, in contrast to the full length FE65L2, I-214 did not influence secretion of A β 40 and A β 42 (Tanahashi and Tabira, 2002). This finding shows that overexpression of the alternatively spliced isoform I-214 may not be simply regarded as positive or negative with respect to its possible role in AD pathogenesis. On one hand I-214 is pro-apoptotic but at the same time it may be protective against A β species.

Apolipoprotein E (APOE)

APOE is considered to be important for the regional vulnerability of neurons in AD (Xu et al., 1999). Moreover, APOE participates in the clearance of A β , protection against A β neurotoxicity, stabilization of microtubules and neuronal repair in a receptor-mediated way (Clatworthy et al., 1999). APOE gene has several alleles, from which the e4 is associated with increased risk and earlier onset of AD possibly because it is accompanied by diminished neuronal activity (Dubelaar et al., 2004), e3 is regarded as neutral, whereas e2 is protective against AD. By means of RNA Seq three isoforms of the APOE gene originating from different transcription sites were identified in the temporal lobe of control and AD patients with APOE e3 genotype. Splice variants APOE 001 and 002 contained all 4 exons and were transcribed from promoter TSS A. APOE 005 was devoid of the first exon since it was generated by an alternative promoter TSS B upstream of the second exon (Twine et al., 2011). In AD cases APOE 001 and 002 were diminished. This decrease was much more pronounced for the APOE 002 isoform that was specific to the temporal cortex. On the contrary, APOE 005 was significantly increased in AD subjects. Accordingly, TSS A promoter was 3-fold down-regulated and TSS B was 26.5-fold up-regulated in the AD group (Twine et al., 2011). Interestingly, differential expression of the APOE isoforms and promoter use with regards to AD was noted in this study only in the temporal cortex and not in the total brain samples. Therefore, it is plausible that alterations in the APOE gene splicing pattern are related to AD neuropathology. The precise role of the APOE splice variants in AD remains to be elucidated.

Alternative Splicing of Other Candidate Genes

Acetylcholinesterase

AD is associated with a pronounced cholinergic deficit resulting from the loss and decreased metabolic activity of cholinergic neurons. These alterations are directly related to learning and memory problems in AD patients (Cummings and Back, 1998; Swaab, 2004). A link between cholinergic dysfunction and amyloid neuropathology is present in the 3' alternative splicing of the acetylcholinesterase (AChE) gene that generates C-terminal isoforms with opposed effects on amyloid fibril deposition (Inestrosa et al., 1996; Berson et al., 2008). In the synaptic AChE-S variant, that is the major transcript in the brain, exon 5 is deleted. The second most common isoform AChE-R (the "readthrough" form) contains a pseudo-intron 4 (I4) between exons 4 and 5 (Grisaru et al., 1999). In vitro, AChE-S increases A β fibril formation and enhances their neurotoxic effects (Inestrosa et al., 1996; Berson et al., 2008). AChE-R appeared to act as an antagonist of AChE-S since it reduced the formation of insoluble A β oligomers and fibrils and abolished A β toxicity to SH-SY5Y cells in a dose-dependent manner (Berson et al., 2008). In double transgenic APPsw/AChE-R mice cortical amyloid plaque accumulation was reduced as compared to single APPsw mice. In AD patients AChE-R protein was decreased in hippocampal neurons. Although, AChE-R mRNA was up-regulated in the same brain region of AD cases, both AChE-R mRNA and protein appeared to be less stable than those of AChE-S variant. It was, therefore, concluded that despite protective direction of 3' alternative splicing of AChE pre-mRNA in AD brain, the protein of the favorable AChE-R isoform may undergo enhanced proteolysis (Berson et al., 2008). It is of particular interest that cholinesterase inhibitors that are used as symptomatic treatment for AD, show differential

effects on the CSF protein levels of AChE-R and AChE-S. In AD patients treated for a year rivastigmine selectively up-regulated the AChE-R, while tacrine increased both AChE-R and AChE-S isoforms. In untreated subjects after 1 year follow-up AChE-R levels diminished while AChE-S concentrations raised. These changes were accompanied by declined MMSE scores (Darreh-Shori et al., 2004). To this end the data indicate that an increase in the neuroprotective AChE-R isoform may be beneficial in the course and treatment of AD.

Tissue Transglutaminase

Tissue transglutaminase (TGase2 or tTG) is the most common member of the transglutaminase family. It has GTP-ase activity, binds both GTP and ATP and is involved in the regulation of signal transduction. Moreover, tTG mediates protein crosslinking and, therefore, is implicated in the pathogenesis of neurodegenerative disorders with excessive protein aggregation (Citron et al., 2002). In AD both tTG enzymatic activity and gene expression are elevated (Kim et al., 1999; Citron et al., 2001; Citron et al., 2002). Furthermore, tTG crosslinks APP and A β , binds to tau and colocalizes with both amyloid plaques and neurofibrillary tangles (Ikura et al., 1993; Ho et al., 1994; Zhang et al., 1998; Citron et al., 2001). Alternative splicing of the tTG pre-mRNA may result in the appearance of long (L) and short (S) variants. Short isoforms lack the GTP-binding domain and are more active in protein crosslinking. In the brain of non-demented subjects only long forms of the tTG were present. In AD brain samples both L and S isoforms were observed suggesting that neurodegenerative events promote tTG alternative splicing (Citron et al., 2001). Moreover, the selective presence of the S tTG variant with unregulated crosslinking only in AD cases indicates that this isoform is largely involved in tau and amyloid neuropathology. This observation further establishes an important link between the AD pathogenesis and alternative splicing of the tTG mRNA.

Estrogen Receptor α

Menopause-related loss of estrogens in women is an important risk factor for AD (Naftolin and Malaspina, 2007; Barron and Pike, 2012). Consistent with this, women have a higher risk for developing AD than men (Craig and Murphy, 2008). Although not without controversy, estrogen therapy initiated during perimenopause within the "critical window" can improve cognitive functions and reduce the risk of dementia (Brinton, 2005; Zandi et al., 2002; Craig and Murphy, 2008). Estrogens mediate their effects mainly via two types of cognate receptors: ER α and ER β . The survey of the literature shows that targeting ER α may offer promising alternative to estrogen therapy to protect cognitive functions in menopausal women (Bailey et al., 2011). In our own studies we found that in AD nuclear expression of ER α is elevated in the neurons of the hypothalamus and basal forebrain involved in the regulation of memory, but was decreased in the hippocampus (Ishunina and Swaab, 2001; Ishunina et al., 2003; Ishunina et al., 2007). Since a lot of changes in the alternatively spliced isoforms are intimately linked to AD (splice variants within plaques, splice variants of enzymes involved in A β processing, i.e., BACE, splice forms of acetylcholinesterase and their selective changes under different cholinesterase inhibitors, (see above)), the idea that the sensitivity to estrogen treatment may be related to ER splice variants was further explored. Investigation of the alternative splicing of the ER α mRNA in different brain areas showed that both the number of ER α splice variants per brain area and mRNA levels of the most common ER α splice forms lacking exon 7 or 2 were significantly diminished in AD cases compared to the non-demented control group. These alterations were more pronounced in AD women than in AD men. Among

numerous ER α isoforms deletion of the exon 7 ($\Delta 7$) appeared to be the most ubiquitous in the human brain (Ishunina and Swaab, 2012). This is a dominant negative variant lacking a substantial portion of the ligand binding domain. Consequently, $\Delta 7$ has minor ability to bind estrogens. Its capacity to bind DNA is also reduced (Bollig and Mikcisek, 2000; Garcia-Pedrero et al., 2003). Nevertheless, $\Delta 7$ may form heterodimers with both canonical ER α and ER β and in this way it significantly suppresses estrogen signaling in different cell types (Garcia-Pedrero et al., 2003; Wong and Weickert, 2009). The percentage of brain areas where $\Delta 7$ was observed as a single form of ER α was nonsignificantly larger in AD (29%) than in control (16%) cases. Moreover, the frequency of complex ER α splice variants with large deletions due to alternative 5' and 3' splice sites inside exons was higher in AD than in control women (Ishunina and Swaab, 2012). With the help of antibodies recognizing specific splice sites of one of the novel ER α splice variants TADDI that we identified in the human hippocampus we further assessed its protein expression in the brain of AD cases. In this isoform 31 bp are deleted from the junction between exons 3 and 4, and 13 nucleotides from exon 2 are inserted into this splice site (Ishunina et al., 2007). In HeLa cells TADDI was dominant negative since it significantly suppressed transcriptional activity of the wtER α (wild type or canonical ER α). However, in dopaminergic M17 cells this splice variant was clearly dominant positive. Its own transcriptional activity was higher than that of the wtER α . Following co-expression with wtER α , estradiol-stimulated transcriptional activity augmented profoundly to the level of TADDI+ wtER α (Ishunina et al., 2013). Immunocytochemical expression of TADDI was clearly decreased in the hippocampus, nucleus basalis of Meynert and hypothalamic tuberomammillary and supraoptic nuclei of AD female subjects compared to control women (Ishunina and Swaab, 2009). Together, these data show that alternative splicing of the ER α mRNA is down-regulated in AD and more significantly in female than in male patients. At the same time, molecular defects in ER α mRNA variants are more severe in AD women than in the female cases of the control group.

Brain Derived Neurotrophic Factor (BDNF)

BDNF promotes survival, metabolic activity, synaptic transmission and excitatory properties of the hippocampal, cortical and basal forebrain cholinergic neurons (reviewed by Garzon et al., 2002). The loss of function and synaptic connectivity of these neurons in AD brain was attributed to the decline in BDNF. Indeed, both mRNA and protein levels of BDNF were significantly decreased in the hippocampus and cortex of AD patients (Phillips et al., 1991; Holsinger et al., 2000; Connor et al., 1997; Hock et al., 2000). Alternative splicing of 6 upstream non-coding exons to the single coding exon 5 produces at least 7 transcripts designates as 1 (exon 1 spliced to exon 5), 2 (exon 2 joined with exon 5), 3 (exon 3 spliced to exon 5), 4 (exon 4 connected to exon 5), 4I (exon 4I joined with exon 5), 4Ia (sequence inside exon 4I spliced to exon 5), 5U (exon 5U joined with exon 5) (Aoyama et al., 2001; Garzon et al., 2002). Splice variants 1, 2 and 3 were significantly down-regulated in the parietal cortex of AD patients, while other isoforms were not changed (Garzon et al., 2002). These data point to the dysregulation of alternative splicing of the BDNF mRNA that may contribute to neurotrophic imbalances in the AD brain.

Tropomyosin Receptor Kinase B (TrkB)

The neurotrophic effects of BDNF are mediated by the tropomyosin receptor kinase B (TrkB) (Patapoutian and Reichardt, 2001). The full-length TrkB-TK⁺ (with intracellular

tyrosine kinase domain (TK+)) is encoded by 24 exons and is expressed primarily in neurons. Two C-terminal truncated isoforms, i.e., TrkB-TK- and TrkB-Shc, are generated by alternative splicing and lack the tyrosine kinase domain. TrkB-TK- is encoded by 16 exons. TrkB-Shc includes exon 19 with associated premature stop codon that is absent in the full-length TrkB-TK+. TrkB-Shc is mainly confined to neurons, whereas TrkB-TK- to glial cells. Both isoforms can dimerize with TrkB-TK+ and function as dominant negative splice variants that reduce neurotrophin signaling (Stoilov et al., 2002; Wong et al., 2012). TrkB-Shc mRNA and protein expression were markedly elevated in the hippocampus of AD cases, whereas the levels of either TrkB-TK+ or TrkB-TK- were not changed. Furthermore, TrkB-Shc appeared to colocalize with TrkB-TK+ and suppressed the BDNF/TrkB-TK+ signaling via the mitogen-activated protein kinase (MEK). Interestingly, the levels of TrkB-Shc mRNA in differentiated neuronal SHSY5Y cells were increased in the presence of fibrillar A β 42 (Wong et al., 2012). Given that overexpression of TrkB-Shc may decrease APP intracellular domain-mediated transcription that is known to promote apoptosis and to contribute to the A β 42 toxicity (Ansaloni et al., 2011; Wong et al., 2012), the reported up-regulation of TrkB-Shc in AD hippocampal neurons may represent a compensatory response and may promote neuronal survival (Wong et al., 2012).

Excitatory Amino Acid Transporter 2 (EAAT2)

Preferential degeneration of glutamatergic neurons in AD leads to an early glutamatergic deficit that is apparent before cognitive decline. Glutamate transport is decreased in the cortex of AD cases (Bert and O'Shea, 2007). This may result in the increased accumulation of glutamate in synaptic clefts and further overstimulation of postsynaptic glutamate receptors causing neuronal loss via excitotoxicity (Masliah et al., 1996). Moreover, cortical neurons receiving glutamatergic projections represent the preferred location of neurofibrillary tangles (Francis, 2003). Glutamate transport is mediated by excitatory amino acid transporters (EAAT) from which the major glial type EAAT2 was related to the decreased glutamate uptake in AD brain tissues. EAAT2 mRNA was down-regulated in the hippocampus and various cortical areas of AD cases (Li et al., 1997; Scott et al., 2011). EAAT2 immunoreactivity was diminished in the AD frontal cortex (Li et al., 1997). Furthermore, EAAT2 was found to co-localize with hyperphosphorylated tau protein in dystrophic neurons (Thal, 2002). Two exon-skipping splice variants of the EAAT2 mRNA lacking exons 7 (EAAT Δ 7) and 9 (EAAT2 Δ 9) were increased in the cerebral cortex of AD patients and showed positive correlation with pathological severity. These isoforms are unable to transport glutamate themselves and suppress glutamate transport capacity of the wtEAAT2 (Scott et al., 2011). Consequently, higher expression of dominant negative EAAT2 splice variants EAAT Δ 7 and EAAT2 Δ 9 suggests their significant contribution to the reduced glutamate transport and increased neuronal death by excitotoxicity in the brain of AD cases. Identification of the EAAT2 Δ 9 protein in dystrophic neuritis and glial cell processes in the vicinity of amyloid plaques in AD brain further supports the role of EAAT2 isoforms in neurodegeneration (Pow and Cook, 2009).

Ion Channels

Alterations in the function of ion channels, like glutamate receptors, nicotinic cholinergic receptors, calcium and potassium channels, have been clearly related to AD pathophysiology. Moreover, AD-associated neurodegeneration was partially linked to the overactivation of N-methyl-D-aspartate receptor followed by the increase in intracellular calcium and oxidative

stress (Heinzen et al., 2007; Hynd et al., 2004). The splice variant microarray showed that 12% of the ion channel genes are alternatively spliced in AD (Heinzen et al., 2007). These changes are briefly summarized in the following table.

Thus, approximately equal proportions of the ion channel genes demonstrate up- or down-regulated alternative splicing in AD. Although the precise role of these changes remains to be elucidated, the data are in agreement with the general idea of this chapter that alternative splicing is dysregulated in AD.

Table 1. Ion channel genes with increased or decreased alternative splicing

Ion channel genes	The number of genes with increased alternative splicing	The number of genes with decreased alternative splicing
Ca ²⁺ channels	2	4
Chloride channels		4
Sodium channels		3
Potassium channels	9	5
GABA receptors	2	1
Ionotropic glutamate receptors	1	2
Two pore segment channels	1	
mucolipin	2	
Ryanodine receptors	2	

Synapsin

Synaptic pathology plays an extremely important role in the pathophysiology of AD since the number of synapses decreases in the AD brain and because A β oligomers bind to synapses involving the latter into the fine structure of senile plaques. Certain synaptic markers, i.e., synapsins and synaptophysin, were reported to be decreased in AD reflecting these synaptic alterations (Duyckaerts et al., 2009). Synapsins represent a group of phosphoproteins that regulate the number of synaptic vesicles available for neurotransmitters (Greengard et al., 1993). Synapsin II is associated with excitatory synapses, whereas synapsin I – with inhibitory ones (Mandell et al., 1992). Synapsin II has two isoforms generated by alternative splicing: synapsin IIa and synapsin IIb. Synapsin IIa is implicated in synaptic plasticity and neurotransmitter release. It has splice variants Ia, IIa, IIIa (I-III) that were all selectively decreased in the entorhinal, but not the in visual cortex of cases with the earliest clinically detectable stages of AD (Ho et al., 2001). Because there was no change in the expression of IIb isoform, it seems reasonable to conclude that these data point again to the dysregulation of alternative splicing in AD. Diminished levels of the synapsin IIa variants were associated with the early phase of AD cognitive decline and were proposed to be useful in the development of novel strategies in the treatment of early AD (Ho et al., 2001).

Regulator of Calcineurin (RCAN1)

RCAN1 might well be implicated in the pathogenesis of AD since it prevents dephosphorylation of the tau protein by calcineurin. This results in the accumulation of

hyperphosphorylated tau and subsequent formation of paired helical filaments and neurofibrillary tangles (Ermak et al., 2001). Consistent with this functional role, RCAN1 gene expression was markedly enhanced in the brain areas of AD patients with pronounced neuropathological changes (Ermak et al., 2001; Harris et al., 2007). Three RCAN1 isoforms generated by alternative splicing are present in the human brain at both the mRNA and protein levels: RCAN1-1L, RCAN1-1S and RCAN1-4. They have the same C-terminus encoded by exons 5, 6 and 7 and differ in the first exon. RCAN1-1L ("Long") contains 55 N-terminal amino acids encoded by exon 1 that are directly spliced to 168 amino acids encoded by exons 5, 6 and 7 (1-5,6,7). RCAN1-1S ("Short") has 29 N-terminal amino acids coded by a part of the exon 1 that are similarly joined to the same C-terminal 168aa (1*-5,6,7). RCAN1-4 does not have the first exon. Its first 29aa are encoded by exon 4 followed by the common C-terminus generated by exons 5, 6 and 7 (4,5,6,7) (Harris et al., 2007). RCAN1-1L is the most abundant isoform in the human brain. It is selectively restricted to neurons and not to glial cells. Moreover, RCAN1-1L mRNA and protein levels were significantly elevated in the brain of AD cases and, in particular, in the brain regions most severely affected by AD neuropathology (i.e., the hippocampus and cerebral cortex) (Harris et al., 2007). These data point to an important functional role of RCAN1-1L splice variant in the formation of neurofibrillary tangles and pathogenesis of AD.

Glial Fibrillar Acidic Protein (GFAP)

GFAP is widely considered as a marker of astrocytes. However, there is clear evidence that this glial protein may appear in the hippocampal neurons of AD patients as a result of alternative splicing alterations. Neuronal GFAP is encoded by two out-of frame splice variants $\Delta 6$ (lacking exon 6) and $\Delta 164$ nt (with 164 nucleotides deletion from exon 6 and 5' portion of the exon 7 due to cryptic donor and acceptor splice sites inside exons 6 and 7). Their translation results in the formation of aberrant GFAP+1 proteins with the out-of-frame C-termini. These isoforms were associated with AD neuropathology since they were present in AD hippocampal neurons and were very rarely observed in the non-demented subjects (Hol et al., 2003). Generation of these mutant proteins was proposed to be due to the initiation of GFAP gene expression in AD neurons with concomitant failure of the proteasomal degradation machinery (Hol et al., 2003).

Ubiquilin-1

Proteasomal disfunction is an important pathological hallmark of AD. It was shown that alternative splicing of the proteins involved in the proteasomal machinery, for example ubiquilin-1, may be affected in AD. Ubiquilin-1 interacts with the proteasome and ubiquitin ligases and is involved in the regulation of protein degradation. In addition, with the help of the C-terminal ubiquitin-associated domain ubiquilin-1 interacts with both PS-1 and PS-2 promoting presenilin accumulation (Mah et al., 2000; Bertram et al., 2005). Therefore, ubiquilin-1 is considered as one of the candidate genes for AD. It has two main transcript variants that are designated as TV1 and TV2. TV1 has 11 exons. In TV2 exon 8 is deleted ($\Delta 8$) as a result of intronic single-nucleotide polymorphism located downstream of exon 8. This polymorphism in turn was linked to the risk allele UBQ-8i on chromosome 9q22. Carriers of one or two copies of the UBQ-8i showed a significant and dose-dependent increase in the risk of AD. In AD patients who have significantly increased frequency of the UBQ-8i allele, the ratio of TV2/TV1 isoforms was elevated (Bertram et al., 2005). Thus, alterations in alternative splicing of ubiquilin-1 mRNA in AD patients lead to the increased expression of the TV2

isoform lacking exon 8 and are associated with the single nucleotide polymorphism in the UBQ-8i risk allele. Since exon 8 encodes a part of the C-terminal domain for the interaction with protein-disulfide isomerase, certain aspects of protein degradation in the proteasome as well as interaction with PS-1 and PS-2 may be altered in AD patients carrying the UBQ-8i risk allele.

CDKN1A Interacting Zinc Finger Protein 1 (CIZ1)

CIZ1 is the DNA replication factor that regulates entry in the S-phase. Inappropriate re-entry of post-mitotic neurons into the cell cycle may alter their function in AD (Nagy, 2000). In-frame deletion of 168 nucleotides from the central portion of the CIZ1 exon 8 generates alternatively spliced isoform CIZ1S lacking 56 amino acids in the second glutamine-rich domain. In the hippocampus of AD cases CIZ1S mRNA was 2.5-fold up-regulated and was expressed at levels comparable with those for canonical CIZ1 mRNA (Mackeprang Dahmcke et al., 2008). These data suggested that increased expression of CIZ1S may contribute to alterations in the regulation of the cell cycle in AD.

Differentially Expressed in Normal and Neoplastic Tissues MAPK-Activating, Death Domain Protein (DENN/MADD)

Splice variants of the DENN/MADD are involved in the cell survival. One of them, IG20 (insulinoma-glucagonoma clone 20) is pro-apoptotic, whereas another one, the DM-SV, has opposing anti-apoptotic function. Knockdown of all of these variants results in apoptosis and tumor shrinkage (Efimova et al., 2004; Lim and Chow, 2002; Mo et al., 2012). In the central nervous system IG20 may be involved in synaptic vesicle cycling (Niwa et al., 2008), whereas the DM-SV was increased in Purkinje cells of the cerebellum following acute hypoxia (Zhang et al., 1998). In the hippocampus of advanced AD cases total DMI (DENN/MADD/IG20) complex is diminished (Del Villar and Millar, 2004). In SH-SY5Y cells oligomeric A β alter the ratio of DMI splice variants DM-SV/IG20 in a time-dependent manner. During first hours of exposure to A β this ratio increases and then after 24 hours becomes reversed. In vitro results are in concordance with the data in postmortem brain tissues. In the nucleus basalis of Meynert and the hippocampus of patients with mild cognitive impairment the ratio DM-SV/IG20 was increased, whereas in the hippocampus of cases with advanced AD the ratio was reversed (Mo et al., 2012). Thus, the peak in DM-SV/IG20 might be protective to A β . However, it decreases during further gradual A β accumulation. It was, therefore, concluded that the DM-SV might be protective against A β neurotoxicity and might reduce the cell death, whereas the IG20 may contribute to selective neuronal vulnerability in AD (Mo et al., 2012).

Diazepam-Binding Inhibitor (DBI)

The DBI is an acyl-CoA binding protein that is involved in fatty acid metabolism and steroidogenesis. It is also implicated in inflammation and apoptosis (Zavala, 1997; Mills et al., 2013). The DBI gene was stimulated by A β in cultured astrocytes (Tokay et al., 2008) and was 3-fold up-regulated in the parietal cortex of AD patients (Mills et al., 2013). Three alternatively spliced isoforms of DBI were identified: DBI-003, DBI-008, DBI-009. DBI-003 is a protein – coding isoform with 4 exons generated by the use of the TSS-B promoter. DBI-008 and DBI-009 are non-coding variants controlled by the different promoter TSS-A. DBI-008 has 3 exons and keeps intronic sequences at the 5' termini of the first and third exons. DBI-009 contains 2

exons with retained introns. DBI-009 was 30-fold increased, DBI-003 was dramatically down-regulated and DBI-008 was not changed in the parietal cortex of AD patients (Mills et al., 2013). Based on experimental evidence (reviewed in Mills et al., 2013) it can be suggested that enhanced expression of the non-coding intron-bearing isoform DBI-009 may affect firing properties of neurons and may be implicated in inflammatory processes in which astrocytes are involved. However, the precise role of the DBI-009 variant in AD pathogenesis remains to be elucidated.

CONCLUSION

This chapter reveals two major ideas concerning alternative splicing in AD. First of all, alternative splicing in the AD affected brain areas is dysregulated. Some genes show increased alternative splicing, others diminished. Certain isoforms of the same mRNA can be elevated, others did not change or even decreased. These data lead to the conclusion that AD is a disease characterized by dysregulated splicing. In support of this idea, Illumina RNA-Seq analysis revealed elevated transcriptome activity in the parietal cortex of AD patients and showed that 4073 alternatively spliced isoforms were up-regulated while 558 were down-regulated (Mills et al., 2013). Misregulation of alternative splicing in AD may result from mutations in the genes themselves or alterations in the regulators of splicing. Mitochondrial damage and the related hypoxic stress are important factors influencing alternative splicing changes in AD (Maracchioni et al., 2007). Second, the extent and severity of molecular defects in alternatively spliced isoforms is generally greater in AD patients. Very often unexpected (“cryptic”) splice sites are present inside classical exons or even within the introns. Together, the data show that alternative splicing and its regulators may be attractive targets for the development of novel treatment options for AD.

REFERENCES

- Annie, M; Stefani, M; Hueber, A; Fischer, F; Paganetti, P. Splicing of intron 3 of human BACE requires the flanking introns 2 and 4. *Biochem Biophys Res Commun*, 2009, 388, 434-438.
- Ansaloni, S; Leung, BP; Sebastian, NP; Samudralwar, R; Gadaleta, M; Saunders, AJ. TrkB isoforms differentially affect AICD production through their intracellular functional domains. *Int J Alzheimers Dis*, 2011, 729382.
- Aoyama, M; Asai, K; Shishikura, T; Kawamoto, T; Miyachi, T; Yokoi, T; Togari, H; Wada, Y; Kato, T; Nakagawara, A. Human neuroblastomas with unfavorable biologies express high levels of brain-derived neurotrophic factor mRNA and a variety of its variants. *Cancer Lett*, 2001, 164, 51-60.
- Bailey, M; Wang, ACJ; Hao, J; Janssen, WGM; Hara, Y; Dumitriu, D; Hof, PR; Morrison, JH. Interactive effects of age and estrogen on cortical neurons: implications for cognitive aging. *Neuroscience*, 2011, 191, 148-158.
- Barrachina, M; Dalfo, E; Puig, B; Vidal, N; Freixes, M; Castano, E; Ferrer, I. Amyloid- β deposition in the cerebral cortex in dementia with Lewy body is accompanied by a relative

- increase in A β PP mRNA isoforms containing the Kunitz protease inhibitor. *Neurochem Int*, 2005, 46, 253-260.
- Barron, AM; Pike, CJ. Sex hormones, aging, and Alzheimer's disease. *Front Biosci*, 2012, 4, 976-997.
- Beart, PM; O'Shea, RD. Transporters for L-glutamate: an update on their molecular pharmacology and pathological involvement. *Br J Pharmacol*, 2007, 150, 5-17.
- Berson, A; Knoblock, M; Hanan, M; Diamant, S; Sharoni, M; Schuppli, D; Geyer, BC; Ravid, R; Mor, TS; Nitsch, RM; Soreq, H. Changes in readthrough acetylcholinesterase expression modulate amyloid-beta pathology. *Brain*, 2008, 131, 109-119.
- Bertram, L; Hiltunen, M; Parkinson, M; Ingelsson, M; Lange, C; Ramasamy, K; Mullin, K; Menon, R; Sampson, AJ; Hsiao, MY; Elliott, KJ; Velicelebi, G; Moscarillo, T; Hyman, BT; Wagner, SL; Becker, KD; Blacker, D; Tanzi, RE. Family-based association between Alzheimer's disease and variants in UBQLN1. *N Engl J Med*, 2005, 352, 884-894.
- Beyreuther, K; Pollwein, P; Multhaup, G; Monnig, U; Konig, G; Dyrks, T; Schubert, W; Masters, CL. Regulation and expression of the Alzheimer's beta/A4 amyloid protein precursor in health, disease, and Down's syndrome. *Ann NY Acad Sci*, 1993, 695, 91-102.
- Bollig, A; Mikcicek, RJ. An estrogen receptor- α splicing variant mediates both positive and negative effects on gene transcription. *Mol Endocrinol*, 2000, 14, 634-649.
- Braak, H; Braak, E. Neuropathological staging of Alzheimer-related changes. *Acta Neuropathol*, 1991, 82, 239-259.
- Brinton, RD. Investigative models for determining hormone therapy-induced outcomes in brain: evidence in support of a healthy cell bias of estrogen action. *Ann NY Acad Sci*, 2005, 1052, 57-74.
- Buee, L; Bussiere, T; Buee-Scherrer, V; Delacourte, A; Hof, PR. Tau protein isoforms, phosphorylation and role in neurodegenerative disorders. *Brain Res Rev*, 2000, 33, 95-130.
- Capell, A; Kaether, C; Edbauer, D; Shirotani, K; Merkel, S; Steiner, H; Haas, C. Nicastrin interacts with gamma-secretase complex components via the N-terminal part of its transmembrane domain. *J Biol Chem*, 2003, 278, 52519-52523.
- Citron, BA; Suo, Z; SantaCruz, K; Davies, PJA; Qin, F; Festoff, BW. Protein crosslinking, tissue transglutaminase, alternative splicing and neurodegeneration. *Neurochem Int*, 2002, 40, 69-78.
- Citron, BA; SantaCruz, KS; Davies, PJA. Festoff, BW. Intron-exon swapping of transglutaminase mRNA and neuronal tau aggregation in Alzheimer's disease. *J Biol Chem*, 2001, 276, 3295-3301.
- Clatworthy, AE; Stockinger, W; Christie, RH; Schneider, WJ; Nimpf, J; Hyman, BT; Rebeck, GW. Expression and altered splicing of apolipoprotein E receptor 2 in brain. *Neuroscience*, 1999, 90, 903-911.
- Collet, J; Fehrat, L; Pollard, H; Ribas de Pouplana, L; Charton, G; Bernard, A; Moreau, J; Ben-Ari, Y; Khrestchatsky, M. Developmentally regulated alternative splicing of mRNAs encoding N-terminal tau variants in the rat hippocampus: structural and functional implications. *Eur J Neurosci*, 1997, 9, 2723-2733.
- Connor, B; Young, D; Yan, Q; Faull, R.L; Synek, B; Dragunow, M. Brain-derived neurotrophic factor is reduced in Alzheimer's disease. *Mol Brain Res*, 1997, 49, 71-81.
- Conrad, C; Zhu, J; Conrad, C; Schoenfeld, D; Fang, Z; Ingelsson, M; Stamm, S; Church, G; Hyman, BT. Single molecule profiling of tau gene expression in Alzheimer's disease. *J Neurochem*, 2007, 103, 1228-1236.

- Cotman, CW; Su, J.H. Mechanisms of neuronal death in Alzheimer's disease. *Brain Pathol*, 1996, 6, 493-506.
- Craig, MC; Murphy, DGM. Alzheimer's disease in women. *Best Pract Res Clin Obst Gynaecol*, 2009, 23, 53-61.
- Cummings, JL; Back, C. The cholinergic hypothesis of neuropsychiatric symptoms in Alzheimer's disease. *Am J Geriatr Psychiatry*, 1998, 6, S64-68.
- Darreh-Shori, T; Hellstrom-Lindahl, E; Flores-Flores, C; Guan, ZZ; Soreq, H; Nordberg, A. Long-lasting acetylcholinesterase splice variations in anticholinesterase-treated Alzheimer's disease patients. *J Neurochem*, 2004, 88, 1102-1113.
- De Jonghe, C; Cruts, M; Rogaeva, EA; Tysoe, C; Singleton, A; Vanderstichele, H; Meschino, W; Dermaut, B; Vanderhoeven, I; Backhovens, H; Vanmechelen, E; Morris, CM; Hardy, J; Rubinsztein, DC; St George-Hyslop, PH; Van Broeckhoven, C. Aberrant splicing in the presenilin-1 intron 4 mutation causes presenile Alzheimer's disease by increased Abeta42 secretion. *Hum Mol Genet*, 1999, 8, 1529-1540.
- Del Villar, K; Miller, CA. Down-regulation of DENN/MADD, a TNF receptor/binding protein, correlates with neuronal cell death in Alzheimer's disease brain and hippocampal neurons. *Proc Natl Acad Sci USA*, 2004, 101, 4210-4215.
- Delatour, B; Mercken, L; El Hachimi, KH; Colle, MA; Pradier, L; Duyckaerts, C. Fe65 in Alzheimer's disease: neuronal distribution and association with neurofibrillary tangles. *Am J Pathol*, 2001, 158, 1585-1591.
- Dubelaar, EJ; Verwer, R.W; Hofman, MA; Van Heerikhuijsen, JJ; Ravid, R; Swaab, DF. ApoE epsilon4 genotype is accompanied by lower metabolic activity in nucleus basalis of Meynert neurons in Alzheimer patients and controls as indicated by the size of the Golgi apparatus. *J Neuropathol Exp Neurol*, 2004, 63, 159-169.
- Duyckaerts, C; Delatour, B; Potier, M-C. Classification and basic pathology of Alzheimer disease. *Acta Neuropathol*, 2009, 118, 5-36.
- Efimova, EV; Al-Zoubi, AM; Martinez, O; Kaithamana, S; Lu, S; Arima, T; Prabhakar, BS. IG20, in contrast to DENN-SV, IMADD splice variants suppresses tumor cell survival, and enhances their susceptibility to apoptosis and cancer drugs. *Oncogene*, 2004, 23, 1076-1087.
- Ermak, G; Morgan, TE; Davies, KJ. Chronic overexpression of the calcineurin inhibitory gene DSCR1 (Adapt 78) is associated with Alzheimer's disease. *J Biol Chem*, 2001, 276, 38787-38794.
- Francis, PT. Glutamatergic systems in Alzheimer's disease. *Int J Geriatr Psychiatry*, 2003, 18, (Suppl1), S15-S21.
- Fukumoto, H; Cheung, BS; Hyman, BT; Irizarry, MC. Beta-secretase protein and activity are increased in the neocortex in Alzheimer disease.
- Garcia-Pedrero, JM; Zuazua, P; Martinez-Campa, C; Lazo, PS; Ramos, S. The naturally occurring variant of estrogen receptor (ER) ERΔE7 suppresses estrogen-dependent transcriptional activation by both wild type ERα and ERβ. *Endocrinology*, 2003, 144, 2967-2976.
- Garzon, D; Yu, G; Fahnstock, M. A new brain-derived neurotrophic factor transcript and decrease in brain-derived neurotrophic factor transcripts 1, 2 and 3 in Alzheimer's disease parietal cortex. *J Neurochem*, 2002, 82, 1058-1064.

- Ghersi, E; Noviello, C; D'Adamio, LD. Amyloid- β protein precursor (A β PP) intracellular domain-associated protein-1 proteins bind to A β PP and modulate its processing in an isoform-specific manner. *J Biol Chem*, 2004, 279, 49105-49112.
- Giannakopoulos, P; Kovari, E; French, LE; Viard, I; Hof, PR; Bouras, C. Possible neuroprotective role of clusterin in Alzheimer's disease: a quantitative immunocytochemical study. *Acta Neuropathol*, 1998, 95, 387-394.
- Glatz, DC; Rujescu, D; Tan, Y; Berendt, F.J; Hartmann, AM; Faltraco, F; Rosenberg, C; Hulette, C; Jellinger, K; Hampell, H; Riederer, P; Moller, H-J; Andreadis, A; Henkel, K; Stamm, S. The alternative splicing of tau exon 10 and its regulatory proteins CLK2 and TRA2-BETA1 changes in sporadic Alzheimer's disease. *J Neurochem*, 2006, 96, 635-644.
- Goedert, M; Spillantini, MG; Jakes, R; Rutherford, D; Crowther, RA. Multiple isoforms of human microtubule-associated protein tau: sequences and localization in neurofibrillary tangles of Alzheimer's disease. *Neuron*, 1989, 3, 519-526.
- Greengard, P; Valtorta, F; Czernik, AJ; Benfenati, F. Synaptic vesicle phosphoproteins and regulation of synaptic function. *Science*, 1993, 259, 780-785.
- Grisaru, D; Sternfeld, M; Eldor, A; Glick, D; Soreq, H. Structural roles of acetylcholinesterase variants in biology and pathology. *Eur J Biochem*, 1999, 264, 672-686.
- Hardy, J; Selkoe, DJ. The amyloid hypothesis of Alzheimer's disease: progress and problems on the road to therapeutics. *Science*, 2002, 297, 353-356.
- Harris, CD; Ermak, G; Davies, KJA. RCAN1-1L is overexpressed in neurons of Alzheimer's disease patients. *FEBS J*, 2007, 274, 1715-1724.
- Heinzen, EL; Yoon, W; Weale, ME; Sen, A; Wood, NW; Burke, JR; Welsh-Bohmer, KA; Hulette, CM; Sisodiya, SM; Goldstein, DB. Alternative ion channel splicing in mesial temporal lobe epilepsy and Alzheimer's disease. *Genome Biology*, 2007, 8, R32.
- Ho, L; Guo, Y; Spielman, L; Petrescu, O; Haroutunian, V; Purohit, D; Czernik, A; Yemul, S; Aisen, PS; Mohs, R; Pasinetti, GM. Altered expression of a-type, but not b-type synapsin isoform in the brain of patients at high risk for Alzheimer's disease assessed by DNA microarray technique. *Neurosci Lett*, 2001, 298, 191-194.
- Hock, C; Heese, K; Hulette, C; Rosenberg, C; Otten, U. Region-specific neurotrophin imbalances in Alzheimer's disease: decreased levels of brain-derived neurotrophic factor and increased levels of nerve growth factor in hippocampus and cortical areas. *Arch Neurol*, 2000, 57, 846-851.
- Hol, EM; Roelofs, RF; Moraal, E; Sonnemans, MAF; Sluijs, JA; Proper, EA; de Graan, PNE; Fischer, DF; van Leeuwen, FW. Neuronal expression of GFAP in patients with Alzheimer pathology and identification of novel GFAP splice forms. *Mol Psychiatry*, 2003, 8, 786-796.
- Holsinger, RM; Schnarr, J; Henry, P; Castelo, VT; Fahnestock, M. Quantitation of BDNF mRNA in human parietal cortex by competitive reverse transcription-polymerase chain reaction: decreased levels in Alzheimer's disease. *Mol Brain Res*, 2000, 76, 347-354.
- Hu, Q; Cool, BH; Wang, B; Hearn, MG; Martin, GM. A candidate molecular mechanism for the association of an intronic polymorphism of FE65 with resistance to very late onset dementia of the Alzheimer type. *Hum Mol Genet*, 2002, 11, 465-475.
- Hu, Q; Jin, LW; Starbuck, MY; Martin, GM. Broadly altered expression of the mRNA isoforms of FE65, a facilitator of β amyloidogenesis, in Alzheimer cerebellum and other brain regions. *J Neurosci Res*, 2000, 60, 73-86.

- Hynd, MR; Scott, HL; Dodd, PR. Glutamate-mediated excitotoxicity and neurodegeneration in Alzheimer's disease. *Neurochem Int*, 2004, 45, 583-595.
- Ikura, K; Takahata, K; Sasaki, R. Cross-linking of a synthetic partial-length (1-28) peptide of the Alzheimer/A4 amyloid protein by transglutaminase. *FEBS Lett*, 1993, 326, 109-111.
- Inestrosa, NC; Alvarez, A; Perez, CA; Moreno, RD; Vicente, M; Linker, C; Casanueva, OI; Soto, C; Garrido, J. Acetylcholinesterase accelerates assembly of amyloid-beta-peptides into Alzheimer's fibrils: possible role of the peripheral site of enzyme. *Neuron*, 1996, 16, 881-891.
- Ingelsson, M; Ramasamy, K; Russ, C; Freeman, SH; Orne, J; Raju, S; Matsui, T; Growdon, JH; Frosch, MP; Ghetti, B; Brown, RH; Irizarry, MC; Hyman, BT. Increase in the relative expression of tau with four microtubule binding repeat regions in frontotemporal lobar degeneration and progressive supranuclear palsy brains. *Acta Neuropathol*, 2007, 114, 471-479.
- Ingelsson, M; Ramasamy, K; Cantuti-Castelvetri, I; Skoglund, L; Matsui, T; Orne, J; Kowa, H; Raju, S; Vanderburg, CR; Augustinack, JC; de Silva, R; Lees, AJ; Lannfelt, L; Growdon, JH; Frosch, MP; Standaert, DG; Irizarry, MC; Hyman, BT. No alteration on tau exon 10 alternative splicing in tangle-bearing neurons of the Alzheimer's disease brain. *Acta Neuropathol*, 2006, 112, 439-449.
- Ishunina, TA; Kamphorst, W; Swaab, DF. Changes in metabolic activity and estrogen receptors in the human medial mammillary nucleus: relation to sex, aging and Alzheimer's disease. *Neurobiol Aging*, 2003, 24, 817-828.
- Ishunina, TA; Fischer, DF; Swaab, DF. Estrogen receptor α and its splice variants in the hippocampus in aging and Alzheimer's disease. *Neurobiol Aging*, 2007; 28, 1670-1681.
- Ishunina, TA; Swaab, DF. Decreased alternative splicing of estrogen receptor- α mRNA in the Alzheimer's disease brain. *Neurobiol Aging*, 2012, 33, 286-296.
- Ishunina, TA; Sluiter, AA; Swaab, DF; Verwer, RWH. Transcriptional activity of human brain estrogen receptor- α splice variants: evidence for cell type-specific regulation. *Brain Res*, 2013, 1500, 1-9.
- Ishunina, TA; Swaab, DF. Hippocampal estrogen receptor- α splice variant TADDI in the human brain in aging and Alzheimer's disease. *Neuroendocrinology* 2009; 89: 187-199.
- Ishunina TA; Swaab DF. Increased expression of estrogen receptor α and β in the nucleus basalis of Meynert in Alzheimer's disease. *Neurobiol Aging*, 2001, 22, 417-426.
- Isoe-Wada, K; Urakami, K; Wakutani, Y; Adachi, Y; Arai, H; Sasaki, H; Nakashima, K. Alteration in brain presenilin-1 mRNA expression in sporadic Alzheimer's disease. *Eur J Neurol*, 1999, 6, 163-167.
- Jacobsen, JS; Blume, AJ; Vitek, MP. Quantitative measurement of alternatively spliced amyloid precursor protein mRNA expression in Alzheimer's disease and normal brain by S1 nuclease protection analysis. *Neurobiol Aging*, 1991, 12, 585-592.
- Kim, SY; Grant, P; Lee, JH; Pant, HC. Steinert, P.M. Differential expression of multiple transglutaminases in human brain. Increased expression and cross-linking by transglutaminases 1 and 2 in Alzheimer's disease. *J Biol Chem*, 1999, 274, 30715-30721.
- Konig, G; Beyreuther, K; Masters, CL; Schmitt, HP; Salbaum, JM. PreA4 mRNA distribution in brain areas. *Prog Clin Biol Res*, 1989, 317, 1027-1036.
- LaPointe, NE; Horowitz, PM; Guillozet-Bongaarts, AL; Silva, A; Andreadis, A; Binder, LI. Tau 6D and 6P isoforms inhibit polymerization of full-length tau in vitro. *Biochemistry*, 2009, 48, 12290-12297.

- Lee, SF; Shah, S; Li, H; Yu, C; Han, W; Yu, G. Mammalian APH-1 interacts with presenilin and nicastrin and is required for intramembrane proteolysis of amyloid-beta precursor protein and notch. *J Biol Chem*, 2002, 277, 45013-45019.
- Leskov, KS; Klovov, DY; Li, J; Kinsella, TG; Boothman, DA. Synthesis and functional analyses of nuclear clusterin, a cell death protein. *J Biol Chem*, 2003, 278, 11590-11600.
- Li, S; Mallory, M; Alford, M; Tanaka, S; Masliah, E. Glutamate transporter alterations in Alzheimer disease are possibly associated with abnormal APP expression. *J Neuropathol Exp Neurol*, 1997, 56, 901-911.
- Lim, KM; Chow, VT. Induction of marked apoptosis in mammalian cancer cell lines by antisense DNA treatment to abolish extension of DENN (differentially expressed in normal and neoplastic cells). *Mol Carcinog*, 2002, 35, 110-126.
- Luo, MH; Tse, SW; Memmott, J; Andreadis, A. Novel isoforms of tau that lack the microtubule-binding domain. *J Neurochem*, 2004, 90, 340-351.
- Mackeprang Dahmcke, C; Buchmann-Moller, S; Jensen, NA; Mitchelmore, C. Altered splicing in exon 8 of the DNA replication factor CIZ1 affects subnuclear distribution and is associated with Alzheimer's disease. *Mol Cell Neurosci*, 2008, 38, 589-594.
- Mah, AL; Perry, G; Smith, MA; Monteiro, MJ. Identification of ubiquilin, a novel presenilin interactor that increases presenilin protein accumulation. *J Cell Biol*, 2000, 151, 847-862.
- Mandell, JW; Czernik, AJ; De Camilli, P; Greengard, P; Townes-Anderson, E. Differential expression of synapsins I and II among rat retinal synapses. *J Neurosci*, 1992, 12, 1736-1749.
- Maracchioni, A; Totaro, A; Angelini, DF; DiPenta, A; Bernardi, G; Carri, MT; Achsel, T. Mitochondrial damage modulates alternative splicing in neuronal cells: implications for neurodegeneration. *J Neurochem*, 2007, 100, 142-153.
- Masliah, E; Alford, M; Deteresa, R; Mallory, M; Hansen, L. Deficient glutamate transport is associated with neurodegeneration in Alzheimer's disease. *Ann Neurol*, 1996, 40, 759-766.
- Mills, JD; Janitz, M. Alternative splicing of mRNA in the molecular pathology of neurodegenerative diseases. *Neurobiol Aging*, 2012, 33, 1012.e11-1012.e24.
- Mills, JD; Nalpathamkalan, T; Jacobs, HIL; Janitz, C; Merico, D; Hu, P; Janitz, M. RNA-Seq analysis of the parietal cortex in Alzheimer's disease reveals alternatively spliced isoforms related to lipid metabolism. *Neurosci Lett*, 2013, 536, 90-95.
- Mitsuda, M; Yamagata, HD; Zhong, W; Aoto, M; Akatsu, H; Uekawa, N; Kamino, K; Taguchi, K; Yamamoto, T; Mazuyama, M; Kosaka, K; Takeda, M; Kondo, I; Miki, T. A novel alternative splice variant of nicastrin and its implication in Alzheimer disease. *Life Sci*, 2006, 78, 2444-2448.
- Mo, Y; Williams, C; Miller, CA. DENN/MADD/IG20 alternative splicing changes and cell death in Alzheimer's disease. *J Mol Neurosci*, 2012, 48, 97-110.
- Morihara, T; Katayama, T; Sato, N; Yoneda, T; Manabe, T; Hitomi, J; Abe, H; Imaizumi, K; Tohyama, M. Absence of endoproteolysis but no effects on amyloid beta production by alternative splicing forms of presenilin-1, which lack exon 8 and replace D257A. *Brain Res Mol Brain Res*, 2000, 85, 85-90.
- Mowrer, KR; Wolfe, MS. Promotion of BACE1 mRNA alternative splicing reduces amyloid β -peptide production. *J Biol Chem*, 2008, 283, 18694-18701.
- Naftolin, F; Malaspina, D. Estrogen, estrogen treatment and the post-reproductive woman's brain. *Maturitas*, 2007, 57, 23-26.

- Nagy, Z. Cell cycle regulatory failure in neurones: causes and consequences. *Neurobiol Aging*, 2000, 21, 761-769.
- Niblock, M; Gallo, JM. Tau alternative splicing in familial and sporadic tauopathies. *Biochem Soc Trans*, 2012, 40, 677-680.
- Niwa, S; Tanaka, Y; Hirokawa, N. KIF1B β - and KIF1A-mediated axonal transport of presynaptic regulator Rab3 occurs in a GTP-dependent manner through DENN/MADD. *Nat Cell Biol*, 2008, 10, 1269-1279.
- Nuutinen, T; Suuronen, T; Kauppinen, A; Salminen, A. Clusterin: a forgotten player in Alzheimer's disease. *Brain Res Rev*, 2009, 61, 89-104.
- Ohno, M; Cole, SL; Yasvoino, M; Zhao, J; Citron, M; Berry, R; Disterhoft, JF; Vassar, R. BACE 1 gene deletion prevents neuron loss and memory deficits in 5xFAD APP/PS1 transgenic mice. *Neurobiol Dis*, 2007, 26, 134-145.
- Patapoutian, A; Reichardt, LF. Trk receptors: mediators of neurotrophin action. *Curr Opin Neurobiol*, 2001, 11, 272-280.
- Penna, I; Vassallo, I; Nizzari, M; Russo, D; Costa, D; Menichini, P; Poggi, A; Russo, C; Dieci, G; Florio, T; Cancedda, R; Pagano, A. A novel snRNA-like transcript affects amyloidogenesis and cell cycle progression through perturbation of Fe65L1/APBB2 alternative splicing. *Biochim Biophys Acta*, 2013, 1833, 1511-1526.
- Phillips, HS; Hains, JM; Armanini, M; Laramie, GR; Johnson, SA; Winslow, J.W. BDNF mRNA is decreased in the hippocampus of individuals with Alzheimer's disease. *Neuron*, 1991, 7, 695-702.
- Pow, DV; Cook, DC. Neuronal expression of splice variants of "glial" glutamate transporters in brains afflicted by Alzheimer's disease: unmasking an intrinsic neuronal property. *Neurochem Res*, 2009, 34, 1748-1757.
- Rockenstein, EM; McConlogue, L; Tan, H; Power, M; Masliah, E; Mucke, L. Levels and alternative splicing of amyloid β protein precursor (APP) transcripts in brains of APP transgenic mice and humans with Alzheimer's disease. *J Biol Chem*, 1995, 270, 28257-28267.
- Sabo, SL; Lanier, LM; Ikin, AF; Khorkova, O; Sahasrabudhe, S; Greengard, P; Buxbaum, JD. Regulation of β -amyloid secretion by FE65, an amyloid protein precursor-binding protein. *J Biol Chem*, 1999, 274, 7952-7957.
- Saito, S; Takahashi-Sasaki, N; Araki, W. Identification and characterization of a novel human APH-1b splice variant lacking exon 4. *Biochem Biophys Res Commun*, 2005, 330, 1068-1072.
- Saito, S; Araki, W. Expression profiles of two human APH-1 genes and their roles in formation of presenilin complexes. *Biochem Biophys Res Commun*, 2005, 327, 18-22.
- Sandbrink, R; Monning, U; Masters, CL; Beyreuther, K. Expression of the APP gene family in brain cells, brain development and aging. *Gerontology*, 1997, 43, 119-131.
- Sarasa, M; Sorribas, V; Terrado, J; Climent, S; Palacios, JM; Mengod, G. Alzheimer β -amyloid precursor proteins display specific patterns of expression during embryogenesis. *Mech Dev*, 2000, 94, 233-236.
- Sato, N; Hori, O; Yamaguchi, A; Lambert, JC; Chartier-Harlin, MC; Robinson, PA; Delacourte, A; Schmidt, AM; Furuyama, T; Imaizumi, K; Tohyama, M; Takagi, T. A novel presenilin-2 splice variant in human Alzheimer's disease brain tissue. *J Neurochem*, 1999, 72, 2498-2505.

- Sato, N; Imaizumi, K; Manabe, T; Taniguchi, M; Hitomi, J; Katayama, T; Yoneda, T; Morihara, T; Yasuda, Y; Takagi, T; Kudo, T; Tsuda, T; Itoyama, Y; Makifuchi, T; Fraser, PE; St. George-Hyslop, P; Tohyama, M. Increased production of β -amyloid and vulnerability to endoplasmic reticulum stress by an aberrant spliced form of presenilin 2. *J Biol Chem*, 2001, 276, 2108-2114.
- Scott, HA; Gebhardt, FM; Mitrovic, AD; Vandenberg, RJ; Dodd, PR. Glutamate transporter variants reduce glutamate uptake in Alzheimer's disease. *Neurobiol Aging*, 2011, 32, 553.e1-11.
- Shirotani, K; Edbauer, D; Prokop, S; Haass, C; Steiner, H. Identification of distinct γ -secretase complexes with different APH-1 variants. *J Biol Chem*, 2004, 279, 41340-41345.
- Smith, M; Sharples, RA; Evin, G; McLean, CA; Dean, B; Pavey, G; Fantino, E; Cotton, RGH; Imaizumi, K; Masters, CL; Cappai, R; Culvenor, JG. Expression of truncated presenilin 2 splice variant in Alzheimer's disease, bipolar disorder, and schizophrenia brain cortex. *Mol Brain Res*, 2004, 127, 128-135.
- Stoilov, P; Castren, E; Stamm, S. Analysis of the human TrkB gene genomic organization reveals novel TrkB isoforms, unusual gene length, and splicing mechanism. *Biochem Biophys Res Commun*, 2002, 290, 1054-1065.
- Swaab, DF. The human hypothalamus: basic and clinical aspects. II. Neuropathology of the human hypothalamus and adjacent structures; in Aminoff MJ; Boller F; Swaab DF. (eds.): *Handbook of Clinical Neurology*. Amsterdam, Elsevier, 2004, vol. 80, 3rd series, vol.2, p.632.
- Tanahashi, H; Tabira, T. Three novel alternatively spliced isoforms of the human beta-site amyloid precursor protein cleaving enzyme (BACE) and their effect on amyloid beta-peptide production. *Neurosci Lett*, 2001, 307, 9-12.
- Tanahashi, H; Tabira, T. Characterization of an amyloid precursor protein-binding protein Fe65L2 and its novel isoforms lacking phosphotyrosine-interaction domains. *Biochem J*, 2002, 367, 687-695.
- Tanaka, S; Liu, L; Kimura, J; Shiojiri, S; Takahashi, Y; Kitaguchi, N; Nakamura, S; Ueda, K. Age-related changes in the proportion of amyloid precursor protein mRNAs in Alzheimer's disease and other neurological disorders. *Brain Res Mol Brain Res*, 1992, 15, 303-310.
- Thakur, MK; Mani, S.T. Estradiol regulates APP mRNA alternative splicing in the mice brain cortex. *Neurosci Lett*, 2005, 381, 154-157.
- Thal, D.R. Excitatory amino acid transporter EAAT-2 in tangle-bearing neurons in Alzheimer's disease. *Brain Pathol*, 2002, 12, 405-411.
- Tokay, T; Hachem, R; Masmoudi-Kouki, O; Gandolfo, P; Desrues, L; Leprince, J; Castel, H; Diallo, M; Amri, M; Vaudry, H; Tonon, MC. Beta-amyloid peptide stimulates endozepine release in cultured rat astrocytes through activation of N-formyl peptide receptors. *Glia*, 2008, 56, 1380-1389.
- Twine, NA; Janitz, K; Wilkins, MR; Janitz, M. Whole transcriptome sequencing reveals gene expression and splicing differences in brain regions affected by Alzheimer's disease. *PLoS ONE*, 2011, 6, e16266.
- Uversky, V. Intrinsic disorder in proteins associated with neurodegenerative diseases. *Front Biosci*, 2009, 14, 5188-5238.
- Wang, JZ; Liu, F. Microtubule-associated protein tau in development, degeneration and protection of neurons. *Prog Neurobiol*, 2008, 85, 148-175.
- Wolfe, MS. Tau mutations in neurodegenerative diseases. *J Biol Chem*, 2009, 284, 6021-6025.

- Wong, J; Weickert, CS. Transcriptional interaction of an estrogen receptor splice variant and ErbB4 suggests convergence in gene susceptibility pathways in schizophrenia. *J Biol Chem*, 2009, 284, 18824-18832.
- Wong, J; Higgins, M; Halliday, G; Garner, B. Amyloid beta selectively modulates neuronal TrkB alternative transcript expression with implications for Alzheimer's disease. *Neuroscience*, 2012, 210, 363-374.
- Xu, PT; Gilbert, JR; Qiu, HL; Ervin, J; Rothrock-Christian, TR; Hulette, C; Schmechel, DE. Specific regional transcription of apolipoprotein E in human brain neurons. *Am J Pathol*, 1999, 154, 601-611.
- Zandi, PP; Carlson, MC; Plassman, BL; Welsh-Bohmer, KA; Mayer, LS; Steffens, DC; Breitner, JC. Cache County Memory Study Investigators. Hormone replacement therapy and incidence of Alzheimer disease in older women: the Cache County study. *JAMA*, 2002, 288, 2123-2129.
- Zavala, V. Peripheral benzodiazepine receptors and the cytokine network: studies in the immune system and in the brain. *Biol Psychiatry*, 1997, 42, 82S-83S.
- Zhang, W; Johnson, BR; Suri, DE; Martinez, J; Bjornsson, T.D. Immunohistochemical demonstration of tissue transglutaminase in amyloid plaques. *Acta Neuropathol*, 1998, 96, 395-400.
- Zhang, Y; Zhou, L; Miller, CA. A splicing variant of a death domain protein that is regulated by a mitogen-activated kinase is a substrate for c-Jun N-terminal kinase in the human central nervous system. *Proc Natl Acad Sci USA*, 1998, 95, 2586-2591.
- Zohar, O; Pick, CG; Cavallaro, S; Chapman, J; Katzav, A; Milman, A; Alkon, DL. Age-dependent differential expression of BACE splice variants in brain regions of tg2576 mice. *Neurobiol Aging*, 2005, 26, 1167-117.

Chapter 20

PLANT RNA-BINDING PROTEINS IMPLICATE MRNA PROCESSING IN ABSCISIC ACID (ABA) RESPONSES

Raquel F. Carvalho, Vera S. Nunes and Paula Duque*

Instituto Gulbenkian de Ciência, Oeiras, Portugal

ABSTRACT

Posttranscriptional control of gene expression is crucial for biological processes. In particular, alternative splicing allows the same gene to produce multiple proteins and is thus a key generator of functional complexity. This posttranscriptional regulatory mechanism is a prevalent feature of eukaryotic genomes, being currently estimated to occur in over 50% of plant genes. RNA-binding proteins (RBPs) are known to control many aspects of RNA metabolism, from pre-mRNA splicing to the transport and stability of mRNA transcripts. The Arabidopsis and rice genomes contain about 200-250 genes predicted to encode RBPs, but few of these proteins have been characterized in plants.

As sessile organisms, plants are continuously exposed to environmental challenges that affect their growth and development. The phytohormone abscisic acid (ABA) is crucial in the coordination of plant responses to various abiotic stress factors. Interestingly, several RNA-metabolism genes have recently been implicated in the ABA pathway, providing a link between mRNA processing and ABA-mediated plant stress responses. Indeed, the loss- or gain-of-function of genes encoding different classes of RBPs directly involved in constitutive and alternative pre-mRNA splicing, such as snRNP factors or SR and hnRNP proteins, has been shown to result in striking ABA-response plant phenotypes. Functional roles in ABA biosynthesis or signaling have also been reported for other RBPs, including cap-binding proteins, RNA helicases, pentatricopeptide repeat proteins or poly(A) processing enzymes. Taken together, these data support the notion that posttranscriptional networks act as central coordinators of plant stress responses, namely by targeting key components of ABA signal transduction machinery. Future identification of the direct targets of these RBPs should uncover the molecular mechanisms underlying the mode of action of these proteins in the regulation of the ABA pathway.

* Corresponding Author's Email: duquep@igc.gulbenkian.pt.

1. INTRODUCTION

The regulation of gene expression is of crucial importance for biological processes in all living organisms. It can be modulated at different steps, from transcriptional initiation to mRNA processing and from there to the post-translational modification of a protein. Post-transcriptional gene regulation in eukaryotes involves several processes, including precursor-mRNA (pre-mRNA) splicing, capping and polyadenylation, or nucleocytoplasmic transport, stabilization and degradation of mRNA. RNA-binding proteins (RBPs) interact with single/double stranded RNA molecules in order to regulate these different cellular processes. Several RNA-binding motifs have been identified (reviewed in Burd and Dreyfuss, 1994; Lunde et al., 2007), with the most widely spread being the RNA-Recognition motif (RRM) followed by the K homology (KH) domain. Other RNA-binding motifs include glycine-rich, arginine-rich and RNA-binding domains, as well as zinc-fingers, DEAD boxes and RGG-boxes. The Arabidopsis genome contains more than 200 genes predicted to encode RBPs (Lorkovic and Barta, 2002; Lorkovic, 2009), whereas about 250 RBP genes have been identified in rice (Cook et al., 2010).

RNA splicing is the process that removes the noncoding introns from the pre-mRNA and joins the coding exons in order to obtain a mature functional mRNA that can later be translated into a protein. The splicing reaction is carried out by a large protein complex, the spliceosome, composed by five small nuclear ribonucleoproteins (snRNPs) and a large number of non-snRNP proteins. The efficient excision of introns from the pre-mRNA relies on specific consensus sequences such as splice sites located at the intron/exon junctions. However, these sequences are short and degenerate and usually not sufficient to determine the assembly of a functional spliceosome. Additional sequence information as well as protein interactions are necessary to activate their use. Indeed, apart from splice sites, other *cis*-elements within exons and introns are essential for the accuracy of splice site selection. These regulatory sequences, called exonic or intronic splicing enhancers (ESEs or ISEs) and silencers (ESSs or ISSs), are bound and activated by one or more of several related splicing factors, thus determining the strength of nearby splice sites.

Alternative splicing occurs when splice sites are differentially recognized, thus allowing for different rearrangements of a gene's coding fragments and generating multiple forms of mature mRNA from the same pre-mRNA molecule. In this way, a single gene may lead to the production of more than one polypeptide, greatly upscaling the genome coding capacity and providing a versatile means of regulating gene expression. In animals, alternative splicing has been found to determine such diverse biological processes as gender specification in *Drosophila* (Bell et al., 1991) or sound frequency recognition in the avian cochlea (Ramanathan et al., 1999) and in humans it is well known that its misregulation can lead to many important diseases (reviewed in Blencowe, 2006; Ward and Cooper, 2010). In fact, splice-site choice must be tightly regulated in time and space. Because splicing factors bind to numerous weakly conserved sequences, a single protein can regulate multiple target genes. In animal systems, a small number of proteins has been repeatedly found to be responsible for the regulation of a large number of alternative splicing events, and most known regulators of alternative splicing are ubiquitously expressed instead of tissue-specific (reviewed in Stamm et al., 2005; Nilsen and Graveley, 2010).

In the last decade great advances have been made regarding the analysis of alternative splicing, especially in plants, where studies were limited to individual genes. The most recent estimates, based on next-generation sequencing, indicate that at least 61% of the intron-containing genes in *Arabidopsis* (Marquez et al., 2012) as well as 48% and 55% of the rice (Lu et al., 2010) and barley (International Barley Genome Sequencing et al., 2012) genomes, respectively, are alternatively spliced. These figures are still lower than in humans, where 95% of the multi-exon genes are predicted to undergo alternative splicing (Pan et al., 2008; Wang et al., 2008), but are most likely to increase as more plant RNA-sequence data become available, particularly from various tissues, cell types and developmental stages or from plants grown under different environmental conditions.

Stress in plants can be defined as any external factor exerting a negative impact on their survival, growth or development. As sessile organisms, plants are continuously exposed to unfavourable environments that result in some degree of stress. The strategies they employ to respond to these adverse conditions can be highly complex, involving changes at the transcriptome, cellular and physiological levels. The phytohormone abscisic acid (ABA) plays key regulatory roles in many plant biological processes, from embryo maturation, seed development and germination to root growth, leaf abscission, fruit ripening and the regulation of stomatal apertures. Importantly, ABA action not only involves developmental and physiological regulation but is also crucial in the coordination of various stress signal transduction pathways, with a wide variety of abiotic factors, such as drought, desiccation, high salinity and cold or heat stress being known to trigger an increase in endogenous ABA levels (reviewed in Tuteja, 2007; Hong et al., 2013). Furthermore, genetic screens based on seed germination or on the use of reporter gene systems have allowed the isolation of a myriad of mutants affected in ABA responses.

In recent years, many RNA-metabolism genes have been described as components of ABA signaling pathways, strongly suggesting that posttranscriptional networks act as central coordinators of plant abiotic stress responses by targeting, for example, key components of the ABA signal transduction machinery. This chapter will review the available functional evidence linking splicing factors and other RBPs to the regulation of plant ABA stress responses (Table 1).

2. SPLICING FACTORS

2.1. snRNP Factors

snRNPs are found in the nucleus of eukaryotic cells and consist of one small RNA molecule (snRNA), a core of seven Sm proteins, which form a ring structure with a hole in the middle through which a U-rich sequence in the snRNA passes, and additional snRNP-specific proteins (reviewed in Valadkhan and Jaladat, 2010). An exception occurs for the U6 snRNA, which lacks an Sm site and therefore formation of the corresponding U6 snRNP involves the association of seven so-called Sm-like proteins (LSm). LSm proteins are structurally related to Sm proteins and modulate various aspects of RNA metabolism including splicing, nuclear export and degradation (Salgado-Garrido et al., 1999; reviewed in He and Parker, 2000; Valadkhan and Jaladat, 2010). In mammals and yeast, the LSm2-LSm8 heptameric complex

functions in U6 snRNP biogenesis for pre-mRNA splicing and therefore associates directly or indirectly with several splicing factors, whereas the LSM1-LSM7 complex has a role in mRNA degradation through decapping of mRNAs, leading to 5' to 3' exonucleolytic cleavage (Achsel et al., 1999; Bouveret et al., 2000; reviewed in He and Parker, 2000). Molecular and functional analyses of the Arabidopsis LSM gene family has shown that plant LSM proteins, similarly to those in yeast and animals, are organized in two different heptameric complexes, being essential for the correct turnover and splicing of selected development-related mRNAs and thus for normal Arabidopsis development (Perea-Resa et al., 2012).

In a genomewide gene expression study, Hoth et al. (2002) identified more than one thousand ABA-responsive genes, including novel ABA targets. More detailed analyses revealed several of these ABA-regulated genes to encode snRNP proteins, with Sm core genes as well as U5 and U4/U6 snRNP-specific protein genes being highly represented (Raab and Hoth, 2007). Among these genes, *AtPRP4/AtSAP60*, encoding a U4/U6 snRNP-specific protein, showed the strongest ABA regulation and was found to be important for seed development (Raab and Hoth, 2007).

In 2001, Xiong et al. described an Arabidopsis mutant, *sad1* (*super-sensitive to ABA and drought 1*), showing enhanced sensitivity to ABA and osmotic stress during seed germination and root growth, deficient drought-induced ABA biosynthesis, and altered expression of stress-response genes. The *SAD1* sequence encodes an LSM snRNP protein similar to LSM5 proteins from various organisms ranging from yeast to human (Xiong et al., 2001a). As the sole LSM5 ortholog present in the Arabidopsis genome, *SAD1* likely regulates different aspects of mRNA metabolism, such as splicing, export and degradation. Drought and ABA treatments trigger the upregulation of ABA biosynthesis genes, with ABA enhancing particularly those encoding the last two enzymes in the ABA biosynthesis pathway (Seo et al., 2000; Xiong et al., 2001a; Xiong et al., 2001b). This suggests the existence of a feedback regulatory mechanism whereby an initial increase in ABA levels induced by drought conditions may result in the speed up of the last steps of ABA biosynthesis (Xiong et al., 2001a). *SAD1* is likely to play a critical role in regulating this feedback loop as the *sad1* mutant shows a lower degree of transcript accumulation for ABA biosynthetic genes under both drought and ABA treatments.

The floral initiator Shk1 Kinase Binding Protein1 (SKB1), also named Protein Arginine Methylationtransferase5 (PRMT5), is known to play an important role in plant development and stress responses by controlling gene expression both at the transcriptional and posttranscriptional levels (Deng et al., 2010; Sanchez et al., 2010; Zhang et al., 2011). SKB1 appears to affect gene transcription by methylating histone4 arginine3 (H4R3) and to function posttranscriptionally by methylating the Sm-like U6 snRNP protein LSM4 (Zhang et al., 2011). Mutations in the *LSM4* gene lead to altered sensitivities to both ABA and salt stress during germination and seedling development, phenotypes also exhibited by *SKB1* loss-of-function mutants (Zhang et al., 2011). RNA-seq transcriptome analyses revealed that SKB1 deficiency causes splicing defects in hundreds of genes involved in multiple biological processes, including stress responses (Deng et al., 2010; Sanchez et al., 2010; Zhang et al., 2011). The *lsm4* mutant also shows splicing defects in several genes, many of which are also affected by the *skb1* mutation, indicating that methylation of LSM4 may regulate splice site selection (Zhang et al., 2011). Interestingly, methylated levels of LSM4 are low under normal conditions but increase upon salt or ABA stress, improving splicing efficiency of certain genes or altering splicing variants of others (Zhang et al., 2011).

Another snRNP-associated protein implicated in ABA responses is STA1 (Stabilized 1). The Arabidopsis *STA1* gene encodes an mRNA splicing factor homologous to the human U5 snRNP-associated 102-kDa protein and the yeast pre-mRNA splicing factors Prp1p and Prp6p. Indeed, *stabilized1* (*sta1*) mutant plants are defective in the splicing of the stress-induced *COR15A* gene, corroborating a role for STA1 in pre-mRNA splicing (Lee et al., 2006). Furthermore, these mutants are altered in their response to ABA, salt and cold stress, showing hypersensitivity to ABA during germination and root growth and being more affected by LiCl during root growth and more chilling-sensitive during plant development than wild-type plants (Lee et al., 2006).

2.2. SR and hnRNPs Proteins

The two major classes of RBPs shown to be active in alternative splicing are members of the serine/arginine-rich (SR) and of the heterogeneous nuclear ribonucleoprotein (hnRNP) protein families. SR proteins belong to a highly conserved family of spliceosomal proteins, which have been shown in animal systems to play vital roles in both constitutive and alternative splicing by influencing the most crucial steps of spliceosome assembly (Wu and Maniatis, 1993; Kohtz et al., 1994; Manley and Tacke, 1996; Shen and Green, 2004). These essential splicing regulators share a characteristic structural domain organization and contain one or two N-terminal RRM and a C-terminal arginine/serine-rich (RS) domain. On the other hand, the hnRNP family comprises a more diverse group of RBPs, which usually contain RRM or other functionally equivalent motifs such as K-homology (KH) domains. Given their ability to associate with RNA and single-stranded (ss)DNA, they play important roles in multiple aspects of nucleic acid metabolism (reviewed in Krecic and Swanson, 1999; Martinez-Contreras et al., 2007; Han et al., 2010). Among their functions in RNA metabolism, hnRNPs have been described to utilize a variety of strategies to control splice site selection in a manner that is important for both alternative and constitutive pre-mRNA splicing. Until recently, it was generally accepted that hnRNP and SR proteins typically acted as repressors and activators of splicing, respectively, but it is now becoming evident that regulatory factors act according to the context they are in. For example, Xue et al. (2009) have shown that the polypyrimidine tract-binding protein (PTB/hnRNPI), a well-characterized hnRNP splicing regulator in mammals, is able to either activate or repress selected splice sites depending on the position of their binding site and interactions with other factors.

Earlier reports on the functional characterization of SR proteins showed them to be important for normal plant development. Analyses of transgenic Arabidopsis overexpressor lines of two SR genes, *AtSR30* and *AtRS2Z33*, revealed severe developmental abnormalities (Lopato et al., 1999; Kalyna et al., 2003). Likewise, a knockout mutant for *SR45*, a plant-specific SR-related protein, which instead of one RS domain like SR proteins includes two, one on either side of the RRM, was reported to exhibit pleiotropic defects in developmental processes (Ali et al., 2007).

The involvement of SR and SR-like splicing factors in plant stress responses is beginning to emerge. *SR45* has been shown to negatively regulate glucose and ABA signaling during early seedling development (Carvalho et al., 2010). In the presence of exogenous glucose, the *sr45-1* mutant shows a drastic delay in the greening and expansion of cotyledons under light conditions as well as severe inhibition of hypocotyl elongation in dark grown seedlings.

Accordingly, the *sr45-1* mutation confers enhanced glucose-responsive gene expression. Mutant seedlings also show enhanced sensitivity to ABA during cotyledon development and display an overaccumulation of this phytohormone in response to glucose, which correlates with a marked induction of ABA biosynthesis gene expression. Importantly, recent phenotypical characterization of the first loss-of-function mutant for a plant SR protein gene has revealed that the *Arabidopsis* SCL30a negatively regulates ABA signaling to control salt and osmotic stress responses during seed germination (Carvalho et al., unpublished results). Therefore, although responding to distinct external signals (high sugars and osmotic/salt stress), both the SR45 and SCL30a splicing factors converge on negatively regulating the ABA pathway during early plant development. This may allow seed germination and seedling growth under moderate stress, thus promoting plant tolerance to unfavorable environmental conditions.

Members of the hnRNP family have also been implicated in ABA signaling. The guard-cell-localized *Vicia faba* AAPK (ABA-Activated Protein Kinase), which regulates plasma membrane ion channels and thereby stomatal aperture, interacts with and phosphorylates AKIP1 (AAPK-interacting protein 1), an RBP homologous to the human hnRNP A/B (Li et al., 2000; Li et al., 2002). Upon ABA treatment, AKIP1 is phosphorylated by AAPK, which allows it to bind the mRNA of dehydrin *DHN1*, a stress-regulated protein thought to promote stability of enzymes and membranes under stress conditions (Li et al., 2002). A role for the AKIP protein in mRNA processing is suggested by the observation that ABA induces the relocation of this protein into nuclear speckles (Li et al., 2002), interchromatin granule clusters and storage/recycling sites for splicing factors delivering them to nearby active sites of transcription (reviewed in Misteli et al., 1997; Lamond and Spector, 2003). Similarly, the *Arabidopsis thaliana* AKIP1 homolog UBA2a (poly(U)-Binding Associated protein) also reorganizes within the nucleus in structures resembling speckles in response to ABA (Riera et al., 2006). UBA2a is an interacting partner of another hnRNP protein, UBP1 (Poly(U)-Binding Protein 1), which is involved in stimulating pre-mRNA splicing (Lambermon et al., 2000). However, UBA2a has been shown to have no effect on splicing efficiency of reporter RNAs, functioning rather in the recognition of U-rich sequences in plant 3'-UTRs and contributing to the stabilization of mRNAs in the nucleus (Lambermon et al., 2002). Contrary to AKIP1, UBA2a does not interact with and is not phosphorylated by the Open Stomata 1 (OST1) kinase, the *Arabidopsis* ortholog of AAPK, suggesting that the physiological relationship between kinases and their respective RBPs has not been conserved between *Vicia faba* and *Arabidopsis thaliana* (Riera et al., 2006).

GRP7 is a glycine-rich RNA-binding protein also showing high homology to the human hnRNP A/B (Lorkovic and Barta, 2002; Wang and Brendel, 2004) that has been recently identified as a novel splicing regulator in *Arabidopsis* (Streitner et al., 2012). This hnRNP-like protein is regulated by the circadian clock and autoregulates its expression by influencing alternative splicing of its own pre-mRNA (Staiger et al., 2003) as well as of other splicing regulators or RBPs (Streitner et al., 2012). *GRP7* was reported to be strongly repressed by ABA and osmotic stress, with a loss-of-function mutant for this gene displaying hypersensitivity to ABA during seed germination (Cao et al., 2006). However, functional analyses of other *GRP7* mutant alleles has generated conflicting data concerning these altered ABA responses (reviewed by Lorkovic, 2009). In fact, a later study by Kim et al. (2008) found no ABA-related phenotypes during germination and seedling growth for either *GRP7* knockout/knockdown plants or overexpressor lines and showed that *GRP7* affects stomatal closure during drought stress in an ABA-independent manner.

Another Arabidopsis glycine-rich type hnRNP, atRZ-1a, was first described as carrying out prominent roles in plant cold and freezing tolerance (Kim et al., 2005), functioning as an RNA chaperon during the cold adaptation process (Kim and Kang, 2006). The RZ-1a protein contains RRM s at the N-terminus and a glycine-rich region interspersed with CCHC-type zinc fingers at the C-terminus. Expression of the *RZ-1a* gene is repressed by ABA (Kim et al., 2005). Under either ABA or glucose treatment, *RZ-1a* loss-of-function mutants germinate earlier than wild-type plants, whereas *RZ-1a* overexpressing lines display retarded germination and cannot develop cotyledons (Kim et al., 2007). These genotypes exhibited similar responses under salt and dehydration stress, which could indicate that RZ-1a negatively regulates seed germination and early plant growth under these conditions in an ABA-dependent manner (Kim et al., 2007). The transcript levels of ABA biosynthesis genes remained unaltered in mutant and overexpressor plants, but changes in several germination-responsive genes were observed under salt and dehydration stress as well as upon ABA and glucose treatments, suggesting that RZ-1a affects germination under different environmental conditions by modulating the expression of these genes (Kim et al., 2007).

2.3. Other Splicing Regulators

Sugliani et al. (2010) have described yet another splicing factor implicated in ABA signaling — Suppressor of ABI3-5 (SUA). SUA is an RBP with two RRM s that interacts with the U2AF pre-spliceosomal component (Sugliani et al., 2010). This protein shares high homology with the human tumor suppressor RBM5 (RNA-binding motif protein 5), which belongs to the spliceosome complex and regulates alternative splicing of apoptosis genes (Bonnal et al., 2008). On the other hand, the *Absciscic Acid Insensitive3* (*ABI3*) gene is a major regulator of seed maturation, and mutations in *ABI3* lead to seed insensitivity to ABA during germination, desiccation intolerance and reduced longevity (Ooms et al., 1993). Sugliani et al. (2010) have shown that the pre-mRNA of the *ABI3* transcription factor possesses a cryptic intron that is alternatively spliced leading to the occurrence of two transcripts, one encoding the full-length protein and the other a truncated form. Functional analyses in the *abi3-5* and *sua-1* loss-of-function mutants indicate that SUA influences seed maturation by suppressing splicing of the cryptic *ABI3* intron. Higher abundance of SUA will favor cryptic intron retention and increase full-length ABI3 protein levels during seed maturation.

3. NON-SPLICING PROTEINS

Other RBPs not specifically or directly involved in mRNA splicing have also been associated to ABA responses in plants.

3.1. CAP Binding Proteins

The RNA cap structure, characteristic of all RNA polymerase II transcripts, consists of 7-methyl guanosine linked via a 5'–5' triphosphate bridge to the first transcribed nucleotide. Most

of the roles described for the mRNA cap are mediated by specific factors called cap-binding proteins (CBPs), which recognize and bind to this structure. In mammals, several different nuclear CBPs have been described but only two of 20 kDa (CBP20) and 80 kDa (CBP80) have been characterized in detail. In order to bind the 5' terminal cap structure of the mRNA, these two proteins need to be together in the Cap Binding Complex (CBC). This heterodimeric nuclear complex offers protection for decapping enzymes, enhances splicing and promotes the first-round of translation (Izaurrealde et al., 1994; Lewis et al., 1996a; Lewis et al., 1996b; Flaherty et al., 1997; Lewis and Izaurrealde, 1997; Cheng et al., 2006).

In plants, defects in the CBC have been described to induce ABA-hypersensitive responses. The *abh1* (ABA hypersensitive 1) mutant was isolated from activation-tagged Arabidopsis lines based on its ABA hypersensitive inhibition of seed germination (Hugouvieux et al., 2001). The mutation was mapped to the *ABH1/CBP80* gene, which is homologous to the human *CBC80* encoding the large subunit of the CBC. Besides the germination phenotype, the *abh1* mutant also shows reduced wilting during drought, ABA-hypersensitive stomatal closing and enhanced ABA-induced guard-cell calcium elevations (Hugouvieux et al., 2001; reviewed in Wasilewska et al., 2008). Similarly, RNAi-mediated silencing of the CBP80 gene in potato plants resulted in increased drought tolerance (Pieczynski et al., 2013). These transgenic plants display enhanced stomatal closure when exposed to increased ABA concentrations, probably the predominant factor underlying their improved resistance to water stress.

Hugouvieux et al. (2001) also identified an Arabidopsis homolog of the smaller subunit of the CBC (AtCBP20), showing with yeast-two hybrid assays that it can interact with ABH1/CBP80. A year later, Kmiecik et al. (2002) cloned and characterized these two Arabidopsis CBP proteins, describing CBP20 to contain a canonical RNA-binding domain (RBD) and CBP80 a MIF4G domain thought to be involved in both protein-protein and protein-RNA interactions. Analogously to the *cbp80/abh1* mutant, the *cbp20* knockout also showed alterations of responses to ABA and drought (Papp et al., 2004).

The phenotypes caused by the *abh1* lesion may result from inadequate processing of specific mRNAs encoding key regulators of certain pathways. Indeed, the *abh1* mutation causes a reduction in the steady state levels of a limited number of mRNAs, including ABA-related genes such as the *AtPP2C* gene encoding a protein phosphatase 2C involved in the regulation of ABA signal transduction (Hugouvieux et al., 2001). It also affects the correct splicing of the first intron of the MADS box transcription factor *Flowering Locus C* (*FLC* gene), which may explain the early-flowering phenotype that has also been described as a consequence of disruption of the *ABH1* gene (Kuhn et al., 2007).

Importantly, the Arabidopsis CBC has also been implicated in alternative splicing (Raczynska et al., 2010). Splicing profiles were compared between wild type plants and the *cbp20* and *cbp80/abh1* single and double mutants by use of an RT-PCR alternative splicing panel (Simpson et al., 2008). Using this tool, Raczynska et al. (2010) were able to analyze 435 alternative splicing events in Arabidopsis, concluding that both ABH1/CBP80 and CBP20 affect alternative splicing of various stress-related genes, preferentially at the 5' splice site of the first intron. This is in agreement with the current model for animal systems that the CBC promotes 5' splice site selection within the first intron of pre-mRNAs.

3.2. RNA Helicases and Pentatricopeptide Repeat Proteins

RNA helicases function as molecular motors that rearrange RNA secondary structure or dissolve RNA-protein complexes. They are hence involved in many steps of RNA metabolism, ranging from transcription to RNA degradation, including pre-mRNA splicing, ribosome biogenesis, translation initiation and organellar gene expression (reviewed in Linder, 2006). In yeast, the DEAD (Asp-Glu-Ala-Asp) box-type helicases Sub2, Prp28 and Prp5 are required for *in vivo* splicing (reviewed in Staley and Guthrie, 1998) and in higher eukaryotes, the RNA helicases p68 and p72 are involved in alternative mRNA splicing (Honig et al., 2002; Liu, 2002; Guil et al., 2003).

In plants, RNA helicase gene families are larger and more diversified than in other systems, with 115 and 113 members in rice and Arabidopsis, respectively (Linder and Owtrim, 2009; Umate et al., 2010). Very little is known about the function of these families in plants, but the functional characterization of a few of their members has implicated them in abiotic stress responses, namely through the ABA pathway. Just recently, a plant DEAD box RNA helicase, RCF1 (Regulator of CBF gene expression 1), has been implicated in RNA splicing — *rcf1-1* mutant plants display mis-spliced cold-responsive genes under cold stress (Guan et al., 2013).

The Arabidopsis gene *LOS4* (*low expression of osmotically responsive genes 4*) encodes a DEAD box RNA helicase, with a mutant allele for this gene, *cryophyte*, being isolated in a genetic screen due to its enhanced cold tolerance (Gong et al., 2005). Besides being more resistant to chilling than wild-type plants, *cryophyte* mutants also display a hypersensitive response to ABA during germination and early seedling development. The observation that the *cryophyte* mutant accumulated stronger polyadenylation signals (hallmarks of translationally-competent mRNAs) in nuclei than did wild-type cells led to the suggestion that the *cryophyte* mutation influences mRNA export of some gene(s) involved in ABA signal transduction (Gong et al., 2005).

Two other Arabidopsis RNA helicases, Stress Response Suppressor 1 (STRS1) and STRS2, have been proposed as negative regulators of multiple abiotic stress responses, acting through both ABA-dependent and ABA-independent pathways (Kant et al., 2007). Mutations in either gene render increased tolerance to salt, osmotic and heat stresses and, consistent with their stress-tolerant phenotypes, *strs* mutants exhibit enhanced expression of both ABA-dependent and ABA-independent stress-responsive genes (Kant et al., 2007). Intriguingly, despite inducing upregulation of ABA-dependent stress-response genes, the absence of the STRS1 and STRS2 proteins did not result in ABA hypersensitivity — in fact, the *strs* mutants appear to be insensitive to the ABA inhibition of seed germination. Although their mode of action is yet unclear, STRS1 and STRS2 present themselves as regulatory nodes linking salt, osmotic, heat and ABA signaling networks.

Most mitochondrial proteins are encoded by nuclear genes, but a small fraction is encoded by the mitochondrial genome. The processing of mitochondrial RNAs, including intron splicing and RNA editing, requires the action of nuclear-encoded proteins such as RNA helicases and Pentatricopeptide Repeat (PPR) proteins (reviewed in O'Toole et al., 2008; Schmitz-Linneweber and Small, 2008; Kohler et al., 2010; Liu et al., 2010; He et al., 2012). The *ABA overly-sensitive 5* (*abo5*) and *abo6* mutants were isolated in a genetic screen for ABA-mediated inhibition of primary root growth, with the mutations being mapped to a PPR protein and a DEXH (Asp-Glu-X-His) Box RNA helicase, respectively (Liu et al., 2010; He et al., 2012). Both these proteins are involved in the correct splicing of mitochondrial RNAs in Arabidopsis.

PPR proteins form a large family of nuclear-encoded RBPs, particularly prevalent in terrestrial plants, which are mostly targeted to chloroplasts and mitochondria. Liu et al. (2010) have shown that a mutation in the *ABO5* PPR gene impairs splicing of the *nad2* NADH dehydrogenase subunit 2 of the mitochondrial complex I in the electron transport chain and leads to ABA hypersensitive phenotypes during early seedling development, such as delayed cotyledon greening and root growth. These might partially result from hyperaccumulation of Reactive Oxygen Species (ROS) in the *abo5* mutant (Liu et al., 2010), as disruption of ROS-producing genes impairs ABA inhibition of seed germination and root growth (Kwak et al., 2003). In addition, like many other ABA-related mutants, *abo5* shows enhanced sensitivity to high sugar concentrations (Liu et al., 2010). At the molecular level, not only did the *abo5* mutation impair splicing in the *nad2* gene, but also resulted in increased complex I and IV mitochondrial gene expression upon ABA treatment, despite the fact that *abo5* mutants are impaired in ABA-regulated gene expression (Liu et al., 2010).

The DEXH box RNA helicase encoded by the *ABO6* gene is also crucial for proper ABA signal transduction (He et al., 2012). When *abo6* mutants are exposed to increased levels of ABA, they exhibit various ABA-hypersensitive phenotypes, namely reduced seed germination and early seedling growth as well as reduced primary root growth and stomatal opening when compared to wild type plants (He et al., 2012). Similarly to what happens in *abo5*, ABA also enhances the production of ROS in *abo6* mutants. Moreover, He et al. (2012) observed that the *abo6* mutants display reduced content of both auxins and auxin carriers, probably due to ABA-mediated ROS overproduction. Accordingly, exogenous application of auxin rescued the altered *abo6* ABA responses to both root growth and germination. Finally, the *abo6* mutation was found to impair splicing of the mitochondrial complex I *NAD1b*, *NAD1c*, *NAD2a*, *NAD4* and *NAD5a* genes and to induce the accumulation of other mitochondrial transcripts (He et al., 2012).

3.3. Other RNA-Binding Proteins

In a screen for ABA-hypersensitive *Arabidopsis* transposon insertion lines, Lu and Fedoroff (2000) isolated the *hyponastic leaves 1* (*hyl1*) mutant displaying increased sensitivity to the inhibition of seed germination and root growth by ABA. The *HYL1* gene encodes a nuclear-localized RNA-binding protein that specifically binds double-stranded (ds)RNA (Lu and Fedoroff, 2000). The *hyl1* mutation causes an increase in basal abundance of ABA-inducible genes, namely of those encoding the ABI5 transcription factor as well as two mitogen-activated protein kinase (MAPK) proteins, which have been shown to phosphorylate ABI5 (Lu et al., 2002). This suggests that the ABA hypersensitivity of the *hyl1* mutant is due to elevated expression of stress MAPK signaling components and that the HYL1 protein could be involved in either transcription regulation or mRNA stability (reviewed in Lu and Fedoroff, 2000; Lu et al., 2002; Kuhn and Schroeder, 2003).

Disruption of another RNA processing component, the poly(A)-specific ribonuclease (PARN), a deadenylase involved in the first step of mRNA degradation, also causes abnormal ABA responses in *Arabidopsis*. Nishimura et al. (2005) characterized in detail a leaky transposon insertion mutant for this gene, *ahg2-1* (*ABA-hypersensitive germination 2*), and found that it is oversensitive to ABA both during germination and at adult rosette stages. The *ahg2-1* mutation causes more ABA to accumulate in seeds and later in development, as well as

in response to the osmotic stressor mannitol, which could explain the observed ABA phenotype. In addition, *ahg2-1* mutant plants are also hypersensitive to salicylic acid (SA) and resistant to drought stress (Nishimura et al., 2005). At the molecular level, microarray analysis showed several stress-induced genes to be upregulated in the mutant. Clustering analysis suggested that this pattern is very similar to that of upregulated genes in ABA- or SA-treated wild-type plants (Nishimura et al., 2005). The authors speculate that PARN/AHG2 has a pivotal role in downregulating target genes by destabilizing mRNAs to maintain proper stress responses. Interestingly, the *PARN/AHG2* gene undergoes alternative splicing, but the physiological roles of the variant transcripts have not yet been investigated.

A further study providing evidence for a connection between RNA metabolism and ABA signaling describes the involvement of the RNA-binding protein Tudor staphylococcal nuclease (TSN) in plant stress responses (dit Frey et al., 2010). TSN is a multifunctional protein that in humans has been implicated in a variety of cellular processes, including gene transcription and pre-mRNA splicing (Yang et al., 2007; Gao et al., 2012). Both the sequence and the domain architecture of TSN proteins are well conserved in eukaryotes. The Arabidopsis genome contains two highly similar genes, designated *TSN1* and *TSN2*, encoding Tudor proteins with four staphylococcal nuclease (SN) domains, followed by a Tudor domain connected to a fifth SN domain. The Tudor domain was originally identified as a region of 50 amino acids found in the *Drosophila* Tudor protein, and has been subsequently found within a number of proteins involved in RNA binding (reviewed in Pek et al., 2012). Phenotypical analysis of a null *tsn1 tsn2* double mutant, obtained by crossing the single T-DNA insertion *tsn1* and *tsn2* mutants, determined TSN to function in the control of RNA stability during plant stress responses (dit Frey et al., 2010; reviewed in Ambrosone et al., 2012; Nakaminami et al., 2012). In fact, germination, seedling growth and survival under high salinity stress were drastically affected in *tsn1 tsn2* plants, with these double mutants also displaying hypersensitivity to ABA during germination (dit Frey et al., 2010).

A putative dsRNA binding protein, FIERY2/CTD-phosphatase-like 1 (FRY2/CPL1), has also been implicated in ABA and stress responses (Xiong et al., 2002). FRY2/CPL1 possesses two dsRNA-binding domains and a conserved region with similarity to the catalytic domain of RNA polymerase II Carboxy-terminal Domain (CTD) phosphatases found in yeast and animals. CTD phosphorylation is known to influence cotranscriptional pre-mRNA processing, namely splicing and poly(A) site cleavage (Bird et al., 2004). The presence of this catalytic domain in FRY2 suggests that it may possess catalytic functions, and indeed Koiwa et al. (2002) have shown that the FRY2 recombinant protein exhibits phosphatase activity. As FRY2 contains dsRNA-binding domains, its enzymatic activity could be regulated by dsRNA (Xiong et al., 2002). *fry2* mutations result in the induction of the Drought Responsive/C-Repeat (DRE/CRT) class of stress-responsive genes upon cold, salt and ABA treatments. Furthermore, *fry2* mutants display insensitivity to salt and ABA during seed germination and early seedling development, but are hypersensitive to ABA during root elongation, and mutant seedlings also show increased susceptibility to freezing damage (Xiong et al., 2002). FRY2 is a negative transcriptional regulator of stress-related genes playing different roles and exhibiting ABA responsiveness during various developmental stages.

Recently, Jung et al. (2013) selected an ABA-regulated RBP, which they named ARP1 (ABA-regulated RNA-binding Protein), from the GENEVESTIGATOR expression database for detailed functional characterization. The *ARP* gene, downregulated by ABA, encodes a nuclear protein containing a well-conserved RRM and is present in Arabidopsis as a unique

gene. To assess the potential role of this RBP in ABA responses, Jung et al. (2013) analyzed ARP1-overexpressing transgenic *Arabidopsis* plants and a loss-of-function T-DNA insertion mutant, *arp1*. Upon exogenous ABA application, similar phenotypes were observed for both the null-mutant and the overexpression lines — hypersensitivity to ABA during seed germination and cotyledon greening (Jung et al., 2013). In accordance with the ABA germination phenotype, several germination-responsive genes were upregulated in *arp1* and ARP1-overexpression plants (Jung et al., 2013). The fact that ARP1 overexpression induced the same effects as knocking out the corresponding gene suggests that overaccumulation of ARP1 inhibits its normal function and thus that levels of its transcript need to be tightly regulated (Jung et al., 2013). The findings of this recent study once again provide evidence for involvement of an RBP in ABA and abiotic stress responses in plants.

CONCLUSION

Studies on the RNA regulation of plant responses to abiotic stress have shown considerable progress in the past decade. As reviewed here, the functional and molecular characterization of several plant RBPs has provided evident links between mRNA processing mechanisms and ABA-mediated plant stress responses. In particular, the loss- or gain-of-function of genes encoding various splicing-related factors, such as snRNPs or SR and hnRNP proteins, have been shown to result in striking ABA-response phenotypes at various stages of plant development. Interestingly, the same has been observed when knocking out other RBP genes, such as those coding for CBPs, RNA helicases, PPR proteins, poly(A) processing enzymes or dsRBPs. The fact that members of a wide range of classes of RNA-metabolism proteins appear to be indispensable for adequate ABA biosynthesis or signal transduction supports a global role for posttranscriptional control of gene expression in regulating plant responses to environmental cues.

To explain the well-defined plant phenotypes associated with altered RBP levels and dissect the role of these factors in the response to stress, the cellular and molecular mechanisms underlying their mode of action in regulating the ABA pathway will need to be uncovered. A key step towards this goal will be the identification of the direct RNA targets and ligands of these RBPs. Importantly, Streitner et al. (2012) recently succeeded in validating putative RNA targets for the hnRNP GRP7 protein by means of RNA immunoprecipitation (RIP) of ribonucleoprotein complexes. Using this RIP approach, the authors were able to confirm direct binding of GRP7 to seven target transcripts that had been identified as showing reciprocal splicing patterns in GRP7 mutant and overexpressing lines using a high-resolution RT-PCR-based alternative splicing panel. Future similar approaches hold much promise in unravelling the mechanistic basis of posttranscriptional regulation of plant ABA-mediated stress responses.

Table 1. Plant RNA-binding Proteins Involved in ABA Stress Responses

	Gene name	Locus ID	Domains (#)	Associated phenotypes	References
Splicing factors					
<i>snRNP factors</i>					
Sm-like snRNP protein (homolog of human LSM5 and yeast Lsm5)	<i>SAD1/LSM5</i>	At5g48870	Sm (1)	<i>sad1</i> mutant hypersensitive to ABA and drought stress during germination and root growth as well as deficient in drought-induced ABA biosynthesis	Xiong et al., 2001a
Sm-like U6 snRNP protein (homolog of human Lsm4 and yeast Lsm4)	<i>LSM4</i>	At5g27720	Sm (1)	<i>lsm4</i> mutant hypersensitive to ABA and salt stress during germination and seedling development	Zhang et al., 2011
U5-snRNP specific protein (homolog of human U5 snRNP-associated protein Prp6b-like and yeast Prp splicing factors)	<i>STA1/U5-102kDa</i>	At4g03430	TPR (3) HAT (15) TPR-like (2) Prp1_N (1)	<i>sta1</i> mutant hypersensitive to ABA and LiCl salt stress during germination and root growth as well as over sensitive to chilling	Lee et al., 2006
<i>SR and hnRNPs proteins</i>					
SR-like protein	<i>SR45</i>	At1g16610	RRM (1), RS (2)	<i>sr45-1</i> mutant hypersensitive to ABA and glucose during early seedling development	Carvalho et al., 2010
SR protein	<i>SCL30a</i>	At3g13570	RRM (1) RS (1)	<i>sc30a-1</i> insertion mutant hypersensitive to ABA as well as to salt and osmotic stress during seed germination	Carvalho et al., unpublished
hnRNP (homolog of human hnRNP A/B)	<i>AKIP1</i>	AKIP1 (<i>Vicia faba</i>)	RRM (1)	ABA activates the RNA-binding activity of AKIP1 and induces its relocation into nuclear speckles	Li et al., 2002
hnRNP (GRP) (homolog of <i>V. faba</i> AKIP1 and human hnRNP A/B)	<i>UBA2a</i>	At3g56860	RRM (2)	ABA induces relocation of UBA2a into nuclear speckles	Riera et al., 2006
hnRNP (GRP) (homolog of human hnRNP A/B)	<i>GRP7</i>	At2g21660	RRM (1)	<i>grp7-1</i> mutant hypersensitive to ABA during germination; conflicting data with other mutant alleles	Cao et al., 2006; Kim et al., 2008
hnRNP (GRP) (homolog of human hnRNP G)	<i>RZ1-a/hnRNP-G2</i>	At3g26420	RRM (1) ZnD (1)	<i>RZ1a</i> knockout mutants/overexpressor lines insensitive/hypersensitive to ABA and glucose as well as to salt and dehydration stress during germination and seedling growth	Kim et al., 2007
<i>Other splicing regulators</i>					
RBP, splicing factor (homolog of human RBM5)	<i>SUA</i>	At3g54230	RRM (2) ZnD (1)	ABA seed maturation-related phenotype	Sugliani et al., 2010

Non-splicing proteins					
Cap-binding proteins					
nuclear cap-binding protein (homolog of human CBP80 and yeast Cbc1)	<i>ABH1/CBP80</i>	At2g13540	MIF4G (1) ARM (3)	<i>abh1</i> mutant hypersensitive to ABA during germination and stomatal closure as well as tolerant to drought stress	Hugouvieux et al., 2001
nuclear cap-binding protein (homolog of human CBP20 and yeast Cbc2p)	<i>CBP20</i>	At5g44200	RRM (1)	<i>cbp20</i> mutant hypersensitive to ABA during germination and stomatal closure as well as tolerant to drought stress	Papp et al., 2004
RNA helicases and Pentatricopeptide Repeat proteins					
DEAD Box RNA helicase	<i>cryophyte/LOS4</i>	At3g53110	DEAD Box (8)	<i>cryophyte/los4-2</i> mutant hypersensitive to ABA during germination and early seedling development as well as tolerant to low temperatures	Gong et al., 2005
DEAD Box RNA helicases	<i>STRS1</i> <i>STRS2</i>	At5g08620 At5g08620	DEAD Box (9)	<i>strs1</i> , <i>strs1a</i> , <i>strs2</i> , <i>strs2a</i> mutants insensitive to ABA, salt and osmotic stress during germination	Kant et al., 2007
PPR protein	<i>ABO5</i>	At1g51965	PPR (6)	<i>abo5</i> mutant hypersensitive to ABA during early seedling development and root growth and to sugars during early growth	Liu et al., 2010
DEXH Box RNA helicase	<i>ABO6</i>	At5g04895	dsRBD (1) DEXH Box (1)	<i>abo6</i> mutant hypersensitive to ABA during germination, early seedling development, root growth and stomatal closure as well as tolerant to drought stress	He et al., 2012
Other RNA-binding proteins					
dsRBP	<i>HYL1</i>	At1g09700	dsRBD (2)	<i>hyl1</i> mutant hypersensitive to ABA during germination and root growth as well as less sensitive to auxins and cytokinins	Lu and Fedoroff, 2000
Poly(A) Ribonuclease	<i>PARN/AHG2</i>	At1g55870	Ribonuclease HD	<i>ahg2-1</i> mutant hypersensitive to ABA and salicylic acid during germination and plant development as well as tolerant to drought stress	Nishimura et al., 2005
Tudor RBP	<i>TSN1</i> <i>TSN2</i>	At5g073550 At5g61780	SN (5) Tudor (1)	<i>tsn1 tsn2</i> double mutants hypersensitive to ABA and salt during germination and seedling growth	dit Frey et al., 2010
putative RBP	<i>FRY2/CPL</i>	At4g21670	dsRBD (2) CTD (1)	<i>fry2</i> mutant insensitive to ABA and salt during germination and early seedling development but hypersensitive to ABA during root growth; also more susceptible to freezing damage	Xiong et al., 2002
putative RBP	<i>ARP1</i>	At3g54770	RRM (1)	<i>ARP1</i> knockout mutants/overexpressor lines hypersensitive to ABA as well as to salt and dehydration stress during germination and seedling growth	Jung et al., 2013

Abbreviations for protein domains are as follows: TPR, Tetratricopeptide; HAT, Half A Tetratricopeptide repeat; PRP1_N, PRP1 splicing factor, N terminal; RRM, RNA-Recognition Motif; RS, arginine/serine dipeptides; ZnD, Zinc-finger domain; MIF4G, Initiation Factor eIF-4 gamma; ARM, ARM Repeat Fold; DEAD box, Asp-Glu-Ala-Asp box; PPR, Pentatricopeptide Repeats; dsRBD, double stranded RNA-binding Domain; DEXH box, Asp-Glu-X-His box; RibonucleaseHD, Ribonuclease H-like domain SN, staphylococcal nuclease; RBP, RNA-binding Protein; CTD, Carboxy-Terminal Domain; RBM5, RNA Binding Motif Protein 5.

REFERENCES

- Achsel, T., Brahms, H., Kastner, B., Bachi, A., Wilm, M. & Luhrmann, R. (1999). A doughnut-shaped heteromer of human Sm-like proteins binds to the 3'-end of U6 snRNA, thereby facilitating U4/U6 duplex formation in vitro. *Embo J*, 18, 5789-5802.
- Ali, G. S., Palusa, S. G., Golovkin, M., Prasad, J., Manley, J. L. & Reddy, A. S. (2007). Regulation of plant developmental processes by a novel splicing factor. *PLoS ONE*, 2, e471.
- Ambrosone, A., Costa, A., Leone, A. & Grillo, S. (2012). Beyond transcription, RNA-binding proteins as emerging regulators of plant response to environmental constraints. *Plant Sci*, 182, 12-18.
- Bell, L. R., Horabin, J. I., Schedl, P. & Cline, T. W. (1991). Positive autoregulation of sex-lethal by alternative splicing maintains the female determined state in *Drosophila*. *Cell*, 65, 229-239.
- Bird, G., Zorio, D. A. & Bentley, D. L. (2004). RNA polymerase II carboxy-terminal domain phosphorylation is required for cotranscriptional pre-mRNA splicing and 3'-end formation. *Mol Cell Biol*, 24, 8963-8969.
- Blencowe, B. J. (2006). *Alternative splicing, new insights from global analyses*. *Cell*, 126, 37-47.
- Bonnal, S., Martinez, C., Forch, P., Bachi, A., Wilm, M. & Valcarcel, J. (2008). RBM5/Luca-15/H37 regulates Fas alternative splice site pairing after exon definition. *Mol Cell*, 32, 81-95.
- Bouveret, E., Rigaut, G., Shevchenko, A., Wilm, M. & Seraphin, B. (2000). A Sm-like protein complex that participates in mRNA degradation. *Embo J*, 19, 1661-1671.
- Burd, C. G. & Dreyfuss, G. (1994). Conserved structures and diversity of functions of RNA-binding proteins. *Science*, 265, 615-621.
- Cao, S., Jiang, L., Song, S., Jing, R. & Xu, G. (2006). AtGRP7 is involved in the regulation of abscisic acid and stress responses in *Arabidopsis*. *Cell Mol Biol Lett*, 11, 526-535.
- Carvalho, R. F., Carvalho, S. D. & Duque, P. (2010). The plant-specific SR45 protein negatively regulates glucose and ABA signaling during early seedling development in *Arabidopsis*. *Plant Physiol*, 154, 772-783.
- Cheng, H., Dufu, K., Lee, C. S., Hsu, J. L., Dias, A. & Reed, R. (2006). Human mRNA export machinery recruited to the 5' end of mRNA. *Cell*, 127, 1389-1400.
- Cook, K. B., Kazan, H., Zuberi, K., Morris, Q. & Hughes, T. R. (2010). RBPDB, a database of RNA-binding specificities. *Nucleic Acids Res*, 39, D301-308.
- Deng, X., Gu, L., Liu, C., Lu, T., Lu, F., Lu, Z., Cui, P., Pei, Y., Wang, B., Hu, S. & Cao, X. (2010). Arginine methylation mediated by the *Arabidopsis* homolog of PRMT5 is essential for proper pre-mRNA splicing. *Proc Natl Acad Sci U S A*, 107, 19114-19119.
- dit Frey, N. F., Muller, P., Jammes, F., Kizis, D., Leung, J., Perrot-Rechenmann, C. & Bianchi, M. W. (2010). The RNA binding protein Tudor-SN is essential for stress tolerance and stabilizes levels of stress-responsive mRNAs encoding secreted proteins in *Arabidopsis*. *Plant Cell*, 22, 1575-1591.
- Flaherty, S. M., Fortes, P., Izaurralde, E., Mattaj, I. W. & Gilmartin, G. M. (1997). Participation of the nuclear cap binding complex in pre-mRNA 3' processing. *Proc Natl Acad Sci U S A*, 94, 11893-11898.

- Gao, X., Zhao, X., Zhu, Y., He, J., Shao, J., Su, C., Zhang, Y., Zhang, W., Saarikettu, J., Silvennoinen, O., Yao, Z. & Yang, J. (2012). Tudor staphylococcal nuclease (Tudor-SN) participates in small ribonucleoprotein (snRNP) assembly via interacting with symmetrically dimethylated Sm proteins. *J Biol Chem*, 287, 18130-18141.
- Gong, Z., Dong, C. H., Lee, H., Zhu, J., Xiong, L., Gong, D., Stevenson, B. & Zhu, J. K. (2005). A DEAD box RNA helicase is essential for mRNA export and important for development and stress responses in Arabidopsis. *Plant Cell*, 17, 256-267.
- Guan, Q., Wu, J., Zhang, Y., Jiang, C., Liu, R., Chai, C. & Zhu, J. (2013). A DEAD box RNA helicase is critical for pre-mRNA splicing, cold-responsive gene regulation, and cold tolerance in Arabidopsis. *Plant Cell*, 25, 342-356.
- Guil, S., Gattoni, R., Carrascal, M., Abian, J., Stevenin, J. & Bach-Elias, M. (2003). Roles of hnRNP A1, SR proteins, and p68 helicase in c-H-ras alternative splicing regulation. *Mol Cell Biol*, 23, 2927-2941.
- Han, S. P., Tang, Y. H. & Smith, R. (2010). Functional diversity of the hnRNPs, past, present and perspectives. *Biochem J*, 430, 379-392.
- He, J., Duan, Y., Hua, D., Fan, G., Wang, L., Liu, Y., Chen, Z., Han, L., Qu, L. J. & Gong, Z. (2012). DEXH box RNA helicase-mediated mitochondrial reactive oxygen species production in Arabidopsis mediates crosstalk between abscisic acid and auxin signaling. *Plant Cell*, 24, 1815-1833.
- He, W. & Parker, R. (2000). Functions of Lsm proteins in mRNA degradation and splicing. *Curr Opin Cell Biol*, 12, 346-350.
- Hong, J. H., Seah, S. W. & Xu, J. (2013). The root of ABA action in environmental stress response. *Plant Cell Rep*.
- Honig, A., Auboeuf, D., Parker, M. M., O'Malley, B. W. & Berget, S. M. (2002). Regulation of alternative splicing by the ATP-dependent DEAD-box RNA helicase p72. *Mol Cell Biol*, 22, 5698-5707.
- Hoth, S., Morgante, M., Sanchez, J. P., Hanafey, M. K., Tingey, S. V. & Chua, N. H. (2002). Genome-wide gene expression profiling in Arabidopsis thaliana reveals new targets of abscisic acid and largely impaired gene regulation in the abi1-1 mutant. *J Cell Sci*, 115, 4891-4900.
- Hugouvieux, V., Kwak, J. M. & Schroeder, J. I. (2001). An mRNA cap binding protein, ABH1, modulates early abscisic acid signal transduction in Arabidopsis. *Cell*, 106, 477-487.
- International Barley Genome Sequencing, C., Mayer, K. F., Waugh, R., Brown, J. W., Schulman, A., Langridge, P., Platzer, M., Fincher, G. B., Muehlbauer, G. J., Sato, K., Close, T. J., Wise, R. P. & Stein, N. (2012). A physical, genetic and functional sequence assembly of the barley genome. *Nature*, 491, 711-716.
- Izaurrealde, E., Lewis, J., McGuigan, C., Jankowska, M., Darzynkiewicz, E. & Mattaj, I. W. (1994). A nuclear cap binding protein complex involved in pre-mRNA splicing. *Cell*, 78, 657-668.
- Jung, H. J., Kim, M. K. & Kang, H. (2013). An ABA-regulated putative RNA-binding protein affects seed germination of Arabidopsis under ABA or abiotic stress conditions. *J Plant Physiol*, 170, 179-184.
- Kalyna, M., Lopato, S. & Barta, A. (2003). Ectopic expression of atRSZ33 reveals its function in splicing and causes pleiotropic changes in development. *Mol Biol Cell*, 14, 3565-3577.

- Kant, P., Kant, S., Gordon, M., Shaked, R. & Barak, S. (2007). Stress Response Suppressor1 and Stress Response Suppressor2, two DEAD-box RNA helicases that attenuate Arabidopsis responses to multiple abiotic stresses. *Plant Physiol*, 145, 814-830.
- Kim, J. S., Jung, H. J., Lee, H. J., Kim, K. A., Goh, C. H., Woo, Y., Oh, S. H., Han, Y. S. & Kang, H. (2008). Glycine-rich RNA-binding protein 7 affects abiotic stress responses by regulating stomata opening and closing in Arabidopsis thaliana. *Plant J*, 55, 455-466.
- Kim, Y. O. & Kang, H. (2006). The role of a zinc finger-containing glycine-rich RNA-binding protein during the cold adaptation process in Arabidopsis thaliana. *Plant Cell Physiol*, 47, 793-798.
- Kim, Y. O., Kim, J. S. & Kang, H. (2005). Cold-inducible zinc finger-containing glycine-rich RNA-binding protein contributes to the enhancement of freezing tolerance in Arabidopsis thaliana. *Plant J*, 42, 890-900.
- Kim, Y. O., Pan, S., Jung, C. H. & Kang, H. (2007). A zinc finger-containing glycine-rich RNA-binding protein, atRZ-1a, has a negative impact on seed germination and seedling growth of Arabidopsis thaliana under salt or drought stress conditions. *Plant Cell Physiol*, 48, 1170-1181.
- Kmiecik, M., Simpson, C. G., Lewandowska, D., Brown, J. W. & Jarmolowski, A. (2002). Cloning and characterization of two subunits of Arabidopsis thaliana nuclear cap-binding complex. *Gene*, 283, 171-183.
- Kohler, D., Schmidt-Gattung, S. & Binder, S. (2010). The DEAD-box protein PMH2 is required for efficient group II intron splicing in mitochondria of Arabidopsis thaliana. *Plant Mol Biol*, 72, 459-467.
- Kohtz, J. D., Jamison, S. F., Will, C. L., Zuo, P., Luhrmann, R., Garcia-Blanco, M. A. & Manley, J. L. (1994). Protein-protein interactions and 5'-splice-site recognition in mammalian mRNA precursors. *Nature*, 368, 119-124.
- Koiwa, H., Barb, A. W., Xiong, L., Li, F., McCully, M. G., Lee, B. H., Sokolchik, I., Zhu, J., Gong, Z., Reddy, M., Sharkhuu, A., Manabe, Y., Yokoi, S., Zhu, J. K., Bressan, R. A. & Hasegawa, P. M. (2002). C-terminal domain phosphatase-like family members (AtCPLs) differentially regulate Arabidopsis thaliana abiotic stress signaling, growth, and development. *Proc Natl Acad Sci U S A*, 99, 10893-10898.
- Krecic, A. M. & Swanson, M. S. (1999). hnRNP complexes, composition, structure, and function. *Curr Opin Cell Biol*, 11, 363-371.
- Kuhn, J. M., Breton, G. & Schroeder, J. I. (2007). mRNA metabolism of flowering-time regulators in wild-type Arabidopsis revealed by a nuclear cap binding protein mutant, abh1. *Plant J*, 50, 1049-1062.
- Kuhn, J. M. & Schroeder, J. I. (2003). Impacts of altered RNA metabolism on abscisic acid signaling. *Curr Opin Plant Biol*, 6, 463-469.
- Kwak, J. M., Mori, I. C., Pei, Z. M., Leonhardt, N., Torres, M. A., Dangel, J. L., Bloom, R. E., Bodde, S., Jones, J. D. & Schroeder, J. I. (2003). NADPH oxidase AtrbohD and AtrbohF genes function in ROS-dependent ABA signaling in Arabidopsis. *Embo J*, 22, 2623-2633.
- Lambermon, M. H., Fu, Y., Wicczorek Kirk, D. A., Dupasquier, M., Filipowicz, W. & Lorkovic, Z. J. (2002). UBA1 and UBA2, two proteins that interact with UBP1, a multifunctional effector of pre-mRNA maturation in plants. *Mol Cell Biol*, 22, 4346-4357.
- Lambermon, M. H., Simpson, G. G., Wicczorek Kirk, D. A., Hemmings-Mieszczak, M., Klahre, U. & Filipowicz, W. (2000). UBP1, a novel hnRNP-like protein that functions at multiple steps of higher plant nuclear pre-mRNA maturation. *Embo J*, 19, 1638-1649.

- Lamond, A. I. & Spector, D. L. (2003). Nuclear speckles, a model for nuclear organelles. *Nat Rev Mol Cell Biol*, 4, 605-612.
- Lee, B. H., Kapoor, A., Zhu, J., Zhu, J. K. (2006). STABILIZED1, a stress-upregulated nuclear protein, is required for pre-mRNA splicing, mRNA turnover, and stress tolerance in Arabidopsis. *Plant Cell*, 18, 1736-1749.
- Lewis, J. D., Gorlich, D. & Mattaj, I. W. (1996a). A yeast cap binding protein complex (yCBC) acts at an early step in pre-mRNA splicing. *Nucleic Acids Res*, 24, 3332-3336.
- Lewis, J. D. & Izaurralde, E. (1997). The role of the cap structure in RNA processing and nuclear export. *Eur J Biochem*, 247, 461-469.
- Lewis, J. D., Izaurralde, E., Jarmolowski, A., McGuigan, C. & Mattaj, I. W. (1996b). A nuclear cap-binding complex facilitates association of U1 snRNP with the cap-proximal 5' splice site. *Genes Dev*, 10, 1683-1698.
- Li, J., Kinoshita, T., Pandey, S., Ng, C. K., Gygi, S. P., Shimazaki, K. & Assmann, S. M. (2002). Modulation of an RNA-binding protein by abscisic-acid-activated protein kinase. *Nature*, 418, 793-797.
- Li, J., Wang, X. Q., Watson, M. B. & Assmann, S. M. (2000). Regulation of abscisic acid-induced stomatal closure and anion channels by guard cell AAPK kinase. *Science*, 287, 300-303.
- Linder, P. (2006). Dead-box proteins, a family affair--active and passive players in RNP-remodeling. *Nucleic Acids Res*, 34, 4168-4180.
- Linder, P. & Owtrim, G. W. (2009). Plant RNA helicases, linking aberrant and silencing RNA. *Trends Plant Sci*, 14, 344-352.
- Liu, Y., He, J., Chen, Z., Ren, X., Hong, X. & Gong, Z. (2010). ABA overly-sensitive 5 (ABO5), encoding a pentatricopeptide repeat protein required for cis-splicing of mitochondrial nad2 intron 3, is involved in the abscisic acid response in Arabidopsis. *Plant J*, 63, 749-765.
- Liu, Z. R. (2002). p68 RNA helicase is an essential human splicing factor that acts at the U1 snRNA-5' splice site duplex. *Mol Cell Biol*, 22, 5443-5450.
- Lopato, S., Kalyna, M., Dörner, S., Kobayashi, R., Krainer, A. R. & Barta, A. (1999). atSRp30, one of two SF2/ASF-like proteins from Arabidopsis thaliana, regulates splicing of specific plant genes. *Genes Dev*, 13, 987-1001.
- Lorkovic, Z. J. (2009). Role of plant RNA-binding proteins in development, stress response and genome organization. *Trends Plant Sci*, 14, 229-236.
- Lorkovic, Z. J. & Barta, A. (2002). Genome analysis, RNA recognition motif (RRM) and K homology (KH) domain RNA-binding proteins from the flowering plant Arabidopsis thaliana. *Nucleic Acids Res*, 30, 623-635.
- Lu, C. & Fedoroff, N. (2000). A mutation in the Arabidopsis HYL1 gene encoding a dsRNA binding protein affects responses to abscisic acid, auxin, and cytokinin. *Plant Cell*, 12, 2351-2366.
- Lu, C., Han, M. H., Guevara-Garcia, A. & Fedoroff, N. V. (2002). Mitogen-activated protein kinase signaling in postgermination arrest of development by abscisic acid. *Proc Natl Acad Sci U S A*, 99, 15812-15817.
- Lu, T., Lu, G., Fan, D., Zhu, C., Li, W., Zhao, Q., Feng, Q., Zhao, Y., Guo, Y., Huang, X. & Han, B. (2010). Function annotation of the rice transcriptome at single-nucleotide resolution by RNA-seq. *Genome Res*, 20, 1238-1249.

- Lunde, B. M., Moore, C. & Varani, G. (2007). RNA-binding proteins, modular design for efficient function. *Nat Rev Mol Cell Biol*, 8, 479-490.
- Manley, J. L. & Tacke, R. (1996). SR proteins and splicing control. *Genes Dev*, 10, 1569-1579.
- Marquez, Y., Brown, J. W., Simpson, C., Barta, A. & Kalyna, M. (2012). Transcriptome survey reveals increased complexity of the alternative splicing landscape in Arabidopsis. *Genome Res*, 22, 1184-1195.
- Martinez-Contreras, R., Cloutier, P., Shkreta, L., Fisette, J. F., Revil, T. & Chabot, B. (2007). hnRNP proteins and splicing control. *Adv Exp Med Biol*, 623, 123-147.
- Misteli, T., Caceres, J. F. & Spector, D. L. (1997). The dynamics of a pre-mRNA splicing factor in living cells. *Nature*, 387, 523-527.
- Nakaminami, K., Matsui, A., Shinozaki, K. & Seki, M. (2012). RNA regulation in plant abiotic stress responses. *Biochim Biophys Acta*, 1819, 149-153.
- Nilsen, T. W. & Graveley, B. R. (2010). Expansion of the eukaryotic proteome by alternative splicing. *Nature*, 463, 457-463.
- Nishimura, N., Kitahata, N., Seki, M., Narusaka, Y., Narusaka, M., Kuromori, T., Asami, T., Shinozaki, K. & Hirayama, T. (2005). Analysis of ABA hypersensitive germination2 revealed the pivotal functions of PARN in stress response in Arabidopsis. *Plant J*, 44, 972-984.
- O'Toole, N., Hattori, M., Andres, C., Iida, K., Lurin, C., Schmitz-Linneweber, C., Sugita, M. & Small, I. (2008). On the expansion of the pentatricopeptide repeat gene family in plants. *Mol Biol Evol*, 25, 1120-1128.
- Ooms, J., Leon-Kloosterziel, K. M., Bartels, D., Koornneef, M. & Karssen, C. M. (1993). Acquisition of Desiccation Tolerance and Longevity in Seeds of Arabidopsis thaliana (A Comparative Study Using Absciscic Acid-Insensitive abi3 Mutants). *Plant Physiol*, 102, 1185-1191.
- Pan, Q., Shai, O., Lee, L. J., Frey, B. J. & Blencowe, B. J. (2008). Deep surveying of alternative splicing complexity in the human transcriptome by high-throughput sequencing. *Nat Genet*, 40, 1413-1415.
- Papp, I., Mur, L. A., Dalmadi, A., Dulai, S. & Koncz, C. (2004). A mutation in the Cap Binding Protein 20 gene confers drought tolerance to Arabidopsis. *Plant Mol Biol*, 55, 679-686.
- Pek, J. W., Anand, A. & Kai, T. (2012). Tudor domain proteins in development. *Development*, 139, 2255-2266.
- Perea-Resa, C., Hernandez-Verdeja, T., Lopez-Cobollo, R., del Mar Castellano, M. & Salinas, J. (2012). LSM proteins provide accurate splicing and decay of selected transcripts to ensure normal Arabidopsis development. *Plant Cell*, 24, 4930-4947.
- Pieczynski, M., Marczewski, W., Hennig, J., Dolata, J., Bielewicz, D., Piontek, P., Wyrzykowska, A., Krusiewicz, D., Strzelczyk-Zyta, D., Konopka-Postupolska, D., Krzeslowska, M., Jarmolowski, A. & Szweykowska-Kulinska, Z. (2013). Down-regulation of CBP80 gene expression as a strategy to engineer a drought-tolerant potato. *Plant Biotechnol J*, 11, 459-469.
- Raab, S. & Hoth, S. (2007). A mutation in the AtPRP4 splicing factor gene suppresses seed development in Arabidopsis. *Plant Biol (Stuttg)*, 9, 447-452.
- Raczynska, K. D., Simpson, C. G., Ciesiolka, A., Szewc, L., Lewandowska, D., McNicol, J., Szweykowska-Kulinska, Z., Brown, J. W. & Jarmolowski, A. (2010). Involvement of the nuclear cap-binding protein complex in alternative splicing in Arabidopsis thaliana. *Nucleic Acids Res*, 38, 265-278.

- Ramanathan, K., Michael, T. H., Jiang, G. J., Hiel, H. & Fuchs, P. A. (1999). A molecular mechanism for electrical tuning of cochlear hair cells. *Science*, 283, 215-217.
- Riera, M., Redko, Y. & Leung, J. (2006). Arabidopsis RNA-binding protein UBA2a relocalizes into nuclear speckles in response to abscisic acid. *FEBS Lett*, 580, 4160-4165.
- Salgado-Garrido, J., Bragado-Nilsson, E., Kandels-Lewis, S. & Seraphin, B. (1999). Sm and Sm-like proteins assemble in two related complexes of deep evolutionary origin. *Embo J*, 18, 3451-3462.
- Sanchez, S. E., Petrillo, E., Beckwith, E. J., Zhang, X., Rugnone, M. L., Hernando, C. E., Cuevas, J. C., Godoy Herz, M. A., Depetris-Chauvin, A., Simpson, C. G., Brown, J. W., Cerdan, P. D., Borevitz, J. O., Mas, P., Ceriani, M. F., Kornblihtt, A. R. & Yanovsky, M. J. (2010). A methyl transferase links the circadian clock to the regulation of alternative splicing. *Nature*, 468, 112-116.
- Schmitz-Linneweber, C. & Small, I. (2008). Pentatricopeptide repeat proteins, a socket set for organelle gene expression. *Trends Plant Sci*, 13, 663-670.
- Seo, M., Peeters, A. J., Koiwai, H., Oritani, T., Marion-Poll, A., Zeevaart, J. A., Koornneef, M., Kamiya, Y. & Koshiba, T. (2000). The Arabidopsis aldehyde oxidase 3 (AAO3) gene product catalyzes the final step in abscisic acid biosynthesis in leaves. *Proc Natl Acad Sci U S A*, 97, 12908-12913.
- Shen, H. & Green, M. R. (2004). A pathway of sequential arginine-serine-rich domain-splicing signal interactions during mammalian spliceosome assembly. *Mol Cell*, 16, 363-373.
- Simpson, C. G., Fuller, J., Maronova, M., Kalyna, M., Davidson, D., McNicol, J., Barta, A. & Brown, J. W. (2008). Monitoring changes in alternative precursor messenger RNA splicing in multiple gene transcripts. *Plant J*, 53, 1035-1048.
- Staiger, D., Zecca, L., Wiczeorek Kirk, D. A., Apel, K. & Eckstein, L. (2003). The circadian clock regulated RNA-binding protein AtGRP7 autoregulates its expression by influencing alternative splicing of its own pre-mRNA. *Plant J*, 33, 361-371.
- Staley, J. P. & Guthrie, C. (1998). Mechanical devices of the spliceosome, motors, clocks, springs, and things. *Cell*, 92, 315-326.
- Stamm, S., Ben-Ari, S., Rafalska, I., Tang, Y., Zhang, Z., Toiber, D., Thanaraj, T. A. & Soreq, H. (2005). Function of alternative splicing. *Gene*, 344, 1-20.
- Streitner, C., Koster, T., Simpson, C. G., Shaw, P., Danisman, S., Brown, J. W. & Staiger, D. (2012). An hnRNP-like RNA-binding protein affects alternative splicing by in vivo interaction with transcripts in Arabidopsis thaliana. *Nucleic Acids Res*, 40, 11240-11255.
- Sugliani, M., Brambilla, V., Clercx, E. J., Koornneef, M. & Soppe, W. J. (2010). The conserved splicing factor SUA controls alternative splicing of the developmental regulator ABI3 in Arabidopsis. *Plant Cell*, 22, 1936-1946.
- Tuteja, N. (2007). Abscisic Acid and abiotic stress signaling. *Plant Signal Behav*, 2, 135-138.
- Umate, P., Tuteja, R. & Tuteja, N. (2010). Genome-wide analysis of helicase gene family from rice and Arabidopsis, a comparison with yeast and human. *Plant Mol Biol*, 73, 449-465.
- Valadkhan, S. & Jaladat, Y. (2010). The spliceosomal proteome, at the heart of the largest cellular ribonucleoprotein machine. *Proteomics*, 10, 4128-4141.
- Wang, B. B. & Brendel, V. (2004). The ASRG database, identification and survey of Arabidopsis thaliana genes involved in pre-mRNA splicing. *Genome Biol*, 5, R102.
- Wang, E. T., Sandberg, R., Luo, S., Khrebukova, I., Zhang, L., Mayr, C., Kingsmore, S. F., Schroth, G. P. & Burge, C. B. (2008). Alternative isoform regulation in human tissue transcriptomes. *Nature*, 456, 470-476.

- Ward, A. J. & Cooper, T. A. (2010). The pathobiology of splicing. *J Pathol*, 220, 152-163.
- Wasilewska, A., Vlad, F., Sirichandra, C., Redko, Y., Jammes, F., Valon, C., Frei dit Frey, N. & Leung, J. (2008). An update on abscisic acid signaling in plants and more. *Mol Plant*, 1, 198-217.
- Wu, J. Y. & Maniatis, T. (1993). Specific interactions between proteins implicated in splice site selection and regulated alternative splicing. *Cell*, 75, 1061-1070.
- Xiong, L., Gong, Z., Rock, C. D., Subramanian, S., Guo, Y., Xu, W., Galbraith, D. & Zhu, J. K. (2001a). Modulation of abscisic acid signal transduction and biosynthesis by an Sm-like protein in Arabidopsis. *Dev Cell*, 1, 771-781.
- Xiong, L., Ishitani, M., Lee, H. & Zhu, J. K. (2001b). The Arabidopsis LOS5/ABA3 locus encodes a molybdenum cofactor sulfuryase and modulates cold stress- and osmotic stress-responsive gene expression. *Plant Cell*, 13, 2063-2083.
- Xiong, L., Lee, H., Ishitani, M., Tanaka, Y., Stevenson, B., Koiwa, H., Bressan, R. A., Hasegawa, P. M. & Zhu, J. K. (2002). Repression of stress-responsive genes by FIERY2, a novel transcriptional regulator in Arabidopsis. *Proc Natl Acad Sci U S A*, 99, 10899-10904.
- Xue, Y., Zhou, Y., Wu, T., Zhu, T., Ji, X., Kwon, Y. S., Zhang, C., Yeo, G., Black, D. L., Sun, H., Fu, X. D. & Zhang, Y. (2009). Genome-wide analysis of PTB-RNA interactions reveals a strategy used by the general splicing repressor to modulate exon inclusion or skipping. *Mol Cell*, 36, 996-1006.
- Yang, J., Valineva, T., Hong, J., Bu, T., Yao, Z., Jensen, O. N., Frilander, M. J. & Silvennoinen, O. (2007). Transcriptional co-activator protein p100 interacts with snRNP proteins and facilitates the assembly of the spliceosome. *Nucleic Acids Res*, 35, 4485-4494.
- Zhang, Z., Zhang, S., Zhang, Y., Wang, X., Li, D., Li, Q., Yue, M., Zhang, Y. E., Xu, Y., Xue, Y., Chong, K. & Bao, S. (2011). Arabidopsis floral initiator SKB1 confers high salt tolerance by regulating transcription and pre-mRNA splicing through altering histone H4R3 and small nuclear ribonucleoprotein LSM4 methylation. *Plant Cell*, 23, 396-411.

Chapter 21

COMPREHENSIVE ANALYSES OF ALTERNATIVE EXONS IN NEURONALLY DIFFERENTIATED P19 CELLS

Hitoshi Suzuki and Toshifumi Tsukahara*

School of Materials Science, Japan Advanced Institute of Science and Technology
Nomi, Ishikawa, Japan

ABSTRACT

Alternative splicing occurs in most human genes and contributes to protein diversity by producing multiple mRNAs from each gene. Many of these alternative isoforms are expressed in a spatio-temporal manner and play important functional roles in many biological processes including neuronal events. Here, neuronal-specific splicing was comprehensively investigated by using P19 cells. GeneChip Exon Array analyses were performed of total RNA purified from cells during different stages of the differentiation process. Nine filtering conditions were used to efficiently and readily extract alternative exon candidates. A total of 262 candidate exons (236 genes) were obtained. Semi-quantitative RT-PCR results of 30 randomly selected candidates suggested that the expression levels of 87% of the candidates were at least 2-fold different between undifferentiated and differentiated cells. Gene ontology and pathway analyses also showed that many of these 236 candidate genes played important roles in neuronal events. These results suggested that this novel method to determine alternative exons was successful and efficient. In addition to the known neuronal functions, informatics analyses demonstrated that alternative splicing events of 11 candidate genes played important cell cycle functions. These results suggest that this novel method also provides a way to determine the functional roles of previously unknown alternative splicing events.

* Corresponding Author's Email: tukahara@jaist.ac.jp.

INTRODUCTION

The human genome project estimates that only approximately 23,000 human genes encode proteins (1). Therefore, the number of human genes is only slightly higher than the number of genes found in lower eukaryotes such as worms. However, more than 90% of human genes are subjected to alternative splicing (2). This percentage of alternative exons in humans is remarkably higher than the percentage in lower eukaryotes (3). Therefore, alternative splicing is very important for higher eukaryotes. Alternative splicing changes the usage of exons and generates multiple different transcripts from a single gene. Each of these transcripts is translated into a different protein isoform. Thus, alternative splicing enhances proteomic diversity and supports increased complexity in higher eukaryotes (4). Many alternative isoforms play different or even opposed functions to other isoforms generated from the same gene. Furthermore, some of these alternative isoforms play very important roles in several biological processes (5, 6). Many alternative splicing events occur in specific spatio-temporal manners and some of these splicing events are controlled by alternative splicing regulators. Several splicing regulators (such as nPTB, Nova1, and Fox-1) and their target pre-mRNAs are characterized in neuronal tissue (7-9). Thus, in depth understandings of the alternative splicing networks are very important since these splicing networks are responsible for a large portion of the proteins that are generated from the gene expression network (10).

There are five basic types of alternative splicing: exon skipping, 5' splice site (ss) selection, 3' ss selection, mutually exclusive exons, and intron retention (4). These are the basic types of splicing and involve differential uses of splice sites. In addition to these basic types of splicing, multiple polyadenylation (poly-A) sites and multiple promoters can also result in different mRNA transcripts being generated from the same gene. 3' ss selection and/or multiple poly-A signals can result in alternative termination points for transcripts. By contrast, multiple promoters can cause transcription to initiate from different points, which results in the generation of transcripts with different 5' exons (5, 11). These different types of alternative splicing can sometimes occur together and result in complicated splicing. The differential usage of exons through alternative splicing, multiple poly-A sites, and multiple promoters results in changes to partial regions of the amino acid codes.

In the past several years, several new technologies (such as microarrays and high-throughput sequencing) were developed that permit genome-wide analyses of alternative splicing [reviewed in (12)]. There are two main types of microarrays used to detect alternative splicing events: junction arrays and exon arrays. Probes for exon arrays are designed to detect each individual exon of known or predicted genes. By contrast, probes for junction arrays are designed to detect exon-exon junctions. GeneChip Exon Arrays (supplied by Affymetrix) are accessible and convenient since the protocols used for sample preparation are very similar to those used for sample preparation for standard gene expression profiling with GeneChip Gene Arrays. In fact, a study found that Gene Arrays are comparable to Exon Arrays for gene expression profiling (13). Thus, Exon Arrays are one of the more convenient tools to investigate genome-wide alternative splicing events. In the last few years, several studies were published that analyzed alternative splicing in exon arrays. However, only a few studies found potential alternative exons in exon arrays and then validated the results by RT-PCR (14-18). Moreover, only a couple of reports validated potential alternative exons and then subjected these exons to gene ontology (GO) analyses or pathway analyses (18).

P19 cells (P19 mouse embryonic carcinoma cells) are multi-potent cells that can be differentiated into neuronal cells or cardiomyocytes depending on the concentration of retinoic acid in the medium (19, 20). We previously reported how neuronal-specific splicing of myocyte enhancer factor-2c (Mef2c) exon β is regulated in P19 cells (21). Mef2 proteins are a family of transcription factors that play important roles in muscle differentiation and synaptogenesis (22, 23). We found that Forkhead box-1 (Fox-1) promotes inclusion of Mef2c exon β via the GCAUG sequence that is located in the intron downstream of Mef2c exon β (21). These results suggested that the machinery necessary for alternative splicing is present in P19 cells and that P19 cells provide a useful model system to study alternative splicing events. There are likely to be many alternative splicing events that are regulated by transcription factors in a neuronal-specific manner, in addition to Mef2c exon β and Fox-1, when P19 cells are differentiated into neuronal cells. Furthermore, other regulators of these splicing events may be expressed in a neuronal-specific manner.

In this study, we attempted to comprehensively investigate neuronal splicing events in P19 cells. We performed GeneChip Exon Arrays using total RNA purified from undifferentiated P19 cells (Day 0), neuronal differentiated cells (Day 7), and early glial stage cells (Day 10). We generated nine filtering conditions for the probe sets based on their annotations, estimated gene expression signals, splicing indices (SIs), detection above background (DABG) values, and alternative splicing predictions. All of these conditions were based on simple parameters that avoided complicated statistics and used freely provided annotations. We extracted 262 candidate differentially alternatively spliced (DAS) exons. Furthermore, we used RT-PCR to confirm that there were greater than 2-fold differences in the expression levels of these candidates between undifferentiated and neuronal differentiated cells. The results of the RT-PCR experiments suggested that approximately 87% of the 262 candidate DAS exons were subjected to alternative splicing in a neuronal-specific manner. These results suggest that this novel method is useful to extract candidate exons. Finally, we used informatics approaches such as GO analyses, text mining, and pathway analyses to confirm that the candidate exons were important for neuronal events.

MATERIALS AND METHODS

Cell Culture and RNA Purification

P19 cells were maintained in Minimum Essential Medium, α (α -MEM; Sigma) supplemented with 10% fetal bovine serum (FBS; Sigma). To induce neuronal cell differentiation, P19 cells (1×10^5 cells/mL) were treated with 1 μ M of all trans-retinoic acid (RA) for 4 days in the 10 cm petri dish (Falcon) with α -MEM containing 10% FBS as described previously (21, 24). Total RNAs were collected from undifferentiated cells (Day 0), from cells during neuronal differentiation (Day 1, 4, 7), and from cells in the early glial stage (Day 10) using RNeasy (Qiagen).

Exon Arrays

cDNA and cRNA syntheses were performed according to the manufacturer's instructions using total RNA from undifferentiated cells, neuronal differentiated cells, and early glial stage cells. These samples were then hybridized to GeneChip Mouse Exon 1.0 ST Arrays (Exon Array, Affymetrix). Hybridization, washing, and scanning were performed with the GeneChip Fluidics Station and the GeneChip Scanner (Affymetrix) according to the manufacturer's instructions. The accession number for the Exon Array data is GSE23710.

Signal Estimations

Signal data were acquired in CEL files and were sketched, normalized, and summarized with the Probe Logarithmic Intensity Error method for exon level intensities and the IterPLIER method for gene level intensities using the Affymetrix Expression Console. In both cases, the core option, which limits the analyses to probe sets that are mapped to RefSeq and the mRNA of full length coding sequences, was used. The presence or absence of each probe set was determined by the DABG p-values. The signal estimates of the three time points with biological duplicates (Days 0, 7, and 10) were summarized with various annotations supplied by Affymetrix such as probe set positions, transcript accessions, and probe sequences.

Extraction of Candidate Exons and Genes

In an exon array, there are multiple probe sets for each exon and the probe set intensities represent the expression levels of exons. Normalized intensity (NI) values, which are the ratios of the probe set intensities to the gene intensities, were estimated for the three time points (NI_{d0}, NI_{d7}, and NI_{d10}). Splicing index (SI) values were calculated by taking the log ratios (base 2) of the averaged NIs of the biological duplicates. Probe sets that had high potential to cross-hybridize based on the annotations were removed from the analyses. The splicing patterns of the exons were predicted with UCSC BLAT (Blast Like Alignment Tool) searches (<http://genome.ucsc.edu/cgi-bin/hgBlat>). An automatic system to predict alternative splicing was established with several probe-filtering procedures (<http://www.w-fusion.com/E/>) using the results of the UCSC BLAT searches and BLAST (Basic Local Alignment Tool) searches. A detailed explanation of the extraction method for the candidate exons can be found in the Results section.

A gene was defined as expressed when more than 50% of the probe sets for that gene had DABG p-values less than 0.05 in both biological duplicates of a time point. Only genes that were defined as expressed at either Day 0 or Day 7 were used for the analyses. Two categories of expressed genes were created for the gene expression analyses. These categories were differentially expressed (DEX)-up and DEX-down. Gene expression levels in the DEX-up group at Day 7 were 2–10-fold greater than the expression levels at Day 0, and were also greater than the expression levels at Day 10. By contrast, the gene expression levels in the DEX-down group at Day 7 were 2–10-fold less than the expression levels at Day 0, and were also less than the expression levels at Day 10.

Table 1. Percentages of the relative amounts of alternative exons were compared between Day 0 and Day 7

Gene	Probeset ID	SI _{d7}	Exon type	Neural AS	Change of PCR product (%/%)	Length (nt)	Novel splicing
<i>Abi2</i>	4987109	1.67	Exon skip	+	7.44 Hold d7/d0 of 329nt	329, 146	-
<i>Aff3</i>	5291502	3.81	Exon skip	++	53.87 Hold d7/d0 of 178nt	178, 102	-
<i>Ank3</i>	5060679	-3.30	3'sss	+	4.78 Hold d0/d7 of 703nt	703, 115	-
<i>Ap1p2</i>	5477899	2.04	Exon skip	+	5.88 Hold d7/d0 of 197nt	197, 157	-
<i>Arhgef12</i>	5209805	-1.44	Alt terminator	-	1.90 Hold d0/d7 of 111nt	166, 111	-
<i>Atf2</i>	5194737	1.85	Exon skip	+	5.45 Hold d7/d0 of 335nt	335, 272, 195, 132	AB575970
<i>Atp1a3</i>	5385869	-2.71	Retained intron	++	16.88 Hold d0/d7 of 446nt	446, 109	-
<i>Bai2</i>	5286600	-2.05	Exon skip, Mutual	+	5.20 Hold d7/d0 of 173nt	209, 173, 110	-
<i>Bcl11a</i>	4969426	-1.43	5'sss, Alt terminator	+	2.82 Hold d7/d0 of 153nt	201, 153, 101	-
<i>Clt</i>	4399984	3.05	Exon skip, Mutual	+	5.93 Hold d7/d0 of 204nt	204, 168, 150, 114	AB575971
<i>Dlg3</i>	5344442	5.39	Exon skip, Mutual	++	52.04 Hold d7/d0 of 198nt	252, 198, 156, 152	AB575972
<i>Epb4.1l3</i>	5517310	2.76	Exon skip	+	5.78 Hold d7/d0 of 264nt	339, 264, 141	-
<i>Epb4.1l3</i>	4866829	2.04	Exon skip	+	2.54 Hold d7/d0 of 224nt	224, 127	-
<i>Fam113a</i>	4648309	-2.98	Exon skip, 3'sss	+	3.74 Hold d7/d0 of 102nt	363, 182, 102	AB575973
<i>Fchs2</i>	4408143	3.36	Exon skip	++	19.52 Hold d7/d0 of 235nt	235, 163	-
<i>Gnao1</i>	4489479	-1.62	3'sss, Alt terminator	+	2.12 Hold d0/d7 of 189nt	189, 124	-
<i>Kif1b</i>	4352234	2.15	Exon skip	+	5.70 Hold d7/d0 of 193nt	235, 193, 157, 115	AB575974
<i>Kif21b</i>	5130589	2.85	Exon skip	++	111.60 Hold d7/d0 of 311nt	311, 131	-
<i>Kif2a</i>	5345537	4.23	Exon skip	++	135.44 Hold d7/d0 of 231nt	231, 117	-
<i>Neo1</i>	5491604	1.77	3'sss	+	6.84 Hold d7/d0 of 153nt	153, 105	-
<i>Neo1</i>	4604088	5.90	Exon skip	++	31.00 Hold d7/d0 of 151nt	151, 118	-
<i>Plekha5</i>	5343636	1.59	Exon skip, 5'sss, 3'sss	++	17.93 Hold d7/d0 of 346nt	346, 230, 154	AB575975
	4740762	2.06					
<i>Rbm9</i>	5479607	2.04	Alt promoter	+	7.41 Hold d0/d7 of 251nt	251, 157, 134	-
<i>Rnf138</i>	4323558	2.31	Exon skip	-	1.17 Hold d7/d0 of 317nt	317, 209	-
<i>Rufy3</i>	4321833	-1.85	Exon skip, Alt promoter	+	2.01 Hold d0/d7 of 157nt	157, 103	-
<i>Snap91</i>	4960676	2.78	Exon skip	++	14.98 Hold d0/d7 of 162nt	246, 162	-
<i>Tmem219</i>	5381697	-1.91	Alt promoter, Exon skip, 3'sss	-	1.58 Hold d0/d7 of 293nt	263, 94	-
<i>Tpd52l2</i>	4473030	1.86	Exon skip	++	12.47 Hold d7/d0 of 185nt	185, 158, 116	-
<i>Trprkb</i>	5267179	-2.75	Exon skip	-	1.00 Hold d7/d0 of 104nt	188, 104	-
<i>Zfp326</i>	5023736	2.31	Exon skip, 5'sss, 3'sss,	+	2.94 Hold d0/d7 of 123nt	529, 426, 265,	AB575976-
			Retained intron			240, 123	AB575978

The differences in the relative amounts of alternative exons are indicated by: (-): less than 2-fold or the change in Day 10 is larger than Day 7 against Day 0; (+): more than 2-fold but less than 10-fold; (++): more than 10-fold. D0: percentage of Day 0; D7: percentage of Day 7; Br: percentage of brain tissue; He: percentage of heart tissue; Ki: percentage of kidney tissue; Li: percentage of liver tissue; Sp: percentage of spleen tissue; and Sk: percentage of skeletal muscle.

Semi-Quantitative RT-PCR

Total RNA was prepared from P19 cells as described above. Total RNA from mouse tissues was commercially available (Ambion, Funakoshi). cDNA was made from 2 µg of total RNA using SuperScript III (Invitrogen) and 0.5 µg of oligo dT primer in a 20 µL reaction. The cDNA was then used as the template for the PCR assays. The PCR assays were performed with GoTaq polymerase (Promega) and specific primers. The sequences of the DNA primers, the number of cycles, and the annealing temperatures that were used for the candidates are described in Supplemental Table 1. The primer sequences, the number of cycles, and the annealing temperatures for β -Actin and GluR1 were previously described (21). PCR products were analyzed with 6% polyacrylamide gels. The gels were stained with SYBR Green I (Takara), and the images were analyzed with a LAS-3000 imaging system (Fuji Film). The PCR product sequences were confirmed with a 3100 DNA sequencer (ABI). The accession numbers of the sequences for the novel alternative variants are listed in Table 3. Densitometry measurements were made with MultiGauge v 3.0 software (Fuji Film) and each experiment was performed more than three times to confirm reproducibility.

GO Analyses, Pathway Analyses, and Text Mining

GO analyses of the genes with candidate exons, the genes in the DEX-up group, and the genes in the DEX-down group were performed. Furthermore, statistically significant biological process terms were obtained with Fisher's exact tests (p -values ≤ 0.01) using Pathway Studio (Ariadne Genomics) for pathway analyses and text mining.

RESULTS

Extraction of Alternative Exon Candidates

While most exons are constitutive exons, alternative splicing occurs in more than 90% of human genes (2). Most of the probes that we analyzed were constitutive exons because we used core exon probes. Therefore, we generated nine filtering conditions to extract exons that were candidates for alternative splicing. (Condition 1) We removed the first and last probe sets of each gene because the outermost probe sets are usually designed to assess potential exons that are outside of the genes. (Condition 2) We excluded probe sets with sequences longer than 500 nucleotides (nt) because the length of most human exons is shorter than 300 nt and probe sets with longer sequences indicate that the untranslated region is included. (Condition 3) Probe sets that had high potential to cross-hybridize based on the annotations were removed from the analyses. With these three filtering conditions, 221,336 core exon probe sets for 16,661 genes were reduced to 159,552 probe sets for 15,238 genes.

(Condition 4) To identify alternative exon splicing events, we only included genes that were expressed at both Day 0 and Day 7, and exons that had detectable expression levels at one of these two time points. The reliability of gene expression levels was determined with DABG p -values of probe sets as described in the Materials and Methods. (Condition 5) For exon level

filtering, we only included probe sets that displayed significant DABG p-values ($p < 0.05$) at either Day 0 or Day 7. These filtering conditions reduced the number of candidate probe sets to 94,780. The Id7 and A-values of these probe sets were plotted, and most were clustered in the middle of the graph where $SI_{d7} = 0.00$. Therefore, these probe sets had small absolute SI_{d7} values (ABS_SI_{d7}) and were not related to alternatively spliced exons at Day 0 and Day 7. (Condition 6) We established a threshold for ABS_SI_{d7} (± 1.35 of SI_{d7} ; 2.55-fold). The SI value of a probe set represents the difference in the exon expression levels between two time points adjusted to the gene expression levels.

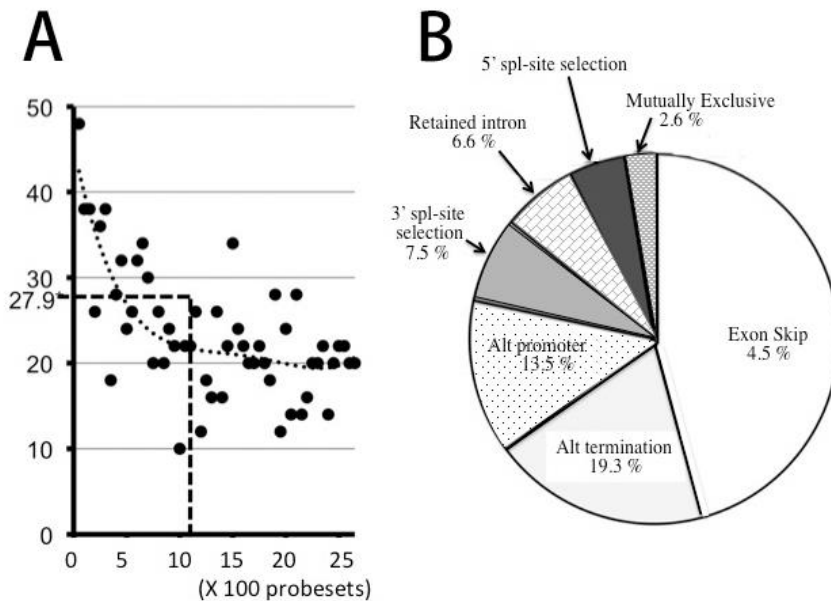


Figure 1. Statistical analyses of candidate exons. (A) Percentages of the predicted alternative exons. A total of 2,650 probe set ID sequences with ABS_SI_{d7} values that were greater than 1.0 for the eighth condition were searched. The seventh condition was excluded. It was then predicted whether or not the probe sets that were found represented alternative exons. Hole dots show percentages of alternative exons in each of the 50 probe sets from the largest ABS_SI_{d7} to approximately 1.0. The broken line shows the approximation curve. The vertical broken line shows the threshold that was used ($ABS_SI_{d7} = 1.35$). The horizontal broken line (27.9%) shows the percentage of predicted alternative exons based on the eighth condition of Table 1. (B) Percentages of the different types of alternative splicing in the DAS exons. Potential alternative splicing events were analyzed for 262 probe set ID sequences through UCSC BLAT searches.

However, the SI values of probe sets do not consider relationships between adjacent exons. Therefore, we also extracted adjacent downstream or upstream probe sets that were expressed (no cross-hybridization; DABG p-values ≤ 0.05). (Condition 7) If at least one of the adjacent probe sets had an ABS_SI_{d7} less than 0.667, then the probe sets were retained. Theoretically a probe set indicates the border between alternative and constitutive exons. Thus, because most probe sets are designed to assess constitutive exons, it is possible that there were many constitutive exons in the remaining 1,107 exons. (Condition 8) We searched the sequences of the remaining probe sets in the UCSC BLAT genome and predicted whether the probe set sequences indicated alternative exons. To this end, a probe set was categorized as an alternative

exon if the sequence aligned with the alternative exon region in the UCSC gene. In addition, if a probe set sequence aligned with multiple fragments of the mRNA sequence in the Genbank and EST databases and also skipped multiple other fragments, then the probe set was categorized as an alternative exon candidate. We also generated an automated system to evaluate the sequences and confirm the results (data not shown). Approximately 28% of the remaining probe sets (309 probe sets) were predicted to be alternative exons (Figure 1A). (Condition 9) Finally, we compared Sid7 and Sid10. We retained all of the probe sets that had a greater SI on Day 7 than on Day 10 when Sid7 was positive. We also kept the probe sets that had a lower SI on Day 7 than on Day 10 when Sid7 was negative. With all of these filtering conditions, we extracted a total of 262 probe sets. Within these 262 probe sets, all of the different types of alternative splicing events could be found (Figure 1B). These 262 probe sets were defined as DAS exons that were potentially alternatively spliced in a neuronal-specific manner.

Semi-Quantitative RT-PCR of the Candidate Exons

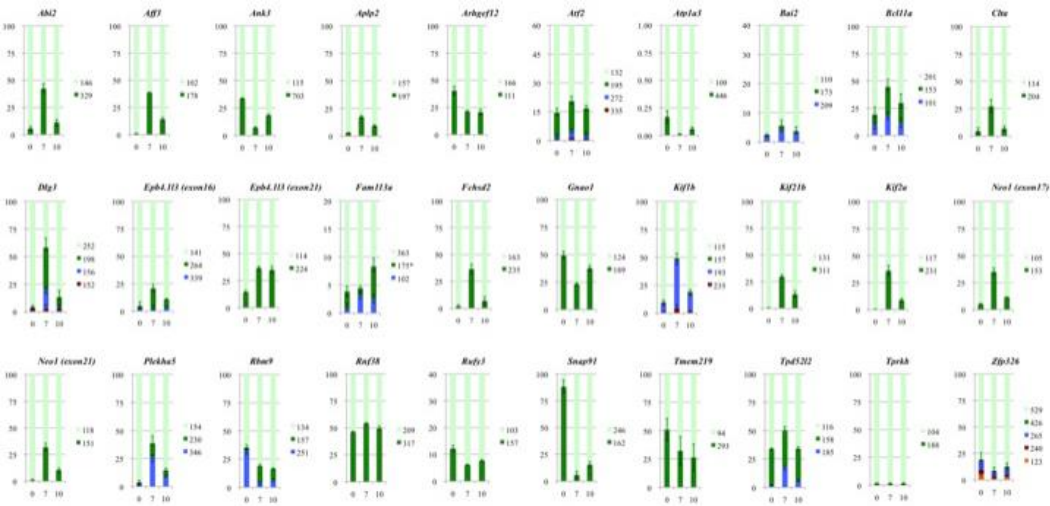


Figure 2. Semi-quantitative RT-PCR results. The results of semi-quantitative RT-PCR experiments and subsequent densitometric analyses were plotted. The amounts of each PCR product were divided by the total amounts in each lane. The bars and error bars in the histogram indicate the percentages and the standard errors. The numbers on the right side of graph indicate the lengths of the PCR products. The numbers at the bottom of the graph show the time points. Day 0 indicates the undifferentiated stage, Day 7 indicates the neuronal stage, and Day 10 indicates the early glial stage.

To examine whether the potential DAS exons actually were subjected to alternative splicing during neuronal differentiation of P19 cells, we performed semi-quantitative RT-PCR experiments. We randomly selected 30 of the 262 exons for these experiments. Typical patterns of the RT-PCR results are shown in Figure 2. Densitometric analyses showed that the alternative splicing of 13 of these 30 exons was greater than 10-fold different between undifferentiated and neuronally differentiated P19 cells (Day 0 and Day 7). In addition, the alternative splicing of 13 other exons was greater than 2-fold but less than 10-fold different

between undifferentiated and neuronally differentiated P19 cells (Figure 2 and Table 1). The increased and decreased expression levels of these 26 exons were essentially consistent with the directions that were indicated by the SI d7 values. The alternations of these 26 exons were greater on Day 7 than on Day 10. These results were consistent with the ninth condition of Table 1. Taken together, the RT-PCR data showed that neuronal cell stage-specific alternative splicing occurred in 26 of 30 exons. If extrapolated to all of the candidate exons, these data suggest that 87% of the 262 candidate DAS exons were subjected to alternate splicing in a neuronal-specific manner.

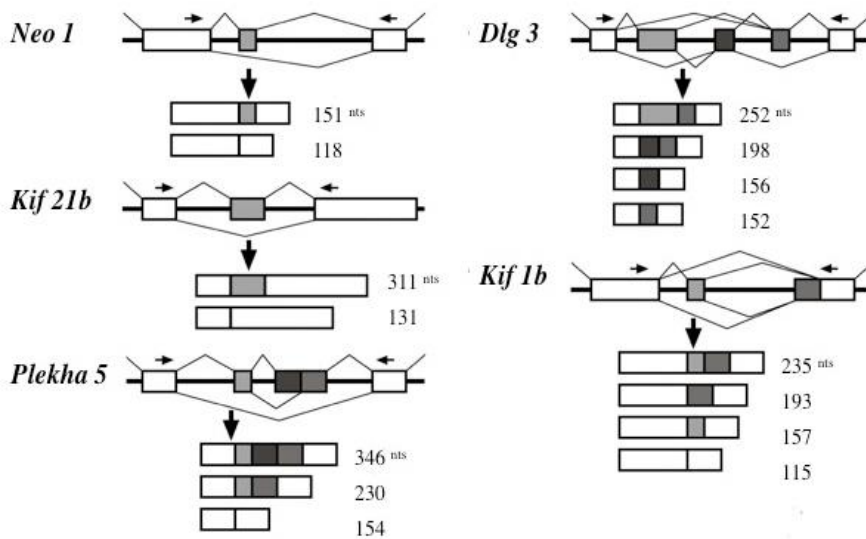


Figure 3. Schematic representations of alternative splicing. Thirty exons were randomly selected from 262 DAS exons. Five of these 30 exons (*Neo1* exon 17, *Kif21b*, *Plekha5*, *Dlg3*, and *Kif1b*) are shown in this figure. The other exons are shown in Figure 2 and Table 1. The boxes and middle lines indicate exons and introns, respectively. The gray color indicates possible alternative exons. Arrows indicate the locations of the primer annealing sites. The numbers indicate the lengths of the PCR products. The sequences of the PCR products were confirmed by sequencing analyses.

We next assessed these 26 exons by RT-PCR of adult mouse brain tissue. Seven of these 26 exons were alternative splicing events that were previously undiscovered splicing patterns in mouse (*Kif1b*, *Dlg3*, *Plekha5*, *Atf2*, *Clta*, *Fam113a*, and *Zfp326*; Figure 3 and Table 1). We found a previously unknown exon, previously unknown splice sites, and previously unknown combinations of exons. Furthermore, the *Dlg3* variant was previously unknown. The other six variants were previously detected in adult mouse brains. Interestingly, the percentages of alternative variants of 17 of these 26 exons were greater than 2-fold different between brain tissue and other tissues that showed the most similar patterns to the brain tissue (Table 1). These results suggest that more than half of the DAS exons discovered in neuronal differentiated P19 cells were changed in a neuronal-specific manner in mouse brains.

GO Analyses of the Candidate Exons

To assess whether the 262 DAS exons (from a total of 236 DAS genes) were important for neuronal events, we performed GO analyses using the Biological Process category. We obtained 72 GO terms ($p\text{-value} \leq 0.01$ in the Fisher's exact test) from the 236 DAS genes and classified the terms into ten groups: 1) neuronal-related processes; 2) differentiation and development; 3) cytoskeleton and cell adhesion; 4) signaling; 5) post-translational regulation; 6) transcription; 7) cell cycle and proliferation; 8) cellular transport; 9) apoptosis; and 10) other GO terms. The neural-related processes terms were the most abundant and accounted for 20.8% of all terms. These terms were followed by the differentiation and development terms (12.5% of all terms), the cytoskeleton and cell adhesion terms (12.5% of all terms), and the signaling terms (9.7% of all terms) (Figure 4A). We next divided the 262 DAS exons into two groups. In the first group, SId7 was greater than +1.35 (DAS-up; 153 genes) and in the second group, SId7 was less than -1.35 (DAS-down; 88 genes). The percentage of neuronal-related processes terms in the DAS-up and DAS-down groups were 17.0% and 21.2%, respectively (Figure 4B). These results suggest that the protein isoforms translated from these alternatively spliced transcripts, which either included or excluded exons, were important for neuronal events.

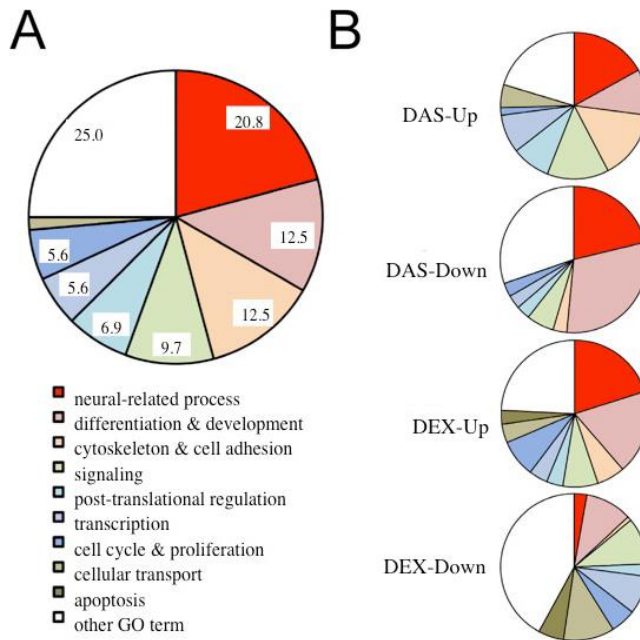


Figure 4. GO analyses. (A) GO analyses of 262 DAS exons (236 DAS genes). Seventy-two GO terms ($p\text{-value} \leq 0.01$; Fisher's T-test) were obtained from the 236 DAS genes. These terms were categorized into 10 groups, and the percentages of all GO terms are shown. (B) GO analyses at the exon and gene expression levels. The 236 DAS genes were divided into two groups. The first group contained 153 DAS-up genes that showed increased expression on Day 7 and 88 DAS-down genes that showed decreased expression on Day 7. The 153 DAS-up genes and 88 DAS-down genes were subjected to GO analyses. In addition, the gene expression levels of 1,666 DEX-up genes were increased (2–10-fold) on Day 7 relative to Day 0 and were also decreased from Day 7 to Day 10. Furthermore, the gene expression levels of 550 DEX-down genes were decreased (2–10-fold) on Day 7 relative to Day 0 and were also increased from Day 7 to Day 10. The 1,666 DEX-up genes and 550 DEX-down were subjected to GO analyses.

In addition to the exon expression profiles, the Exon Arrays also provided gene expression profiles. The Affymetrix Expression Console calculates an expression level for each gene based on the intensities of the probe sets that are designed for exons in the whole transcript. Therefore, we analyzed the gene expression profiles and obtained DEX genes in P19 neuronal cells. The expression levels of 1,666 DEX-up genes were increased more than 2-fold and less than 10-fold from Day 0 to Day 7, and were also decreased from Day 7 to Day 10 (data not shown). The DEX-up gene group had 194 GO terms. The neuronal-related processes terms were the most abundant and accounted for 20.1% of all terms. These terms were followed by differentiation and development terms, which accounted for 18.6% of all terms (Figure 4B). Therefore, these DEX-up results were similar to the DAS results.

We next further examined the relationship between expression profiles and the GO terms. DEX-up and DAS genes were primarily associated with neuronal-related processes GO terms. These results suggested that the DAS genes and the DEX-up genes were important for neuronal events. We also obtained 550 DEX-down genes. The expression levels of the DEX-down genes were decreased more than 2-fold and less than 10-fold from Day 0 to Day 7, and were also increased from Day 7 to Day 10 (data not shown). The DEX-down gene group had 107 GO terms. By contrast to the DEX-up group, the neuronal-related processes terms accounted for only 2.8% of all terms. The cellular transport terms and apoptosis terms accounted for 11.0% and 5.6% of all terms, respectively (Figure 4B). Therefore, the GO terms of DEX-down genes showed a very different pattern than the DAS, DAS-up, DAS-down, and DEX-up genes. These data suggest that the DEX-down genes were not important for neuronal events.

Extraction of Potentially Important Exons by Pathway Analyses and Text Mining

We performed pathway analyses and text mining of the DAS genes with the MedScan text mining technology provided by Pathway Studio (Ariadne Genomics). Sixty-six of 236 DAS genes were identified as well-known genes in neuronal processes because MedScan found published scientific articles demonstrating that these genes played functional roles in neuronal cells or organs (Figures 5A and 5B). By contrast, there were no (or only a few) scientific articles linking the remaining 170 genes to neuronal cells or organs. Therefore, these 170 genes were categorized as genes with no known neuronal functions. These results confirmed the GO analyses and further demonstrated that many of the DAS genes were important for neuronal events. Moreover, 49 of the 236 DAS genes had identified isoforms or splicing variants of transcripts in published scientific articles (Figure 5B). Thirty-four genes had both known neuronal functions and characterized splicing variants (Figure 5A and B). Among these 34 DAS genes, the exons of 15 genes agreed with the published variants (Table 2). Furthermore, different functions of the alternative isoforms were published for nine of these 15 genes (Table 2).

In the GO analyses (Figure 4), DEX-up genes were important for neuronal events and DEX-down genes showed no specific relation to neuronal events. We extracted 279 cell process terms closely related to the DEX-up and the DEX-down genes from neuronal cell, neuronal tissue, and neuronal organ literature using the text mining and Pathway Studio databases (Figure 5C). Ninety terms were subtracted because they were found in both the DEX-up and DEX-down groups. We assumed that the remaining 189 cell process terms were essentially

responsible for neuronal events in differentiated P19 cells and considered these terms to be neuronal cell-specific processes. We next extracted the biological relationships between these 189 cell processes and the 170 DAS genes with no known neuronal functions. Forty-eight of these 170 genes were linked to these cell processes (Figure 5A). Therefore, these 48 genes are potentially important because neuronal-specific isoforms generated by alternative splicing may play essential neuronal roles that have not yet been discovered. Furthermore, 11 of these 48 genes were involved in cell cycle-related events such as G1 and/or S phase-related cell processes (Figure 6). These results are potentially interesting because cell proliferation and cell differentiation are very closely related.

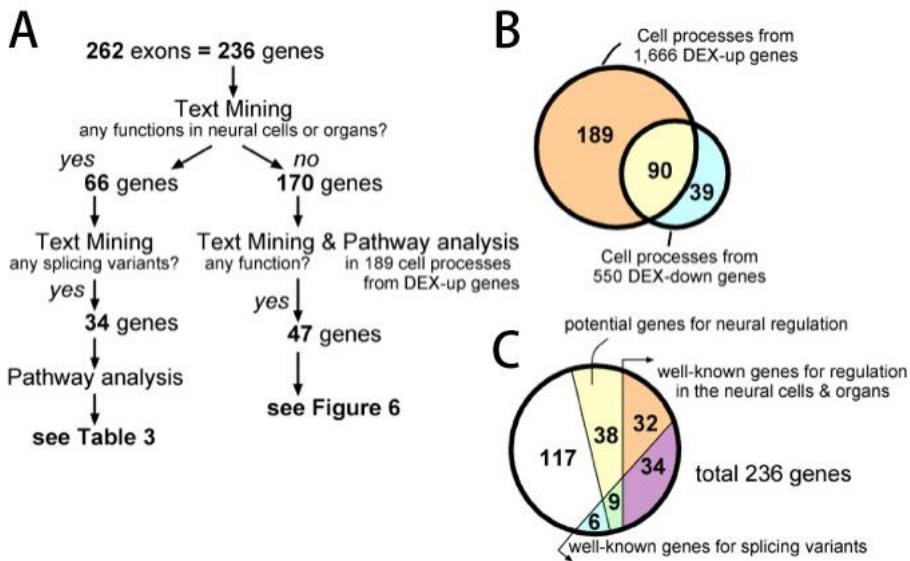


Figure 5. Pathway analyses and text mining results. (A) A flow chart representing how well-studied alternative exons and potentially important exons were extracted. Pathway Studio was used to determine whether there were published scientific reports showing that any of the DAS genes played functional roles in neuronal cells or organs. Of the 236 DAS genes, 66 genes had known neuronal functions. The approach for assessing the remaining 170 genes is described in Figure 5B. Of these 66 genes, 34 genes (40 DAS exons) were reported to have splicing variants or isoforms. These reported variants were compared with the probe set ID sequences of the 34 genes (40 exons). (B) Cell processes were used to assess the 170 genes with no known neuronal functions. The cell processes from the pathway analyses of the DEX-up and DEX-down genes were used (see Figure 4C). The orange and yellow colors indicate the numbers of biological processes from DEX-up genes and the yellow and blue colors indicate the numbers of biological processes from DEX-down genes. Because DEX-up genes were important for neuronal events and the DEX-down genes were not (Figure 4), we predicted that the 189 biological processes that were specific to the DEX-up genes were important for neuronal differentiation. Of the 170 genes with no known neuronal functions (see Figure 5A), 48 genes were matched with the 189 biological processes that were specific to the DEX-up genes. (C) The genes that were linked with neuronal functions. The 66 genes with known neuronal functions are indicated in orange and purple colors. The 48 genes that were linked with the 189 processes specific to the DEX-up genes are indicated in yellow and green colors. The 148 genes that were not linked with the DEX-up gene processes are indicated in white and blue colors. The 49 genes with reported variants or isoforms are indicated in purple, green, and blue colors.

Informatics and Experimental Analyses of Splicing Isoforms

In addition to the gene level analyses described above, we also performed informatics analyses at the exon level. Forty-nine of the 236 DAS genes were identified as isoforms or splicing variants of transcripts in published scientific articles (Figure 5C). Thirty-four genes had both known neuronal functions and characterized splicing variants (Figures 5A and B). Among these 34 DAS genes, the exons of 15 genes agreed with the published variants (Table 2). Furthermore, different functions of the alternative isoforms were published for nine of these 15 genes (Table 2). These nine genes with alternative isoforms that have different functions might be important for neuronal events. Therefore, we analyzed one of these nine genes, *Gnao1*, with semi-quantitative RT-PCR experiments. The RT-PCR results and the SI values indicated that the *Gnao1* B isoform (189 nt) was expressed in undifferentiated P19 cells and that the *Gnao1* A isoform (124 nt) was expressed in P19 neuronal cells (Tables 1 and 3). These data are consistent with a previous report showing that the *Gnao1* A isoform, but not the *Gnao1* B isoform, is required for light responses in retinal bipolar cells (30).

Of the 49 DAS genes previously identified as isoforms or splicing variants of transcripts in published scientific articles, 15 were also found in a group of 85 functional genes (Figure 5C). We analyzed the splicing of one of these 15 genes, *Abi2*, with RT-PCR experiments (Table 1). Detailed expression analyses of *Abi2* splicing variants have not yet been performed. However, the RT-PCR results and the SI values indicated that a skipping isoform (149 nt) of *Abi2* was expressed in undifferentiated P19 cells and that an inclusion isoform (329 nt) was expressed in P19 neuronal cells (Table 2). Interestingly, there is a published article about the neuronal functions of the *Abi2* gene, which found that *Abi2* is involved in learning and memory (31). To test whether *Abi2* splicing is related to these higher brain functions, we investigated alternative splicing of *Abi2* in brains. The inclusion isoform was expressed at higher levels in the frontal cortex than in other neuronal tissues. These preliminary results implied that neuronal regulation of *Abi2* alternative splicing might be important for higher brain functions.

DISCUSSION

For exon array studies, it is essential to validate that alternative exons actually exist in cells with RT-PCR experiments. Our semi-quantitative RT-PCR results found that 87% of the 262 DAS exons were altered in P19 neuronal cells (Table 1). This high validity suggests that our filtering conditions successfully and efficiently extracted DAS exons from Exon Array signals. Another important aspect of exon array studies is that the filtering conditions should be easy to modify. Our filtering conditions consisted of simple parameters and annotations that were easily modifiable. For example, if we wanted to obtain more specific potential DAS exons, we could increase the ABS_SId7 threshold in the sixth condition (Table 1). We could use a threshold of ≥ 3.0 rather than ≥ 1.35 . It is likely that higher ABS_SI thresholds would result in large alternative splicing differences in the RT-PCR experiments (Table 1). As shown in Figure 1A, 40–50% of the probe sets in the higher ABS_SId7 group indicated alternative exons in the UCSC BLAT search. We assumed that there were several alternative splicing events in DAS exons that resulted in high ABS_SI values. Indeed, the DAS exons were validated by RT-PCR to be alternatively spliced in a neuronal-specific manner. Therefore, while this threshold can be

varied, we believe that a threshold of ≥ 1.35 was an appropriate filtering condition. The approximation curve showed that the percentage of the predicted alternative exons decreased in the area with high ABS_SId7 values and the number of non-alternative exons plateaued around 1.35 (Figure 1A). This was because several of the potential DAS exons already accumulated with the ≥ 1.35 threshold. Thus, we believe that the inflection point between the decreasing range and the plateau state is a suitable ABS_SI threshold (Figure 1A).

In addition to the efficient extraction of DAS exons using simple parameters and annotations, we also suggest using accessible software such as the Expression Console and Microsoft Excel even though we generated a program to compare SI values of neighboring probe sets. Open access homology search tools such as NCBI BLAST and UCSC BLAT are useful for predicting alternative exons. These applications save time and money when performing exon array analyses. For all these reasons, we believe that our extraction method provides an efficient way to simply extract DAS exons. Therefore, our methods will be useful for projects that are attempting to perform exon array analyses. In addition, our method provides a way to comprehensively analyze alternative splicing events.

In this study, we found seven previously undiscovered splicing patterns among 30 randomly selected exons from the 262 DAS exons (Figure 2 and Table 1). These results suggest that there are several undetermined alternative splicing events. If we extrapolate our data (7 of 30 exons), it is possible that approximately 23% of the probe sets that were removed by the alternative splicing prediction searches (the eighth condition of the filtering) had novel splicing patterns. Exon and/or intron scoring with exonic splicing enhancer and exonic splicing silencer may accurately predict DAS exons. Thus, we believe that this approach may be effective when attempting to find aberrant splicing events, which are frequently observed in cancer cells. Informatics analyses showed that only 49 of the 236 DAS genes were previously reported as having alternative splicing variants (Figure 5). Although exon level annotations are a relatively new approach, it is likely that simply searching for splicing isoforms is not sufficient to comprehensively assess alternative splicing. Moreover, there are no assembled databases that assess the functional roles of splicing isoforms even though there are several databases that show the functional roles of genes and proteins. However, expression profiles of alternative splicing events are more proportional to proteomics than standard gene expression profiles. In addition, the expression profiles of alternative splicing events are starting to accumulate in databases. Therefore, informatics analyses at the exon level will become more important as better tools become available.

We found that DAS genes had identified isoforms or splicing variants of transcripts in published scientific articles. Moreover, 34 of these 49 genes were known to have functional roles in neuronal cells or organs. Furthermore, different functions of the alternative isoforms were published for nine of these 34 genes (Table 2). Gnao1 is one of these nine genes and the Gnao1 A isoform is required for the neuronal response to light (30). The Gnao1 B isoform (189 nt) was expressed in undifferentiated P19 cells and the Gnao1 A isoform (124 nt) was expressed in P19 neuronal cells (Table 1). Furthermore, Gnao1 A isoform was abundant in brain tissue. Therefore, Gnao1 alternative splicing is likely to be important for neuronal differentiation. Abi2 was one of the 49 genes that were known to have functional roles in neuronal cells and organs. In addition, Abi2 was categorized as one of the 85 functional genes. It is unknown whether the different Abi2 isoforms play different functional roles. A skipping isoform (149 nt) of Abi2 was expressed in undifferentiated P19 cells and an inclusion isoform (329 nt) was expressed in P19 neuronal cells (Table 1). The inclusion isoform was expressed at higher levels in the frontal

cortex than in other neuronal tissues. We assume that alternative splicing regulation of *Abi2* is important for higher brain functions because a previous study showed that *Abi2* is involved in learning and memory (31). Based on these *Abi2* data, we believe that the group of 85 functional genes is the most important group for future investigations because these genes are mostly uncharacterized in neuronal cells or organs.

Informatics analyses showed that many of the DAS and DEX-up genes were important for neuronal events. These results suggested that our extraction of DAS exons from the exon array signals used appropriate filtering conditions and was successful. Based on the GO biological process terms associated with the DEX-up genes, we believe that these genes are responsible for neuronal events in differentiated P19 neuronal cells. On the other hand, the GO terms associated with the DEX-down genes were not neuronal-related terms. It is easy to speculate that the protein products of DEX-down genes show reduced expression in P19 neuronal cells. Reductions or disruptions of DEX-down protein products do not permit these proteins to play any functional roles in the differentiated cells. Therefore, even if the reductions or disruptions of DEX-down protein products are important for cellular events in P19 neuronal cells, it is difficult to demonstrate that disruptions of expression contribute to biological events. Recently developed reverse genetic techniques such as RNA silencing may help to demonstrate whether there is negative linkage between a biological event and disruption of expression levels as seen with the DEX-down genes. Regardless, our GO analyses demonstrated that DEX-up gene-related functions were important for the biological events during neuronal differentiation of the P19 cells. Therefore, we searched for links between the 189 cell process terms from the pathway analyses of the DEX-up genes and the 170 DAS genes with no known functions in neuronal cells or organs (Figure 5A). Forty-eight of these 170 DAS genes were linked to these cell processes, and 11 of these 48 DAS genes play important roles in the G1/S transition. Therefore, alternative splicing of these 11 genes is likely to be important for neuronal differentiation because cell proliferation and cell differentiation are very closely related (Figure 6). There are still several unknowns in alternative splicing research. Based on the results in this chapter, we believe that our informatics approach provides a basis to determine the potentially important DAS exons.

Alternative splicing contributes to protein diversity in higher eukaryotes. It is difficult to obtain comprehensive protein expression profiles. However, the expression profiles of alternatively spliced transcripts can be determined from exon array analyses. Furthermore, the expression profiles of alternatively spliced transcripts are more proportional to proteomics than gene expression profiles. There are several open access databases for functional analyses of genes and proteins, but there are no databases for functional analyses of alternative isoforms. In this study, we converted 262 DAS exons into 236 DAS genes and then applied GO analyses and pathway analyses (Figures 4 and 5). Because there are no databases for functional analyses of alternative splicing isoforms, we analyzed whether any alternative splicing variants of the 236 DAS genes were previously reported by using a text mining method (Figure 5). Forty-nine of the 236 DAS genes had known alternative splicing variants and 34 of these 49 genes played functional roles in neuronal cells or organs. We confirmed that 13 DAS exons among the 34 DAS genes were previously reported (Table 2). Although 34 genes played known functional roles in neuronal cells and organs, there were many more genes (170) that were uncharacterized in neuronal cells and organs. Therefore, our informatics approach used to assess these 170 DAS genes provides a basis for future studies to analyze previously uncharacterized alternative splicing events.

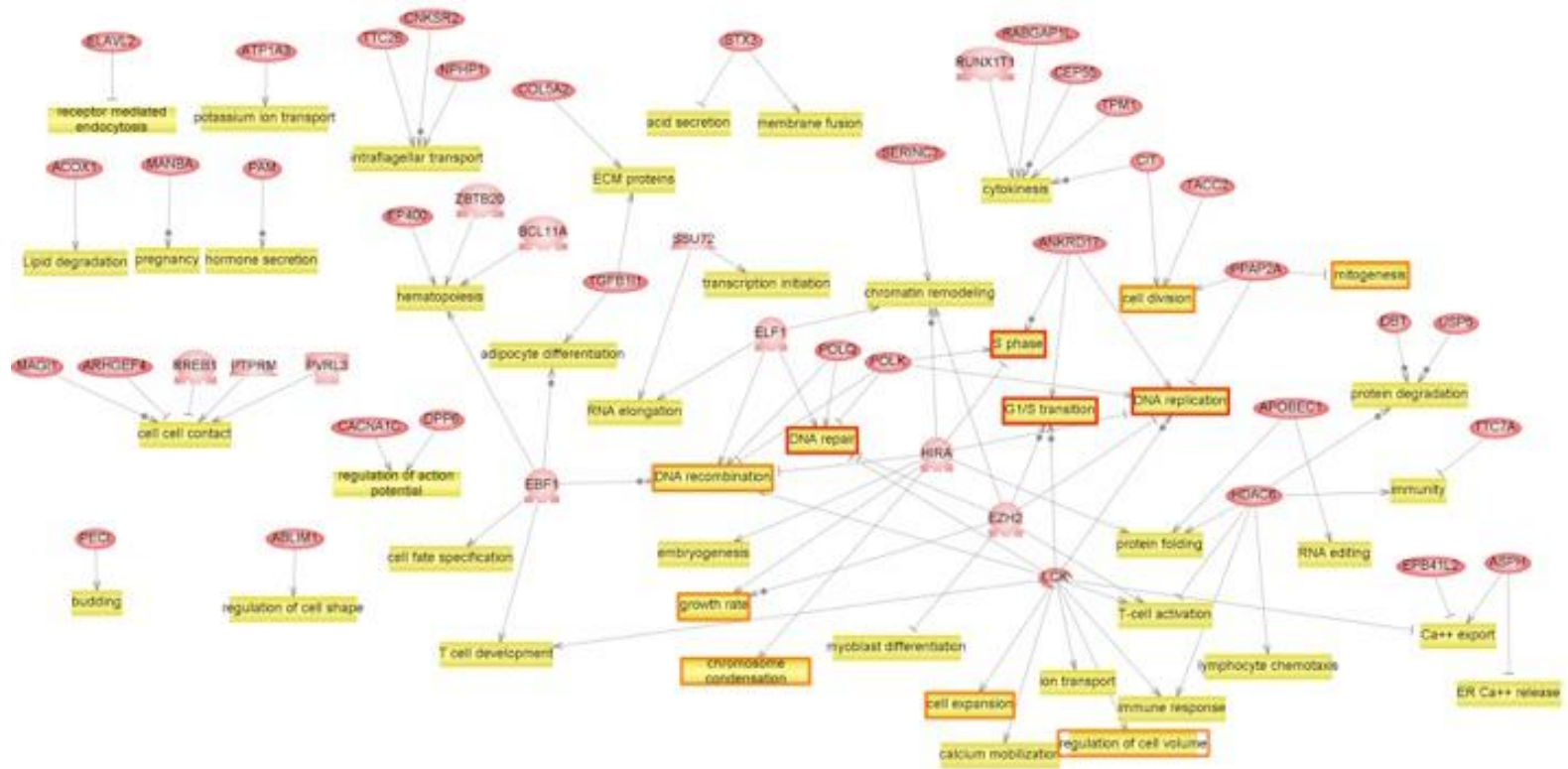


Figure 6. Pathway analyses of the 48 DAS genes that were linked to the biological processes specific to the DEX-up genes. Pathway Studio was used to link the 170 DAS genes with no known neuronal functions to the 189 biological processes that were specific to the DEX-up genes (see Figure 5A). The 48 DAS genes and their linked biological processes were shown above. Many of these genes had known cell cycle functions (indicated by orange and red colors) and specifically functions related to G1 and/or S phase (indicated by red color).

Table 2. Well-studied isoforms of the DAS genes

	"Exon"	Reported regulations	Position indicated by	Isoform name		Distinct function of the isoform	
Protein	Probeset ID	in neural cells or organs	"Probeset" - "Article"	with/without "Exon"	SI _{d7}	Function	PMID
DTNA	5566371	regulation of synapse, etc	same	□ - dystrobrevin-2/1	1.56	viral life cycle	12475945
EPB4.1	5594543	cytoskeleton assembly, etc	same	Epb 4.1 variant1/2	2.70	cytoskeleton assembly	9092575
	5012146		same	Epb 4.1 variant1/3	1.50		9092575
GNAO1	4489479	neurotransmitter uptake	same	Gnao1B/A	-1.62	neurotransmitter uptake	12077185
GPHN	4505349	synaptic transmission, etc	same	Gphn2,4,6/2,6	-1.74	Mo-cofactor synthesis	18411266
KITL	5091217	neurogenesis, etc	same	SCF248/220	1.47	proteolysis	7507105
MTAP2	5123064	cytoskeleton assembly	same	Map2s /c	13.22	neuclear localization	10383434
	5161361	neurite outgrowth, etc.	same	HMW/LMW map2	-3.24		10383434
NFASC	5439373	cell contact,	same	NF186/155	2.63	neurite outgrowth	16314110
	4436418	cell adhesion	same		2.11		16314110
PITX2	4556486	morphogenesis	same	Pitx2c/a	-1.42	morphogenesis	10662647
THRA	4582355	Glucose metabolism,	same	c-erbA □1/2	-1.79	DNA binding, etc	2901090
	4705058	neurogenesis, etc.	same	c-erbA □2/3 (&1)	1.44		2901090
ABCC5	5507375	relaxation	same	ABCC5 SV1, 2, 3/	2.62	not found	N/A
CSF1	5305033	microglia activation, etc	same	Csf1 variant3/1	-2.75		
CUGBP2	4579194	RNA splicing, etc	same	Napor-3/Etr-3	-1.59		
NRCAM	5034147	axon guidance, etc.	same	Nrcam variant1/2	11.80		
SYT7	5121401	vesicle exocytosis,	same	Syt7 variant3 (g)/1	2.58		
	4667795	endocytosis, etc	same	Syt7 variant2/1	2.60		
EPB4.1L3	5517310	organogenesis	same	not determined	2.76		
	4866829		same		2.04		
APLP2	5477899	axon cargo transport, etc	different	Aplp2 variant2/1	2.04	chondroitin modification	17371289
GAS7	4941178	cell growth, etc.	different	Gas7a/b	2.50	neurite outgrowth	15948147
RBM9	5479607	RNA splicing	different	Fox-2f/a	2.04	RNA splicing	17715393
TCF4	4886763	neuron differentiation	different	Mitf-2a/b	1.85	transcription	8631961
CACNA1B	4683815	synaptic transmission, etc	different	Cacna1b variant1/2	-1.35	not found	N/A
CLTA	4399984	neurotransmitter secretion	different	LCa variant2 (&3)/1	3.05		
DLG3	5344442	synaptogenesis	different	Dlg3 variant1/2	5.39		
TSC2	5151584	S phase, etc.	different	Tsc2 isoform 5/4	1.87		

Table 2. (Continued)

	"Exon"	Reported regulations	Position indicated by	Isoform name		Distinct function of the isoform	
Protein	Probeset ID	in neural cells or organs	"Probeset" - "Article"	with/without "Exon"	SI _{d7}	Function	PMID
EGFR	5316652	axonogenesis, etc.	different	Egfr variant2 (&4)/1	2.96		
BCL2L11	5411420	neuronal death, etc	different	undetermined	2.26	N/A	
ENPP3	4509455	cell differentiation	different		3.16		
	4522439		different		2.38		
FRAP1	4375053	synaptic plasticity, etc	different		1.78		
MBNL1	4357388	reflex	different		-2.16		
SFRS8	4922980	RNA splicing	different		1.78		
SPP1	5403558	synaptic plasticity, etc	different		-2.07		
APBB1	5391818	RNA splicing	different		-1.47		
NDRG4	4329704	cell proliferation, etc	different		-1.54		
PDE7A	5142491	memory	different		2.74		
PIM2	4775974	cell proliferation	different		3.42		

The 34 DAS genes (40 DAS exons) found with both known neuronal functions and characterized splicing variants or isoforms (see Figure 5A). The probe set ID sequences of the 40 DAS exons and the published variants of these genes were compared to determine whether the sequences and the variants were the same isoforms. In addition, published reports describing different functions of the different isoforms were searched.

B. J. Blencowe stated that there are four layers of gene expression regulation including the transcriptional network and the alternative splicing network (10). In this chapter, we obtained gene expression profiles and expression profiles of alternatively spliced transcripts. These profiles provided important information about the key regulatory factors and their potential targets in the transcriptional and the alternative splicing networks. Approximately 20 genes that were related to RNA processing were observed in the DEX-up group of genes and 14 of these 20 genes were DAS genes. These genes may play key roles in regulating neuronal alternative splicing. For instance, a previous report found that the exon-inclusion activities of RBM9, which was one of these DAS genes, were reduced in its isoform. Our results also suggested that cell cycle-related genes were important for neuronal events (Figure 6). A few of the cell cycle-specific splicing regulators were previously described. We previously reported that SRp38/NSSR1, which was detected during neuronal differentiation in P19 cells (21). SRp38/NSSR1 represses splicing via dephosphorylation in an M-phase-specific manner. In addition, nucleophosmin is also a cell cycle-dependent splicing regulator. Nucleophosmin is phosphorylated during the G1/S transition and functions as a splicing repressor. We believe that future studies should focus on these splicing regulators and their potential targets to clarify alternative splicing networks during neuronal differentiation of P19 cells.

ACKNOWLEDGMENTS

This research was supported in part by a Grant-in-Aid for Scientific Research from the Japan Society for the Promotion of Science of Japan (to T.T. and H.S.) and by a Grant-in-Aid from the Ministry of Health, Labor, and Welfare of Japan (to H.S.).

REFERENCES

- [1] International Human Genome Sequencing Consortium (2004) Finishing the euchromatic sequence of the human genome. *Nature* **431**, 931–945.
- [2] Wang E.T., Sandberg R., Luo S., Khrebtkova I., Zhang L., Mayr C., Kingsmore S.F., Schroth G.P., and Burge C.B. (2008) Alternative isoform regulation in human tissue transcriptomes. *Nature*, **456**, 470–476.
- [3] Kim E., Magen A., and Ast G. (2007) Different levels of alternative splicing among eukaryotes. *Nucleic Acids Res.*, **35**, 125–131.
- [4] Graveley, B.R. (2001) Alternative splicing: increasing diversity in the proteomic world. *Trends Genet.*, **17**, 100–107.
- [5] Black D.L. (2003) Mechanisms of alternative pre-messenger RNA splicing. *Annu Rev Biochem.* **72**, 291–336.
- [6] Matlin A.J., Clark F., and Smith C.W. (2005) Understanding alternative splicing: towards a cellular code. *Nat Rev Mol Cell Biol.* **6**, 386–398.
- [7] Coutinho-Mansfield G.C., Xue Y., Zhang Y., and Fu X.D. (2007) PTB/nPTB switch: a post-transcriptional mechanism for programming neuronal differentiation. *Genes Dev.*, **21**, 1573–1577.
- [8] Ule, J., Stefani, G., Mele, A., Ruggiu, M., Wang, X., Taneri, B., Gaasterland, T., Blencowe, B.J. and Darnell, R.B. (2006) An RNA map predicting Nova-dependent splicing regulation. *Nature*, **444**, 580–586.
- [9] Zhang, C., Zhang, Z., Castle, J., Sun, S., Johnson, J., Krainer, A.R. & Zhang, M.Q. (2008) Defining the regulatory network of the tissue-specific splicing factors Fox-1 and Fox-2. *Genes Dev.*, **22**, 2550–2563.
- [10] Blencowe B.J. (2006) Alternative splicing: new insights from global analyses. *Cell.* **126**, 37–47.
- [11] Roberts G.C., and Smith C.W. (2002) Alternative splicing: combinatorial output from the genome. *Curr Opin Chem Biol.* **6**, 375–383.
- [12] Hallegger M., Llorian M., and Smith C.W. (2010) Alternative splicing: global insights. *FEBS J.*, **277**, 856–866.
- [13] Ha K.Ch, Coulombe-Huntington J., and Majewski J. (2009) Comparison of Affymetrix Gene Array with the Exon Array shows potential application for detection of transcript isoform variation. *BMC Genomics.*, **10**, 519.
- [14] Clark T.A., Schweitzer A.C., Chen T.X., Staples M.K., Lu G., Wang H., Williams A., and Blume J.E. (2007) Discovery of tissue-specific exons using comprehensive human exon microarrays. *Genome Biol.*, **8**, R64.

- [15] Kwan T., Benovoy D., Dias C., Gurd S., Provencher C., Beaulieu P., Hudson T.J., Sladek R., and Majewski J. (2008) Genome-wide analysis of transcript isoform variation in humans. *Nat Genet.*, **40**, 225–231.
- [16] Xing Y., Stoilov P., Kapur K., Han A., Jiang H., Shen S., Black D.L. and Wong W.H. (2008) MADS: a new and improved method for analysis of differential alternative splicing by exon-tiling microarrays. *RNA*, **14**, 1470–1479.
- [17] Moller-Levet C.S., Betts G.N., Harris A.L., Homer J.J., West C.M., and Miller C.J. (2009) Exon array analysis of head and neck cancers identifies a hypoxia related splice variant of LAMA3 associated with a poor prognosis. *PLoS Comput Biol.*, **5**, e1000571.
- [18] Johnson M.B., Kawasaki Y.I., Mason C.E., Krsnik Z., Coppola G., Bogdanovic D., Geschwind D.H., Mane S.M., State M.W. and Sestan N. (2009) Functional and evolutionary insights into human brain development through global transcriptome analysis. *Neuron.*, **62**, 494–509.
- [19] McBurney, M.W. (1993) P19 embryonal carcinoma cells. *Int. J. Dev. Biol.* **37**, 135–140.
- [20] Skerjanc, I.S. (1999) Cardiac and skeletal muscle development in P19 embryonal carcinoma cells. *Trends Cardiovasc. Med.* **9**, 139–143.
- [21] Ab Hakim, N.H., Kounish, T., Alam, A.H.M.K., Tsukahara, T. and Suzuki, H. (2010) Alternative splicing of Mef2c promoted by Fox-1 during neural differentiation in P19 cells. *Genes Cells*, **15**, 255–267.
- [22] Shen, H., McElhinny, A.S., Cao, Y., Gao, P., Liu, J., Bronson, R., Griffin, J.D. and Wu, L. (2006) The Notch coactivator, MAML1, functions as a novel coactivator for MEF2C-mediated transcription and is required for normal myogenesis. *Genes Dev.* **20**, 675–688.
- [23] Barbosa A.C., Kim M.S., Ertunc M., Adachi M., Nelson E.D., McAnally J., Richardson J.A., Kavalali E.T., Monteggia L.M., Bassel-Duby R., and Olson E.N. (2008) MEF2C, a transcription factor that facilitates learning and memory by negative regulation of synapse numbers and function. *Proc. Natl. Acad. Sci. U.S.A.*, **105**, 9391–9396.
- [24] Jones-Villeneuve, E.M., McBurney, M.W., Rogers, K.A. and Kalnins, V.I. (1982) Retinoic acid induces embryonal carcinoma cells to differentiate into neurons and glial cells. *J. Cell Biol.*, **94**, 253–262.

Chapter 22

IDENTIFICATION OF GENUINE ALTERNATIVE SPLICING VARIANTS FOR RARE OR LONG-SIZED TRANSCRIPTS

Seishi Kato*

Research Institute, National Rehabilitation Center for Persons with Disabilities,
Tokorozawa, Japan

ABSTRACT

Only sequence analysis of full-length transcripts can identify genuine alternative splicing variants. However, it was difficult to obtain full-length cDNAs for rare or long-sized transcripts. Recently, we have developed a powerful method, named a vector-capping method, to construct a size-unbiased full-length cDNA library containing rare or very-long-sized cDNA clones with >10kbp inserts. The characteristic of the full-length cDNA contained in this library is that the intactness of the 5'-end capped site sequence of the cDNA can be assured by the presence of an additional dG at its 5' end. Since this full-length cDNA is derived from a single mRNA, this library enables us to perform in-depth analysis of genuine alternative splicing variants. Using the vector-capping method, we prepared full-length cDNA libraries from human retina-derived cell lines and analyzed the full sequence of the clones. As a result, we found many novel alternative-splicing variants for rare or long-sized transcripts. In this chapter, I show the examples of these variants including very-long-sized transcripts with >7kbp that were identified by us for the first time.

1. INTRODUCTION

The Human Genome Project revealed that the human genome seems to encode only 20,000~25,000 protein-coding genes (International Human Genome Sequencing Consortium,

* Corresponding Author's Email: kato-seishi@rehab.go.jp (National Rehabilitation Center for Persons with Disabilities, 4-1 Namiki, Tokorozawa, Saitama 359-8555, Japan).

2004). The analysis, including the evolutionary conservation, further cut the number of protein-coding genes to ~20,500 (Clamp et al., 2007). This number was unexpectedly small to understand the function of genes underlying the complex biological system in the cell. This issue has been solved by the discovery of diverse transcript variants for each gene. The recent research showed that diverse variants are produced from a single gene locus due to alternative promoter usage (Kimura et al., 2006), alternative splicing (AS) (Modrek et al., 2001), and alternative polyadenylation (Beaudoing et al., 2000). Initially, these variants were identified by mapping of expressed sequence tags (ESTs) to the genome. Recently, a new high-throughput sequencing technology such as mRNA-seq was applied to in-depth sequencing analysis of mRNAs isolated from various tissues and cell lines. These analyses revealed that more than 90% of human genes undergo AS (Wang et al., 2008; Pan et al., 2008).

Since the AS events vary between tissues and between developmental stages, each AS variant should be involved in the regulation of tissue-specific or cell-specific development. To fully understand the relationship between the genetic information encoded by the genome and the biological function of the cell, it is necessary to identify all transcripts including a full set of AS variants. One trial to achieve this purpose was an ENCODE project (ENCODE Project Consortium, 2011). This project adopted tiling DNA microarrays, RNA-seq, cap-analysis of gene expression (CAGE), and paired-end diTag (PET) to determine exonic regions, transcription start sites (TSSs), splice junctions, transcript 3' ends, and polyadenylation sites (Djebali, 2012). However, these protocols produce only partial sequence showing the presence of each site. Patterns of AS and alternative cleavage and alternative polyadenylation were found to be strongly correlated across tissues (Wang et al., 2008). This means that we need to know the precise combination of these alternative sites to determine the correlation between them.

To know the combination of multiple variation sites in a single transcript, the full sequence of the full-length transcript is required. The analysis of full-length transcripts can be achieved by obtaining the corresponding full-length complementary DNA (cDNA). Large-scale sequencing analyses of full-length cDNA clones were carried out using full-length cDNA libraries synthesized with the oligo-capping method (Takeda et al., 2006; Wakamatsu et al., 2009). These analyses identified a large number of AS variants including alternative TSSs.

The conventional methods for synthesizing full-length cDNAs have the following problems: (i) inability to determine whether or not the cDNA starts from a true TSS, (ii) loss of some clones due to restriction enzyme treatment during a cDNA synthesis process, (iii) difficulty in synthesizing long-sized cDNAs. Recently, we have developed a novel method, named a vector-capping method, to overcome these problems (Kato et al., 2005; Kato et al., 2011). Using this method, we prepared full-length cDNA libraries from human retina-derived cell lines. By the large-scale sequencing analysis of these libraries, we identified a lot of novel AS variants (Kato et al., 2005; Oshikawa et al., 2008; Oshikawa et al., 2011). In this chapter, I describe the examples of novel splicing variants for rare or long-sized genes we identified, and I would like to emphasize the importance of identifying a genuine AS variant derived from a single mRNA using full-length cDNA.

2. SYNTHESIS OF FULL-LENGTH cDNA USING THE VECTOR-CAPPING METHOD

We started our investigation 20 years ago by focusing on how to obtain whole human proteins. The strategy we took was to collect all proteins as a form of cDNA. At that time, the Human Genome Project had been launched and one of the projects was to analyze ESTs. However, ESTs composed of cDNA fragments were not suitable to obtain proteins. To achieve our purpose, we needed to obtain a full-length cDNA that contains an intact open reading frame (ORF) to produce the encoded protein. Thus, we developed a novel method to synthesize full-length cDNA based on replacing the cap structure of mRNA by a DNA-RNA chimeric oligonucleotide (Kato et al., 1994). This method enabled us to effectively synthesize full-length cDNAs, but it had drawbacks; it required a lot of mRNA and many reaction steps.

When improving this method, we succeeded in developing the vector-capping method shown in Figure 1 (Kato et al., 2005; Kato et al., 2011). Its process is very simple; the first-strand cDNA is synthesized using a vector primer, and then the vector-cDNA conjugate is circularized. The development of this method is attributed to the discovery of an unexpected reaction: the 3' end of the first-strand cDNA can be ligated to the 5' end of the vector primer using "RNA ligase". Furthermore, we found that the full-length cDNA possesses an additional dG at the 5' end. This additional dG is derived from dC that added to the 3' end of the first-strand cDNA by terminal deoxynucleotidyl transferase activity of reverse transcriptase only when the template mRNA has a cap structure. Thus, the presence of the additional dG at the 5' end assures the intactness of the 5'-end capped site sequence of the cDNA.

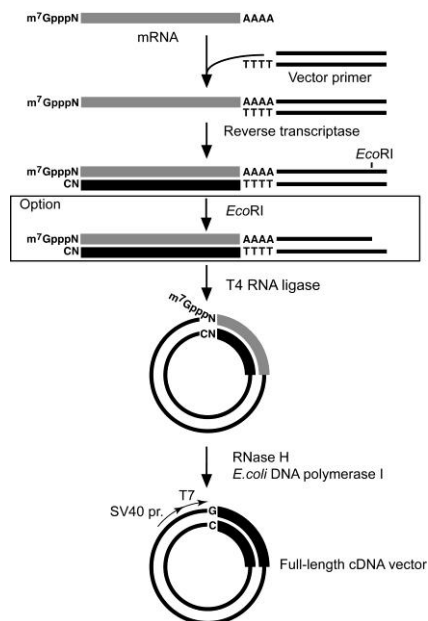


Figure 1. Schematic procedure for the vector-capping method. Several micrograms of total RNA is enough as a starting material. A vector primer has an approximately 60-nucleotide dT tail. The EcoRI digestion step can be omitted. Since the full-length cDNA vector has an SV40 promoter, the encoded protein can be produced in the mammalian cells by introducing the vector into the cells.

The full-length cDNA library constructed using the vector-capping method has the following advantages compared with those by conventional methods:

- (i) the library is composed of genuine full-length cDNA clones of > 95% content,
- (ii) we can identify full-length cDNA by the presence of an additional dG at the 5' end,
- (iii) artificial mutation or deletion seldom occurs because the procedure contains neither PCR nor the restriction enzyme treatment step,
- (iv) the library contains full-length clones for rare or long-sized genes,
- (v) we can easily identify the cDNA for an antisense gene due to the use of the vector primer.

Thus, the resulting cDNA library seems to provide us with the full-length cDNA clones ideal for identifying the alternative TSS, AS, and alternative polyadenylation.

3. ANALYSIS OF FULL-LENGTH cDNA CLONES

3.1. Retina-Derived Full-Length cDNA Libraries

We have been searching for genes responsible for retinitis pigmentosa, which is the major cause of visual impairment among patients visiting our center. To identify a novel candidate gene causing this disease, we identified genes specifically expressed in the retina by analyzing the full-length cDNA libraries that were constructed from human retinal pigment epithelium cell line ARPE-19 and human retinoblastoma cell line Y79 using the vector-capping method (Kato et al., 2005; Oshikawa et al., 2008; Oshikawa et al., 2011). We randomly picked up 100,000 clones from each library and stored them as glycerol stocks. By sequencing the 5' end of approximately 24,000 clones from each library, we identified a total of 39,643 full-length cDNA clones that were classified into 7,067 genes (Oshikawa et al., 2011). In this section, I describe the examples of novel AS variants obtained from the above libraries. Most of full-length clones and the other full-length cDNA libraries remain not fully analyzed: ARPE-19, 52,800 clones; Y79, 52,800 clones; embryonal pluripotent carcinoma cell line NT2/D1, 76,800 clones; human testis, 76,800 clones. If researchers are interested in particular genes, they may find new AS variants by full sequencing of those clones. All clones are available from RIKEN BioResource Center DNA Bank (<http://dna.brc.riken.jp/en/NRCDhuman.html>).

3.2. Characterization of Eye-Specific Genes

3.2.1. Aryl Hydrocarbon Receptor Interacting Protein-Like 1 (*AIPL1*)

Since retinoblastoma cell line Y79 is derived from cone progenitor cells (Xu et al., 2009), Y79 cells expressed various photoreceptor-specific genes. One of abundant eye-specific genes found in our Y79 full-length cDNA libraries was *AIPL1*. *AIPL1* has been identified as a gene responsible for Leber congenital amaurosis (LCA), a severe early-onset retinopathy that leads to visual impairment in infants (Sohocki et al., 2000).

Fifteen clones (C1-C15) encoding AIPL1 were fully sequenced, and their exon-intron structures were determined as shown in Figure 2. Since TSSs were distributed within the range of position -14 to position 13, the same promoter seemed to be used. We identified seven AS variants by the shift of a splicing site, skipping of exon 3, and the alternative use of a polyadenylation signal. The encoded proteins were classified into five isoforms: 384-amino acid (aa) isoform (V1 and V4, 9 clones); 345-aa isoform (V3, 2 clones); 321-aa isoform (V2 and V5, 2 clones); 270-aa isoform (V6, 1 clone); 262-aa isoform (V7, 1 clone). The difference between V1 and V4 and the difference between V2 and V5 were due to the alternative use of the polyadenylation signal. V3 showed the 58-bp downstream shift of the 3' splice site of exon 1 (designated by #7 in Figure 2), resulting in the shift of the initiation codon of the longest ORF from exon 1 to exon 2. This shift causes the loss of the N-terminal 39-aa residues. V2 and V5 lacked exon 3, resulting in the 63-aa deletion from the middle part of the protein. V6 and V7 lacked exon 6 because of the shift of the polyadenylation site, resulting in the deletion of the C-terminal 114-aa residues. Furthermore, V7 showed the 24-bp downstream shift of the 5' splice site of exon 4 (designated by #8), causing the corresponding 8-aa deletion. It should be noted that the Y79 cell-derived transcripts showed five single nucleotide polymorphisms (#1, #2, #4, #5, #6) and 2-bp deletion (#3). As a result, the clones were classified into two haplotypes, and their allelic origin was identified. Half of the clones (C3, C7, C9, C11, C12, C13, C14) were assigned to one haplotype.

The identified variants were compared with RefSeq in GenBank provided by the National Center for Biotechnology Information (NCBI), which is a collection of taxonomically diverse, non-redundant and richly annotated sequences representing naturally occurring molecules of DNA, RNA, and protein (Pruitt et al., 2009). Three RefSeqs were constructed as the transcripts of an *AIPL1* gene. RefSeq1 and RefSeq2 correspond to V1 and V2, respectively. Our collection did not contain the clone corresponding to RefSeq3 that is an AS variant skipping exon 2. In NCBI's GenBank, 13 sequences except for our 14 sequences were registered as *AIPL1* mRNA with an ORF. However, these mRNA sequences had no polyadenylation signal, whereas our clones had a canonical polyadenylation signal, AATAAA (V1-V6) and AGTAAA (V7) followed by a poly(A) tail. The mRNA sequences without a polyadenylation signal were terminated before an A-stretch (A₁₅) at position 1386 of RefSeq1 or an A-rich region (A₆CA₄CA₄CA₅) at position 2131. These clones seem to have been synthesized by priming of the oligo(dT) primer to these A-stretch sites during the first-strand cDNA synthesis. Of these 13 mRNAs, five clones correspond to V1, two clones to V2, and one clone to V6. The remaining five clones were novel variants, suggesting that there would be more variants for *AIPL1* transcripts to be identified.

AIPL1 has been reported to interact with various proteins including NUB1 (Akey et al., 2002), FAT10, FAT10nylated protein, UBA6 (Bett et al., 2012), and the catalytic subunit (alpha) of rod cGMP phosphodiesterase (PDE6A) (Kolandaivelu et al., 2009). AIPL1 is necessary for the proper assembly of functional rod PDE6 subunits (Kolandaivelu et al., 2009) that is a key phototransduction enzyme. AIPL1 is composed of three functional domains: an FKBP-like prolyl peptidyl isomerase (FKBP) domain, a tetratricopeptide (TPR) domain, and Pro-rich domain (PRD).

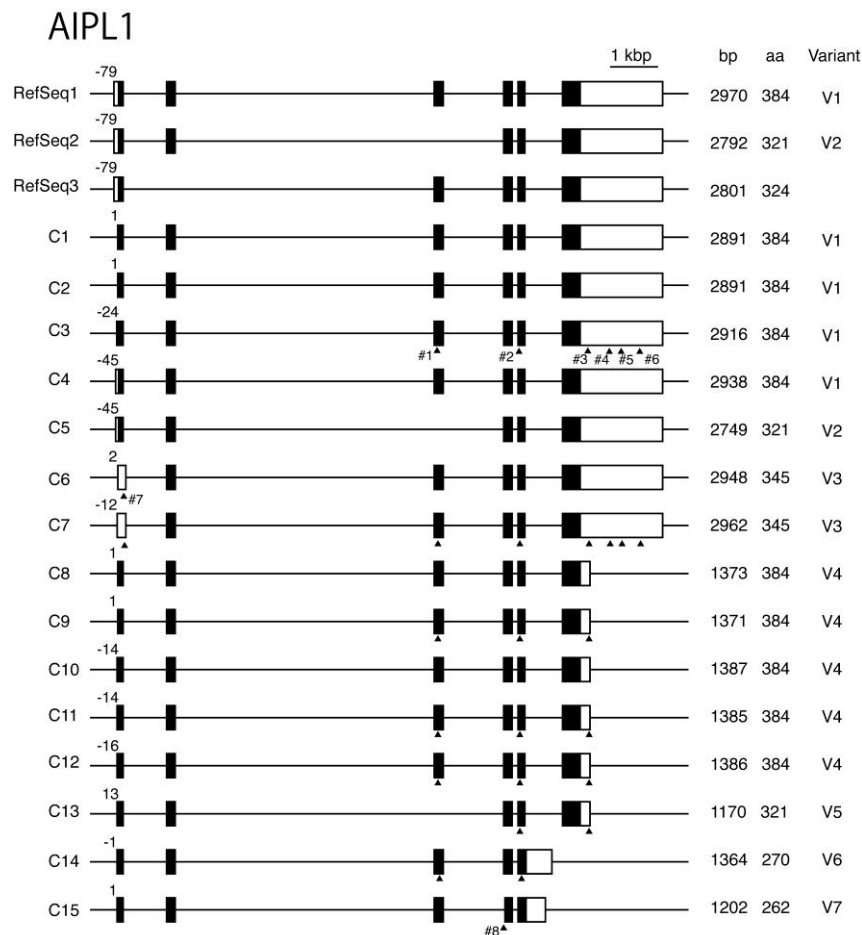


Figure 2. The exon-intron structure of alternative splicing variants for *AIPL1*. Fifteen clones (C1-C15) were classified into seven variants (V1-V7). The relative position of a transcription start site was indicated on exon 1. Arrowheads represent the following sequence variations: #1, SNP (A>G); #2, SNP (A>G); #3, 2-nt deletion (AA>--); #4, SNP (A>G); #5, SNP (A>G); #6, SNP (C>T); #7, downstream splicing site shift (58 bp); #8, downstream splicing site shift (24 bp). RefSeq1, RefSeq2, and RefSeq3 correspond to GenBank Accession No. NM_014336.3, NM_001033054.1, and NM_001033055.1, respectively. Our clones correspond to AB593062.1 - AB593067.1.

V3 encodes an isoform lacking the N-terminal 39-aa residues whose function is unclear. V6 and V7 encode an isoform lacking the C-terminal 114-aa residues corresponding to the PRD carrying a chaperone activity (Li et al., 2013). This isoform might lose function as a chaperone, because the Trp278X mutant lacking the C-terminal 107-aa residues causes LCA (Sohocki et al., 2000). V2 and V5 encode an isoform lacking the FKBP domain. To elucidate why these isoforms lacking a functional domain are produced in the cell, it is necessary to know the role of each isoform in the *AIPL1* regulation system by investigating their localization in the cell or binding activity with other proteins. Although the expression level of each minor variant is low, the total level of minor variants reaches the same level as main variants (V1 and V4), suggesting that isoforms encoded by these minor variants play their own roles in the *AIPL1*-related system. Since the EST database in GenBank (dbEST) contains novel variants, we would find more novel variants by analyzing the full-length cDNA libraries.

3.2.2. LIM Momeobox 3 (*LHX3*)

The Y79 libraries contained many clones encoding various eye-specific transcription factors. A transcription factor *LHX3* is a member of a large protein family that carries a LIM domain, a Cys-rich zinc-binding domain, and is required for pituitary development and motor neuron specification. Two variants encoding human *LHX3* (isoform a and isoform b) have been cloned from pituitary cDNA libraries (Sloop et al., 1999), and they were adopted as RefSeq for *LHX3* genes. The mutation of this gene caused combined pituitary hormone deficiency (CPHD) (Netchine et al., 2000). There is no report for *LHX3* expressed in the eye except for ESTs obtained from eye and pineal gland libraries in dbEST.

Eight clones for *LHX3* were obtained from the Y79 cDNA libraries and four variants were identified as shown in Figure 3. Five clones designated by V1 encoded a 397-aa isoform a. V2 using an alternative promoter had a different exon 1 from V1 and encoded a 402-aa isoform b. As a result, the N-terminal 25-aa sequence of isoform a was different from the N-terminal 30-aa sequence of isoform b. V3 was a novel variant containing an unspliced intron between exon 2 and exon 3, resulting in shortening of the first ORF followed by a longer ORF. The upstream short ORF encoded the N-terminal 89-aa sequence of the isoform a and the downstream long ORF encoded the C-terminal 264-aa sequence of isoform a. V4 using another promoter had a novel exon 1, and encoded a novel 386-aa isoform whose N-terminal 15-aa sequence was different from those of isoform a and b.

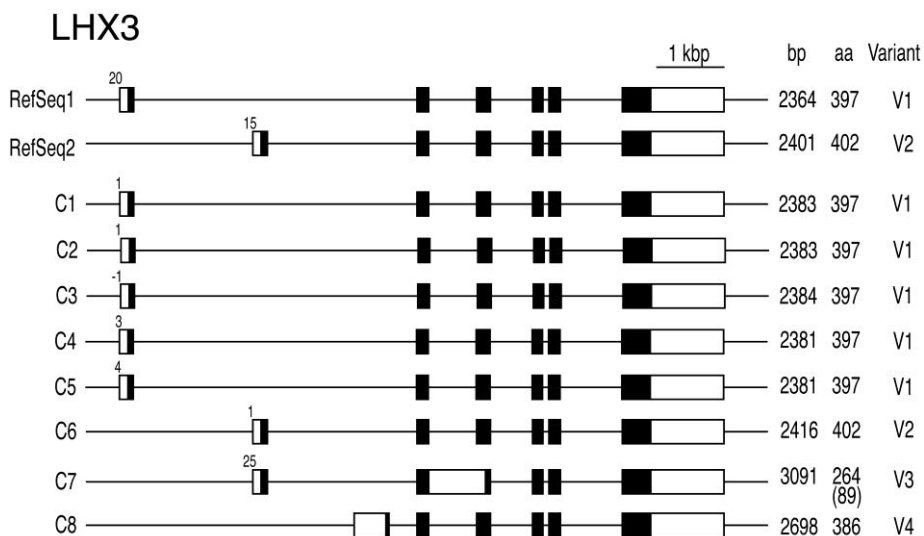


Figure 3. The exon-intron structure of alternative splicing variants for *LHX3*. Eight clones (C1-C8) were classified into four variants (V1-V4). The relative position of a transcription start site was indicated on exon 1. RefSeq1 and RefSeq2 correspond to GenBank Accession No. NM_178138.4 and NM_014564.3, respectively. Our clones correspond to AB593042.1 - AB593055.1.

V1 and V2 correspond to RefSeq1 and RefSeq2, respectively. V3 and V4 are novel variants. In GenBank two mRNA sequences containing ORF were registered except for our six sequences. These mRNAs correspond to isoform a and isoform b that were cloned from human pituitary cDNA libraries (Sloop et al., 1999). These two sequences had no polyadenylation signal maybe due to the use of a random primer in synthesizing cDNA. The dbEST contained 13 sequences: nine from eye, two from pineal gland, and two from brain. All sequences lacked

a sequence corresponding to exon 1. The cDNA libraries used for cloning these ESTs were prepared using conventional methods that contained a NotI or XhoI treatment step. Since the sequence of the full-length cDNA contained NotI and XhoI sites, the fragmentation of synthesized cDNA might occur. Although dbEST contained many ESTs isolated from the Y79 full-length cDNA libraries prepared using the oligo-capping method, there was no EST for *LHX3* in spite of an abundant gene. This was explained by the presence of an SfiI site in the *LHX3* gene, because the oligo-capping protocol contained a step using linkers with the SfiI site (Suzuki and Sugano, 2001).

LHX3 has two LIM domains, a homeodomain, and a C-terminal LHX3-specific domain. Three isoforms for human LHX3 (isoform a, isoform b, short isoform) have been reported (Sloop et al., 2001). Isoforms a and b differ in their N-terminal sequence. The N-terminal sequence of isoform b has been shown to inhibit the binding of LHX3 to DNA. Furthermore, isoform a produced a 264-aa short isoform starting at Met-134 and thus lacking LIM domains. This short isoform showed a transcription factor activity owing to the downstream region including a homeodomain. Interestingly, the longest ORF of the novel variant V3 encoded this short isoform. It is also interesting whether the N-terminal sequence of the novel isoform encoded by V4 affects the activity of LHX3. In the pituitary gland, the expression pattern of V1 and V2 differed between cell lines, suggesting that these variants play a different role in the regulation of gene expression during development of each cell type (Sloop et al., 2001). The role of these variants in development of the retina remains to be solved.

3.2.3. Neural Retina-Specific Leucine Zipper Protein (NRL)

NRL is a basic motif-leucine zipper transcription factor that plays an essential role in the differentiation of photoreceptor cells (Swaroop et al., 1992). The Y79 libraries contained seven clones encoding NRL, which were classified into five variants as shown in Figure 4. V1 corresponded to RefSeq1, and V2 had a shortened 3'-untranslated region (3'-UTR) due to the alternative use of a polyadenylation signal. V3, V4, and V5 used an alternative promoter located between exon 1 and exon 2. Although the 3'-end splice sites of a new exon 1 of these three variants were slightly different, these variants had the same ORF in exon 2, which encoded a 98-aa sequence starting from Met-140 in the isoform encoded by V1. When the expression vectors of these variants were introduced into cultured cells, the corresponding 98-aa short protein was produced (unpublished data). Since this short isoform lacked a minimal transactivation domain (Friedman et al., 2004), it showed no transcriptional activity as expected. The Leu zipper domain has been reported to interact with a CRX homeodomain (Mitton et al., 2000). The short isoform carrying only the Leu zipper domain may be involved in the regulation of complex formation through this domain.

GenBank contained three mRNAs (one corresponding to V1 and two to V2) except for the six sequences we registered. The dbEST contained six V4 sequences and one V5. Although one research group cloned the cDNA corresponding to V4, the authors could not judge whether it was a full-length or truncated one (Wistow et al., 2002). Like this case, when only one cDNA different from known ones is cloned, it is difficult to judge its intactness. Even in such case, our clone can be identified to be a full-length clone by the presence of an additional dG at the 5' end. Since the V1 sequence had SfiI sites, the Y79 cDNA libraries prepared using the oligo-capping method missed cloning the full-length cDNA for *NRL*.

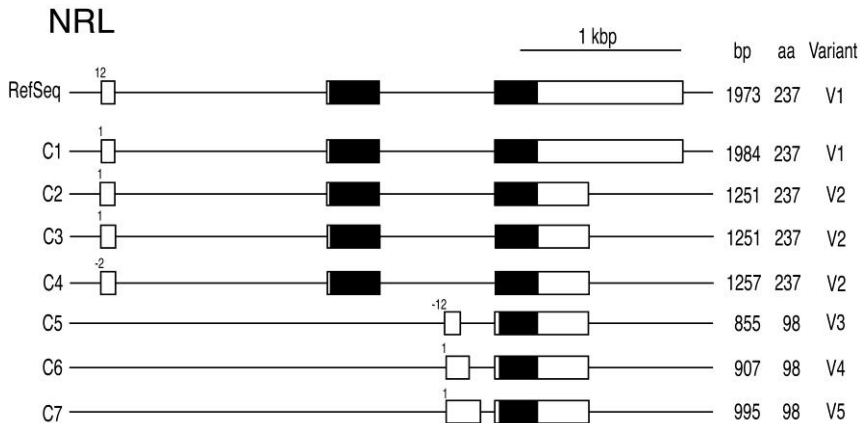


Figure 4. The exon-intron structure of alternative splicing variants for *NRL*. Seven clones (C1-C7) were classified into four variants (V1-V5). The relative position of a transcription start site was indicated on the first exon. RefSeq correspond to GenBank Accession No. NM_006177.3. Our clones correspond to AB593102.1 - AB593104.1.

3.2.4. *OTX2 Antisense RNA 1 (OTX2-AS1)*

Our libraries contained many novel non-coding RNAs including rare variants. As an example of an eye-specific non-coding RNA, we obtained five clones for *OTX2-AS1* from the Y79 libraries. These clones showed a variety of structures as shown in Figure 5. The length of cDNA varied from 303 bp of V2 to 2900 bp of V3 and the splicing pattern varied from clone to clone. Only a part of exon 1 was shared within all variants. The sequences in dbEST are also rich in variety. *OTX2-AS1* is a gene transcribed in the opposite direction at the upstream region of the locus of orthodenticle homeobox 2 (*OTX2*) that is a transcription factor involved in the development of brain and sensory organs (Alfano et al., 2005). Our Y79 libraries also contained three clones for *OTX2*, each of which is an AS variant (data not shown). The exon 1 of a variant of mouse *Otx2-as1* overlapped with the antisense strand of the exon 1 of *Otx2* (Alfano et al., 2005), but there was no human variant whose exon 1 overlapped to *OTX2*. Since all ESTs for *OTX2-AS1* in dbEST were obtained from retina cDNA libraries, this gene might be involved in the development of retina. The presence of diverse AS variants implies the complex regulation system by this gene.

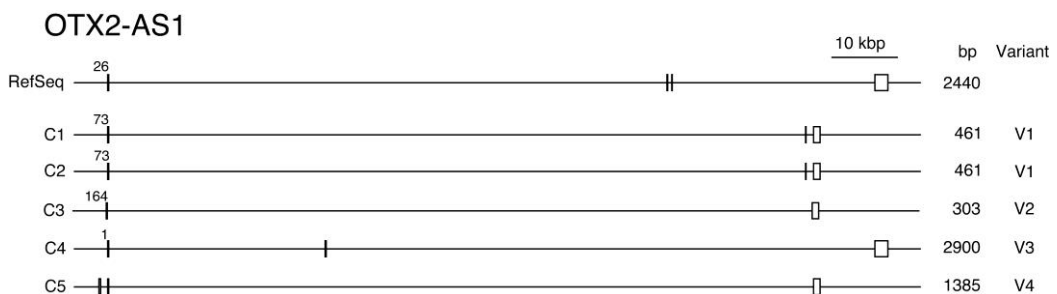


Figure 5. The exon-intron structure of alternative splicing variants for *OTX2-AS1*. Five clones (C1-C5) were classified into four variants (V1-V4). The relative position of a transcription start site was indicated on the first exon. RefSeq correspond to GenBank Accession No. NR_029385.1. Our clones correspond to AB593038.1 - AB593041.1.

3.3. Characterization of Long-Sized Genes

3.3.1. Very-Long-Sized Genes

We succeeded to clone 82 full-length cDNAs with >7kbp from the libraries prepared using the vector-capping method (Oshikawa et al., 2008; Oshikawa et al., 2011). The ARPE-19 libraries contained full-length cDNA clones encoding golgin B1 (*GOLGB1*, 11.2kbp), NEDD4 binding protein 2 (*N4BP2*, 9.7 kbp), acetyl-CoA carboxylase alpha (*ACACA*, 9.5 kbp), filamin B, beta (*FLNB*, 8.0-9.4 kbp), filamin C, gamma (*FLNC*, 9.2 kbp), spectrin, beta, non-erythrocytic 1 (*SPTBN1*, 8.4 kbp), filamin A, alpha (*FLNA*, 8.2 kbp), collagen, type V, alpha 1 (*COL5A1*, 8.1 kbp), spectrin, alpha, non-erythrocytic 1 (*SPTAN1*, 7.8 kbp), fibronectin 1 (*FN1*, 7.8 kbp), myosin, heavy chain 9, non-muscle (*MYH9*, 7.4 kbp), and agrin (*AGRN*, 7.3 kbp). The Y79 libraries contained full-length cDNAs encoding Dmx-like 1 (*DMXL1*, 12.8 kbp), *GOLGB1* (11.1 kbp), SEC16 homolog A (*SEC16A*, 9.0 kbp), *FLNA* (8.4 kbp), eyes shut homolog (*EYS*, 8.0 kbp). Out of these genes, four genes having multiple AS variants were selected and their structures were analyzed below.

3.3.2. Golgin B1 (*GOLGB1*)

GOLGB1 is a huge integral membrane protein located in Golgi, originally named giantin (Linstedt et al., 1993). Two research groups cloned approximately 10-kbp cDNA encoding a protein that reacts with autoantibody contained in sera of patients with chronic rheumatism: mRNA1, 10,295-bp cDNA encoding 3,225-aa protein (Sohda et al., 1994); mRNA2, 10,300-bp cDNA encoding 3,259-aa protein (Seelig et al., 1994). These clones were not derived from a single mRNA. The full sequence was constructed by combining the sequences of cDNA fragments. Thus, it is doubtful whether the sequence reflects the true structure of the AS variant.

Our libraries contained two full-length cDNA clones for *GOLGB1* (V1 from ARPE-19, 11.2 kbp; V2 from Y79, 11.1 kbp). The exon-intron structures of the above four clones were different as shown in Figure 6. In GenBank, RefSeqs seem to be constructed by referring to registered mRNAs including our clones: RefSeq1 to V1; RefSeq2 to mRNA2; RefSeq3 to mRNA1; RefSeq4 to V2. Although exon 1 was shared within all clones, they were all different AS variants encoding the protein with the different number of aa residues. In V1 and mRNA1, the 28-bp downstream shift of the 3' splice site of exon 2 (designated by #1 in Figure 6) caused a frame shift, and thus the initiation codon in exon 3 was used. As a result, the N-terminal sequence was shortened by 39 aa compared with the isoform for V1. Furthermore, V2 lacked a 41-aa sequence corresponding to exon 7 by exon skipping. mRNA2 lacked a 5-aa sequence by the 15-bp downstream shift of the 5' splice site of exon 7 (#2). V1 had 5-aa insertion by the 15-bp downstream shift of the 3' splice site of exon 18 (#3). The dbEST contained ESTs carrying not only these four variations but also other variations including the shift of splice site or skipping of exon 4, 6, 10, 11, 12, 15-21, suggesting the presence of diverse AS variants of *GOLGB1*.

GOLGB1 is an integral membrane protein involved in linkage between a Golgi membrane and a COPI vesicle (Sönnichsen et al., 1998). This protein has no N-terminal secretory signal sequence, but has a C-terminal transmembrane domain. Most of the cytoplasmic part is composed of a coiled-coil structure in which the AS variants had deletion or insertion. This structure is thought to be involved in regulation of retrograde trafficking to the endoplasmic reticulum in Golgi apparatus through binding of small GTPase such as Rab6 and Rab1 (Rosing

et al., 2007). Thus, each AS variant may play a role in the regulation of this trafficking. To elucidate the detailed mechanism, further investigation is necessary using these AS variants.

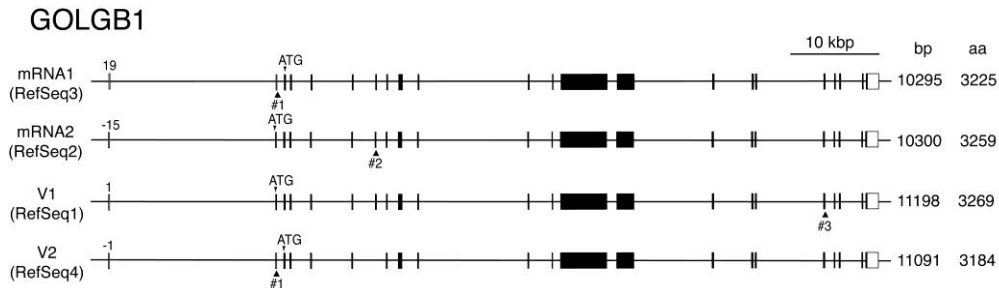


Figure 6. The exon-intron structure of alternative splicing variants for *GOLGB1*. C1 was cloned from the ARPE-19 cDNA library and C2 from the Y79 cDNA library. The relative position of a transcription start site was indicated on exon 1. Arrowheads represent the following sequence variations: #1, the downstream shift of the 3' splice site (28 bp); #2, the downstream shift of the 5' splice site (15bp); #3, the downstream shift of the 5' splice site (15 bp). mRNA1 and mRNA2 correspond to GenBank Accession No. D25542.1 and X75304.1, respectively. Our clones correspond to AB371588.1 and AB593126.1.

3.3.3. Filamin A, Alpha (*FLNA*)

FLNA was most abundantly found in our libraries as a long-sized gene with >7kbp. *FLNA* is an actin-binding protein involved in change in cell shape and migration through crosslinking of actin filaments and linking actin filaments to membrane glycoproteins. ARPE-19 and Y79 libraries contained eight clones (7.3 – 8.2 kbp) and one clone (8.4 kbp), respectively. These clones were classified into four variants as shown in Figure 7. V1 and V2 were main components in ARPE-19 cells. V2 lacked exon 30, resulting in deletion of an 8-aa sequence. V3 lacked exon 38-41 because of AS between the middle splice site in exon 37 and the middle splice site in exon 42. Y79-originated V4 started from a 135-bp upstream TSS compared with V1, resulting in the generation of a novel initiation codon that caused 27-aa extension of the N-terminal sequence.

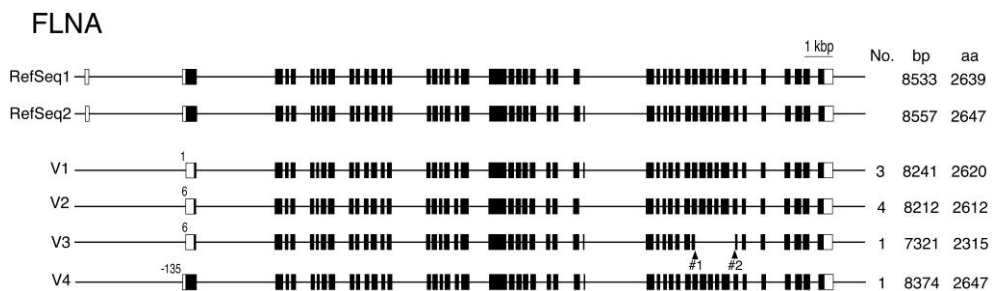


Figure 7. The exon-intron structure of alternative splicing variants for *FLNA*. Eight clones obtained from ARPE-19 were classified into three variants (V1-V3). V4 was obtained from Y79. The relative position of a transcription start site was indicated on the first exon. "No." represents the number of obtained clones. Arrowheads represent the following sequence variations: #1, the 96-bp upstream shift of the 3' splice site of exon 37; #2, the 72-bp downstream shift of the 5' splicing site of exon 42. RefSeq1 and RefSeq2 correspond to GenBank Accession No. NM_001456.3 and NM_001110556.1, respectively. Our clones correspond to AB191259.1 - AB191260.1, AB371574.1- AB371579.1 and AB593010.1.

GenBank contained 3 mRNAs with a full ORF except for the nine clones we registered. The sequence of mRNA1 (X53416) was constructed by combining seven fragments cloned from a human endothelial cell (Gorlin et al., 1990). mRNA2 (AK090427) originated from a single mRNA and had the same initiation codon as V4, but lacked exon 30. Although this clone seems to be near full-length cDNA, the registrant regarded it as a truncated clone maybe because of no evidence for intactness of the 5' end of the cDNA. mRNA3 (GU727643) was synthesized with RT-PCR based on RefSeq. RefSeq1 corresponding to mRNA1 has exon 1, which uses an upstream alternative promoter. Its ORF starts from the same initiation codon with V4. Our nine clones had no exon 1. There were 14 sequences having exon 1 in dbEST. RefSeq2 was constructed based on our clone V1 except for exon 1. The dbEST contained a sequence (CN421698) that has the same deletion as V3.

FLNA has a rod-like structure composed of 24 repeats of the beta-pleated sheet unit: an actin-filament binding domain (Rod1, repeats 1-15), a partner protein binding domain (Rod2, repeats 16-23), a self-assembly domain (repeat 24), and a hinge linking the domains (Hinge-1 and Hinge-2) (Nakamura et al. 2011). V2 lacked the 8-aa residues that were located in the last repeat 15 of Rod1. V3 lacked the 114-aa sequence that was the part of repeats 18 and 19 of Rod2. Since these repeats are known to bind to several partner proteins, these isoforms may lose a function borne by the corresponding repeat.

3.3.4. *Filamin B, Beta (FLNB)*

FLNB and FLNC as well as FLNA are a member of a filamin family. The ARPE-19 libraries contained four *FLNB* clones (8-9 kbp) and one *FLNC* clone (9156 bp). The *FLNC* clone corresponded to RefSeq (data not shown). All clones for FLNB were different AS variants as shown in Figure 8. Four RefSeqs are constructed based on our four clones. They had a similar TSS. V4 had a total of 47 exons and the other three variants lacked exon 26 (93 bp, 31 aa). V2 lacked 11-aa residues due to the 33-bp upstream shift of 3' splice site of exon 31 (designated by #1 in Figure 8). V1 and V4 had a shorter exon 47 due to the use of an alternative polyadenylation signal. GenBank contained two mRNA sequences except for our clones. These two sequences corresponding to V1 were constructed by combining the sequences of cDNA fragments (Takafuta et al., 1998; Xu et al., 1998). Xu et al. (Xu et al., 1998) have reported near full-length cDNA clone (9.5 kb) for V3 derived from a single mRNA. The dbEST contained four clones possessing a shortened exon 31 found in V2 and two clones possessing exon 26 found in V4.

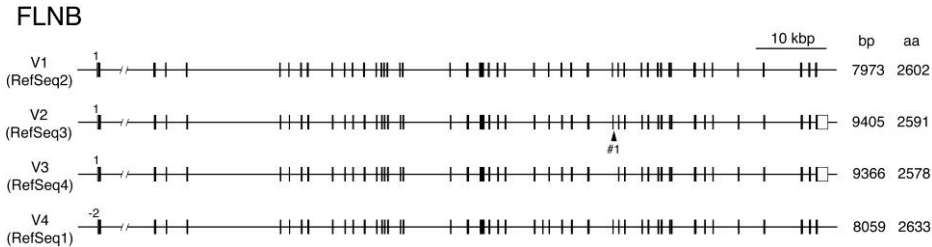


Figure 8. The exon-intron structure of alternative splicing variants for *FLNB*. Four clones were all different variants. Four RefSeqs are constructed based on our four clones as shown in parenthesis. The relative position of a transcription start site was indicated on the first exon. Arrowhead #1 represents the 33-bp upstream shift of the 3' splice site of exon 31. Our clones correspond to AB191258.1 and AB371580.1- AB371582.1.

FLNB has a domain structure similar to FLNA. V4 encoded an isoform possessing 31-aa insertion in the middle part of repeat 31 due to the insertion of exon 26. The isoform encoded by V3 lacked Hinge-1 of 24-aa residues corresponding to exon 31. V2 encoded an isoform lacking the 11-aa C-terminal half of Hinge-1. The deletion of these aa sequences might affect the function of each isoform through the change of binding ability to the partner proteins as well as FLNA. For example, van der Flier et al. showed that an FLNB fragment lacking a part of repeat 19-20 or C-terminal repeat 24 obtained using RT-PCR had a different binding ability to integrin beta subunit (van der Flier, 2002). Furthermore, they showed that the expression pattern of these variants varied from tissue to tissue and during myogenesis. We have to keep in mind that this kind of experiment using RT-PCR shows the expression level of only a partial sequence of transcript and the expression pattern does not reflect the change of the full-length transcript.

3.3.5. Eyes Shut Homolog (*EYS*)

EYS is an extracellular matrix specifically produced in photoreceptor cells (Abd El-Aziz, 2008; Collin et al., 2008). Recently, we showed that one-third of Japanese patients with retinitis pigmentosa had founder mutations in the *EYS* gene. RefSeq1 for the *EYS* gene comprises 43 exons as shown in Figure 9 and the length of mRNA is 11 kb. In GenBank, there are two short RefSeqs terminating by exon 11. RefSeq2 has a long 3'-UTR. RefSeq3 uses an alternative promoter located between exon 2 and exon 3, resulting in the formation of a new exon 1. The Y79 libraries contained two clones only for short variants. Although V1 was a very-long-sized clone with an insert of 7,898 bp, it terminated by exon 11 containing a long 3'-UTR and encoded the N-terminal 594-aa sequence as well as RefSeq2. The exon 11 of RefSeq2 was split due to splicing. V2 terminated by exon 4 and encoded a short isoform of 318 aa. The *EYS* protein comprises 27 EGF-like domains and 5 laminin G-like domains. Thus, the isoform encoded by V1 terminated at the middle of the sixth EGF-like domain, and the isoform for V2 at the middle of the third EGF-like domain. The function of these short forms of the *EYS* protein remains to be solved.

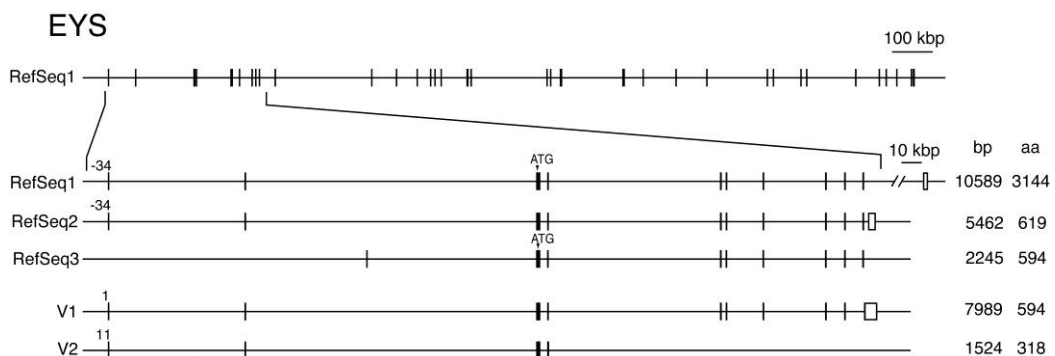


Figure 9. The exon-intron structure of alternative splicing variants for *EYS*. Two variants were cloned from the Y79 cDNA library. The relative position of a transcription start site was indicated on the first exon. RefSeq1, RefSeq2 and RefSeq3 correspond to GenBank Accession No. NM_001142800.1, NM_001142800.2, and NM_198283.1, respectively. Our clones correspond to AB593114.1 and AB593112.1.

4. FULL-LENGTH cDNA LIBRARIES DERIVED FROM OTHER SPECIES

The power of the vector-capping method was first demonstrated by the transcriptome analysis of budding yeast. Miura et al. performed a large-scale analysis of full-length cDNA libraries prepared from budding yeast cells growing exponentially in a minimal medium and meiotic cells (Miura et al., 2006). They identified 11,575 TSSs associated with 3,638 genes, suggesting that most yeast genes have two or more TSSs. They also identified 45 previously undescribed introns, including those spliced alternatively. Furthermore, they found 667 transcripts in the intergenic region and 367 transcripts derived from antisense strands of known genes. These results suggest that many genes remain unidentified even in an intensively analyzed simple organism such as budding yeast.

Since the vector-capping method was published in 2005 (Kato et al., 2005), it has been adopted by various research projects to construct cDNA libraries from various tissues of various species: plants such as burma mangrove (Miyama et al., 2006), miniature tomato (Aoki et al., 2010), Chinese cabbage (Abe et al., 2011), rubber tree (Suzuki et al., 2012); mammals such as macaque monkey (Osada et al., 2009), pig (Uenishi et al., 2012), common marmoset (Tatsumoto et al., 2013); parasites such as *Haemaphysalis* (Zhou et al., 2006), *Echinococcus* (Watanabe et al., 2007), *Babesia* (Aboge et al., 2008); hagfish (Uchida et al., 2010); *Bombyx mori* nucleopolyhedrovirus (Katsuma et al., 2011). In many cases, it seems to have been difficult to obtain a large amount of starting material. The vector-capping method requires only several micrograms of total RNA. This seems to be one reason why this method was adopted to prepare the libraries from the above samples. The cloning ability of a long-sized cDNA was confirmed by the cloning of very-long-sized genes (9.1kb and 9.8kb) that encode egg case silk from a wasp spider (Zhao et al., 2006).

5. PROBLEMS ON IDENTIFICATION OF AS VARIANTS

5.1. AS Variants of Rare or Long-Sized Genes

In the above sections, I have described some problems in identifying AS variants using the conventional methods. A serious problem occurred in the case of a long-sized gene. The combination of alternative promoter usage, multiple alternative splicing sites, and alternative polyadenylation can produce diverse forms of transcripts. The partial sequence analyses using RT-PCR or RNA-seq do not disclose this combination. To determine the precise structure of the AS variant, it is necessary to determine the full sequence of a single mRNA. One solution for this requirement is to determine the full sequence of a full-length cDNA derived from a single mRNA.

Another problem is related to the intactness of the full-length cDNA. In the case of an abundant gene, many cDNA clones can be obtained. If these cDNAs are shown to start at the similar site by comparing their 5'-end sequences, we could regard them as a full-length or near full-length cDNA having a capped site sequence. However, a rare gene may give only one cDNA clone, thus we cannot judge the intactness of this cDNA. The same problem occurs in the case of very short or very long genes. It may be difficult to judge whether the cDNA are derived from intact mRNA or degraded mRNA. The vector-capping method solves these

problems. We can judge the intactness of the cDNA by inspecting the presence of the additional dG at the 5' end of the cDNA.

5.2. Synthesis of Full-Length cDNA

It has been difficult to obtain a full-length cDNA for a rare or long-sized gene using conventional methods. Here, the problems are shown with regard to each step of cDNA synthesis.

(1) Oligo(dT) Priming

The conventional methods usually use an oligo(dT) primer of ~20 nt to synthesize the first-strand cDNA. When mRNA has a short A stretch, the oligo(dT) primer can accidentally hybridize to this site and be used for cDNA synthesis, resulting in missing the downstream part of mRNA to a poly(A) tail. A good example is an *AIPL1* gene shown in section 3.2.1. We observed several other examples of such mispriming (data not shown). The vector-capping method uses a vector primer possessing approximately 60-nt dT at one end of the vector. The long dT tail may rarely prime a short A stretch in mRNA.

(2) Reaction Conditions

According to our experience, the amount of template mRNA, reverse transcriptase and substrate nucleotides seem to be essential factors that are related to biases by the expression level or size of mRNA. Usually the first-strand cDNA synthesis is carried out using several micrograms of poly(A)⁺RNA. In these reaction conditions, most reverse transcriptase and substrate nucleotides seem to be consumed to synthesize cDNA mainly from abundant or short-sized mRNAs, causing biases by the expression-level and size of mRNA. In the vector-capping method, total RNA is used as a template in place of poly(A)⁺RNA to synthesize the first-strand cDNA under the same reaction conditions. Thus, the amount of enzyme and substrate might be enough to synthesize cDNA from rare or long-sized mRNAs. Omitting mRNA purification steps also may help to reduce these biases.

(3) PCR Step

Some conventional methods including the oligo-capping method contain a PCR step in the procedure for preparing the cDNA library. The amplification step by PCR may cause bias by the expression level and the size of mRNA. In fact, when the full-length cDNA libraries prepared from monkey liver and kidney by the oligo-capping method were compared with those prepared by the vector-capping method, the redundancy of the vector-capped libraries is lower than those of the oligo-capped libraries (Osada et al., 2009). To synthesize rare or long-sized cDNAs, the PCR step should be avoided.

(4) Restriction Enzyme Treatment

The conventional methods contain a linker attachment step, in which a oligonucleotide linker with a restriction enzyme site (e.g., NotI, EcoRI, SalI, XhoI, SfiI et al.) are ligated to the double-stranded cDNA and then after cutting by restriction enzyme the cDNA are introduced into a vector. If the cDNA has the same restriction enzyme site as the linker, it is difficult to

obtain full-length cDNA or any cDNA in some cases. The examples are shown in the above sections on *LHX3* and *NRL*. Many clones registered in dbEST seem to be a truncated cDNA that was generated due to this step.

(5) Size Fractionation

Some protocol contains a size fractionation step to remove short cDNA fragments. This step should be avoided because there are many short transcripts having a poly(A) tail. We observed such short full-length cDNAs with < 100 bp (data not shown).

5.3. Vector-Capping Method

The vector-capping method solves all the above problems. Thus, this will be the most effective method to synthesize genuine full-length cDNAs at present. However, this has one limitation. It is difficult to obtain full-length cDNA clones from a low-quality RNA sample containing highly degraded mRNA, because this protocol does not contain a step for experimentally selecting full-length cDNAs, such as a cap-dependent linker ligation in the oligo-capping method. In addition, it requires a lot of labor and cost to search novel AS variants from the vector-capped libraries. This is the case particularly when the target cell expresses genes with low complexity. In that case, we should use a subtraction or normalization protocol together. If the target gene has been decided, we may isolate in advance target cDNA using a probe for the target gene.

CONCLUSION

Here I have demonstrated that the vector-capping method provides us with a high-quality cDNA library composed of genuine full-length cDNA clones derived from a single mRNA and that the obtained clones can be used to effectively identify AS variants. This library contains many full-length cDNA clones for rare or long-sized genes whose intactness is guaranteed. By analyzing these clones, we can identify novel AS variants for rare or long-sized genes that have been difficult to obtain using conventional methods. These results suggest that comprehensive, in-depth analysis of full-length cDNA clones isolated from the vector-capped libraries is the most effective way to identify an entire set of AS variants. Furthermore, these full-length cDNA clones can be used as a resource for producing the encoded proteins. I hope that the vector-capping method will be widely used for analyzing full-length AS variants derived from various tissues of various species.

REFERENCES

- [1] Abd El-Aziz MM, Barragan I, O'Driscoll CA, et al. EYS, encoding an ortholog of *Drosophila* spacemaker, is mutated in autosomal recessive retinitis pigmentosa. *Nat Genet.* 2008;40(11):1285-1287.

- [2] Abe H, Narusaka Y, Sasaki I, Hatakeyama K, et al. Development of full-length cDNAs from Chinese cabbage (*Brassica rapa* Subsp. *pekinensis*) and identification of marker genes for defense response. *DNA Res.* 2011;18(4):277-289.
- [3] Aboqe GO, Jia H, Terkawi MA, et al. Cloning, expression, and characterization of *Babesia gibsoni* dihydrofolate reductase-thymidylate synthase: inhibitory effect of antifolates on its catalytic activity and parasite proliferation. *Antimicrob Agents Chemother.* 2008;52(11):4072-4080.
- [4] Akey DT, Zhu X, Dyer M, et al. The inherited blindness associated protein AIPL1 interacts with the cell cycle regulator protein NUB1. *Hum Mol Genet.* 2002;11(22):2723-2733.
- [5] Alfano G, Vitiello C, Caccioppoli C, et al. Natural antisense transcripts associated with genes involved in eye development. *Hum Mol Genet.* 2005;14(7):913-923.
- [6] Aoki K, Yano K, Suzuki A, et al. Large-scale analysis of full-length cDNAs from the tomato (*Solanum lycopersicum*) cultivar Micro-Tom, a reference system for the Solanaceae genomics. *BMC Genomics.* 2010;11:210.
- [7] Beaudoing E, Freier S, Wyatt JR, Claverie JM, Gautheret D. Patterns of variant polyadenylation signal usage in human genes. *Genome Res.* 2000;10(7):1001-1010.
- [8] Bett JS, Kanuga N, Richet E, et al. The inherited blindness protein AIPL1 regulates the ubiquitin-like FAT10 pathway. *PLoS One.* 2012;7(2):e30866.
- [9] Clamp M, Fry B, Kamal M, et al. Distinguishing protein-coding and noncoding genes in the human genome. *Proc Natl Acad Sci U S A.* 2007;104(49):19428-19433.
- [10] Collin RW, Littink KW, Klevering BJ, et al. Identification of a 2 Mb human ortholog of *Drosophila* eyes shut/spacemaker that is mutated in patients with retinitis pigmentosa. *Am J Hum Genet.* 2008;83(5):594-603.
- [11] Djebali S, Davis CA, Merkel A, et al. Landscape of transcription in human cells. *Nature.* 2012;489(7414):101-108.
- [12] ENCODE Project Consortium. A User's Guide to the Encyclopedia of DNA Elements (ENCODE). *PLoS Biol.* 2011;9(4):e1001046.
- [13] Friedman JS, Khanna H, Swain PK, et al. The minimal transactivation domain of the basic motif-leucine zipper transcription factor NRL interacts with TATA-binding protein. *J Biol Chem.* 2004;279(45):47233-47241.
- [14] Gorlin JB, Yamin R, Egan S, et al. Human endothelial actin-binding protein (ABP-280, nonmuscle filamin): a molecular leaf spring. *J Cell Biol.* 1990;111(3):1089-1105.
- [15] International Human Genome Sequencing Consortium. Finishing the euchromatic sequence of the human genome. *Nature.* 2004;431(7011):931-945.
- [16] Kato S, Ohtoko K, Ohtake H, Kimura T. Vector-capping: a simple method for preparing a high-quality full-length cDNA library. *DNA Res.* 2005;12(1):53-62.
- [17] Kato S, Oshikawa M, Ohtoko K. Full-length transcriptome analysis using a bias-free cDNA library prepared with the vector-capping method. *Methods Mol Biol.* 2011;729:53-70.
- [18] Kato S, Sekine S, Oh SW, Kim NS, et al. Construction of a human full-length cDNA bank. *Gene.* 1994;150(2):243-250.
- [19] Katsuma S, Kang W, Shin-i T, et al. Mass identification of transcriptional units expressed from the *Bombyx mori* nucleopolyhedrovirus genome. *J Gen Virol.* 2011;92(Pt 1):200-203.

- [20] Kimura K, Wakamatsu A, Suzuki Y, et al. Diversification of transcriptional modulation: large-scale identification and characterization of putative alternative promoters of human genes. *Genome Res.* 2006;16(1):55-65.
- [21] Kolandaivelu S, Huang J, Hurley JB, Ramamurthy V. AIPL1, a protein associated with childhood blindness, interacts with alpha-subunit of rod phosphodiesterase (PDE6) and is essential for its proper assembly. *J Biol Chem.* 2009;284(45):30853-30861.
- [22] Li J, Zoldak G, Kriehuber T, et al. Unique Proline-Rich Domain Regulates the Chaperone Function of AIPL1. *Biochemistry.* 2013;52 (12):2089–2096.
- [23] Linstedt AD, Hauri HP. Giantin, a novel conserved Golgi membrane protein containing a cytoplasmic domain of at least 350 kDa. *Mol Biol Cell.* 1993;4(7):679-693.
- [24] Mitton KP, Swain PK, Chen S, Xu S, Zack DJ, Swaroop A. The leucine zipper of NRL interacts with the CRX homeodomain. A possible mechanism of transcriptional synergy in rhodopsin regulation. *J Biol Chem.* 2000;275(38):29794-29799.
- [25] Miura F, Kawaguchi N, Sese J, et al. A large-scale full-length cDNA analysis to explore the budding yeast transcriptome. *Proc Natl Acad Sci U S A.* 2006;103(47):17846-17851.
- [26] Miyama, M., Shimizu H., Sugiyama, M. and Hanagata N. Sequencing and analysis of 14,842 expressed sequence tags of burma mangrove, *Bruguiera gymnorrhiza*. *Plant Science.* 2006;171(2):234-241.
- [27] Modrek B, Resch A, Grasso C, Lee C. Genome-wide detection of alternative splicing in expressed sequences of human genes. *Nucleic Acids Res.* 2001;29(13):2850-2859.
- [28] Nakamura F, Stossel TP, Hartwig JH. The filamins: organizers of cell structure and function. *Cell Adh Migr.* 2011;5(2):160-169.
- [29] Netchine I, Sobrier ML, Krude H, et al. Mutations in LHX3 result in a new syndrome revealed by combined pituitary hormone deficiency. *Nat Genet.* 2000;25(2):182-186.
- [30] Osada N, Hirata M, Tanuma R, et al. Collection of *Macaca fascicularis* cDNAs derived from bone marrow, kidney, liver, pancreas, spleen, and thymus. *BMC Res Notes.* 2009;2:199.
- [31] Oshikawa M, Sugai Y, Usami R, Ohtoko K, Toyama S, Kato S. Fine expression profiling of full-length transcripts using a size-unbiased cDNA library prepared with the vector-capping method. *DNA Res.* 2008;15(3):123-136.
- [32] Oshikawa M, Tsutsui C, Ikegami T, et al. Full-length transcriptome analysis of human retina-derived cell lines ARPE-19 and Y79 using the vector-capping method. *Invest Ophthalmol Vis Sci.* 2011;52(9):6662-6670.
- [33] Pan Q, Shai O, Lee LJ, Frey BJ, Blencowe BJ. Deep surveying of alternative splicing complexity in the human transcriptome by high-throughput sequencing. *Nat Genet.* 2008;40(12):1413-1415.
- [34] Pruitt KD, Tatusova T, Klimke W, Maglott DR. NCBI Reference Sequences: current status, policy and new initiatives. *Nucleic Acids Res.* 2009;37(Database issue):D32-36.
- [35] Rosing M, Ossendorf E, Rak A, Barnekow A. Giantin interacts with both the small GTPase Rab6 and Rab1. *Exp Cell Res.* 2007;313(11):2318-2325.
- [36] Seelig HP, Schranz P, Schröter H, Wiemann C, Renz M. Macroglolgin--a new 376 kD Golgi complex outer membrane protein as target of antibodies in patients with rheumatic diseases and HIV infections. *J Autoimmun.* 1994;7(1):67-91.
- [37] Sloop KW, Dwyer CJ, Rhodes SJ. An isoform-specific inhibitory domain regulates the LHX3 LIM homeodomain factor holoprotein and the production of a functional alternate translation form. *J Biol Chem.* 2001;276(39):36311-36319.

- [38] Sloop KW, Meier BC, Bridwell JL, Parker GE, Schiller AM, Rhodes SJ. Differential activation of pituitary hormone genes by human Lhx3 isoforms with distinct DNA binding properties. *Mol Endocrinol.* 1999;13(12):2212-2225.
- [39] Sohda M, Misumi Y, Fujiwara T, Nishioka M, Ikehara Y. Molecular cloning and sequence analysis of a human 372-kDA protein localized in the Golgi complex. *Biochem Biophys Res Commun.* 1994;205(2):1399-1408.
- [40] Sohocki MM, Bowne SJ, Sullivan LS, et al. Mutations in a new photoreceptor-pineal gene on 17p cause Leber congenital amaurosis. *Nat Genet.* 2000;24(1):79-83.
- [41] Sönnichsen B, Lowe M, Levine T, Jämsä E, Dirac-Svejstrup B, Warren G. A role for giantin in docking COPI vesicles to Golgi membranes. *J Cell Biol.* 1998;140(5):1013-1021.
- [42] Suzuki N, Uefuji H, Nishikawa T, et al. Construction and analysis of EST libraries of the trans-polyisoprene producing plant, *Eucommia ulmoides* Oliver. *Planta.* 2012;236(5):1405-1417.
- [43] Suzuki Y, Sugano S. Construction of full-length-enriched cDNA libraries. The oligo-capping method. *Methods Mol Biol.* 2001;175:143-153.
- [44] Swaroop A, Xu JZ, Pawar H, Jackson A, Skolnick C, Agarwal N. A conserved retina-specific gene encodes a basic motif/leucine zipper domain. *Proc Natl Acad Sci U S A.* 1992;89(1):266-270.
- [45] Takafuta T, Wu G, Murphy GF, Shapiro SS. Human beta-filamin is a new protein that interacts with the cytoplasmic tail of glycoprotein Ibalpha. *J Biol Chem.* 1998;273(28):17531-17538.
- [46] Takeda J, Suzuki Y, Nakao M, et al. Large-scale identification and characterization of alternative splicing variants of human gene transcripts using 56,419 completely sequenced and manually annotated full-length cDNAs. *Nucleic Acids Res.* 2006;34(14):3917-3928.
- [47] Tatsumoto S, Adati N, Tohtoki Y, et al. Development and Characterization of cDNA Resources for the Common Marmoset: One of the Experimental Primate Models. *DNA Res.* 2013;20(3):255-262.
- [48] Uchida K, Moriyama S, Chiba H, et al. Evolutionary origin of a functional gonadotropin in the pituitary of the most primitive vertebrate, hagfish. *Proc Natl Acad Sci U S A.* 2010;107(36):15832-15837.
- [49] Uenishi H, Morozumi T, Toki D, Eguchi-Ogawa T, Rund LA, Schook LB. Large-scale sequencing based on full-length-enriched cDNA libraries in pigs: contribution to annotation of the pig genome draft sequence. *BMC Genomics.* 2012;13:581.
- [50] van der Flier A, Kuikman I, Kramer D, et al. Different splice variants of filamin-B affect myogenesis, subcellular distribution, and determine binding to integrin [beta] subunits. *J Cell Biol.* 2002;156(2):361-376.
- [51] Wakamatsu A, Kimura K, Yamamoto J, et al. Identification and functional analyses of 11,769 full-length human cDNAs focused on alternative splicing. *DNA Res.* 2009;16(6):371-383.
- [52] Wang ET, Sandberg R, Luo S, et al. Alternative isoform regulation in human tissue transcriptomes. *Nature.* 2008;456(7221):470-476.
- [53] Watanabe J, Wakaguri H, Sasaki M, Suzuki Y, Sugano S. Comparasite: a database for comparative study of transcriptomes of parasites defined by full-length cDNAs. *Nucleic Acids Res.* 2007;35(Database issue):D431-438.

- [54] Wistow G, Bernstein SL, Wyatt MK, et al. Expressed sequence tag analysis of human RPE/choroid for the NEIBank Project: over 6000 non-redundant transcripts, novel genes and splice variants. *Mol Vis.* 2002;8:205-220.
- [55] Xu Wf, Xie Z, Chung DW, Davie EW. A novel human actin-binding protein homologue that binds to platelet glycoprotein Ibalph. *Blood.* 1998;92(4):1268-1276.
- [56] Xu XL, Fang Y, Lee TC, et al. Retinoblastoma has properties of a cone precursor tumor and depends upon cone-specific MDM2 signaling. *Cell.* 2009;137(6):1018-1031.
- [57] Zhao AC, Zhao TF, Nakagaki K, et al. Novel molecular and mechanical properties of egg case silk from wasp spider, *Argiope bruennichi*. *Biochemistry.* 2006;45(10):3348-3356.
- [58] Zhou J, Liao M, Hatta T, Tanaka M, Xuan X, Fujisaki K. Identification of a follistatin-related protein from the tick *Haemaphysalis longicornis* and its effect on tick oviposition. *Gene.* 2006;372:191-198.

Chapter 23

AN EPIGENETIC VIEW ON ALTERNATIVE SPLICING

***Gabriele A. Fontana, Aurora Rigamonti
and Silvia M. L. Barabino****

Department of Biotechnology and Biosciences,
University of Milano-Bicocca Piazza della Scienza, Milan, Italy

ABSTRACT

Genome-wide analysis indicate that alternative splicing of mRNA precursors (pre-mRNAs) affects the vast majority of human genes. Alternative splicing provides a fundamental mechanism to increase transcriptome complexity, allowing the production of two or more mRNA variants that often encode proteins with different, sometimes opposite functions. Its importance is underscored by the observation that misregulated alternative splicing can lead to human diseases. Pre-mRNA splicing has long been known to be regulated by *cis*-acting sequence elements and *trans*-acting protein factors. In higher eukaryotes, it mostly occurs co-transcriptionally so that it is not surprising that a role for chromatin and epigenetic factors in the regulation of exon inclusion is now emerging. In this review, we will discuss the most recent findings on the roles played by chromatin structure on the modulation of the cotranscriptional splicing reactions. In particular, we will focus our attention on how the modulation of the transcribing RNA polymerase II, the changes in nucleosome architecture and the presence of different histone modifications contribute to the regulation of the splicing process.

1. INTRODUCTION

Most eukaryotic mRNAs are generated from their primary transcripts (pre-mRNAs) through capping at the 5' end, removal of introns by splicing and 3' end cleavage and polyadenylation. These processes can lead to transcript diversification through the phenomenon of alternative splicing (AS). AS can lead to the specific inclusion or the skipping of exons (or

* Corresponding Author's Email: silvia.barabino@unimib.it (University of Milano-Bicocca, Piazza della Scienza, 2, I-20126 Milano, Italy; Tel: +39-02-6448 3352; Fax: +39-02-6448 3569).

even parts of an exon), to the selection of different 3' terminal exons, or to intron retention. It is likely that more than 95% of all human genes give rise to alternative mRNAs (Pan et al., 2008; Barash et al., 2010). The effect of AS in expanding the protein repertoire might partially underlie the apparent discrepancy between gene number and the complexity of higher eukaryotes. Given the crucial role of AS in the regulation of gene expression, it is not surprising that alterations in this process are associated with cancer and neurodegenerative pathologies, such as spinal muscular atrophy (SMA) and amyotrophic lateral sclerosis (ALS) (for review, David et al., 2010, Cooper et al., 2009).

RNA splicing occurs in the spliceosome, a large complex composed of five small ribonucleoprotein particles (U1, U2, U4/U6 and U5 snRNPs) and many non-snRNPs splicing factors. SR proteins and hnRNP proteins were the first non-snRNPs splicing factors identified. These proteins are components of the basal splicing machinery but, since their concentration can influence splice site selection, they contribute to AS. In addition, a growing list of tissue-specific AS regulators have been identified in recent years (for review see Wahl et al., 2009).

The mature 3' ends of mRNAs, with the exception of replication-dependent histones transcripts, are generated by endonucleolytic cleavage of the pre-mRNA followed by polyadenylation of the upstream cleavage product. Pre-mRNA 3'-end processing requires also several *trans*-acting protein factors (for review see Proudfoot, 2011). A large proportion of mammalian genes also undergoes alternative polyadenylation generating mRNA variants that differ in their coding sequence and/or in their 3' untranslated regions (UTRs), thereby potentially regulating the stability, localization and translation efficiency (for review see Tian et al., 2013)

Although the general mechanisms of pre-mRNA splicing and 3'-end processing have been well studied, how specific exons are chosen and are included in the mature transcript is still not completely clear. Both the splicing machinery and the 3' end processing complex assemble on conserved sequence elements that define the intron-exon junctions, the so-called splice sites (ss), the branch point sequence (BPS), a poorly conserved sequence located near the 3' end of the intron, and the polyadenylation site (PAs, Figure 1). In addition to these core signals, splicing is influenced by other regulatory elements (Wang et al., 2008). These elements are conventionally classified as exonic splicing enhancers (ESEs) or silencers (ESSs) depending whether they function to promote or inhibit inclusion of the exon they reside in, and as intronic splicing enhancers (ISEs) or silencers (ISSs) if they enhance or inhibit usage of adjacent splice sites or exons from an intronic location. These regulatory elements function by recruiting *trans*-acting factors that activate or inhibit splice site recognition and/or spliceosome assembly.

The early steps of spliceosome assembly, which provide the main targets for regulation, involve recognition of the consensus splice sites at both ends of the intron. Members of the SR family of splicing factors play essential roles in the early steps of splice-site recognition. These proteins contain one or two N-terminal RNA recognition motifs (RRMs) that function in sequence-specific RNA binding and a C-terminal domain rich in alternating arginine and serine residues, referred to as RS domain that is required for protein-protein interactions with other RS domains. SR proteins bound to specific RNA sequence elements are thought to recruit key splicing factors enhancing the recognition of splice sites and influencing splice site selection in a concentration-dependent manner (for review, see Zhou et al., 2013). This raises the possibility that tissue-specific expression of SR proteins may drive variation in splicing patterns. In addition, members of the family of hnRNP have also been shown to participate in the regulation of AS. Often these proteins have an antagonistic function to SR proteins. So far, only a few

systems of regulated splice site choice have been genetically or biochemically dissected and most regulatory proteins and sequence elements have not yet been identified.

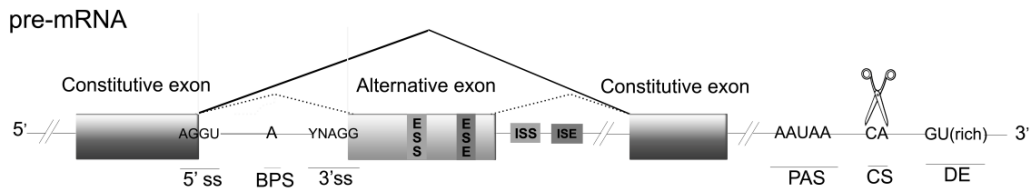


Figure 1. Schematic structure of an eukaryotic pre-mRNA. The pre-mRNA contains different *cis*-acting regulatory elements. The intron sequences are highly variable except for a conserved short region at 5' and 3' ends, called splice sites (ss), and the branch point sequence (BPS) within the intron. Exons and introns also contain splicing enhancers (ESEs and ISE) or silencers (ESSs and ISSs). At the 3' end of the pre-mRNA signals are found that direct cleavage and polyadenylation. The canonical polyadenylation signal consists of a conserved hexameric sequence, termed polyadenylation site (PAS) that precedes the CA dinucleotide where cleavage of the pre-mRNA occurs (cleavage site, CS). The CS is followed by a U- or UG- rich stretch of nucleotides (downstream element, DE).

This review concentrates on recent advances in the study of the cross-talk between chromatin structure and AS regulation. First, it focuses on several recent reports, that found large-scale evidence for a connection between nucleosome positioning and exon-intron architecture. An interesting emerging concept from these studies is that nucleosome positioning may reflect the exon-intron architecture. Then, we review histone modifications that may contribute to splicing regulation. Finally, intragenic DNA methylation and evidence for a role of methylated cytosine (5-mC) in exon definition are reviewed.

2. ALTERNATIVE SPLICING AND RNA POL II ELONGATION RATE

Although initially splicing and 3' end formation have been studied *in vitro* as independent processing events, biochemical, cytological and functional evidence suggest that all the events leading to the synthesis of the mature mRNA are coupled to transcription (for review see Kornblihtt et al., 2013). Several RNA processing factors are recruited on the C-terminal domain of RNA polymerase II (RNA Pol II) and are deposited on the nascent pre-mRNA molecule during transcription elongation. Transcription-coupled AS can be explained by the differential recruitment of AS factors on the transcribing polymerase. The carboxy-terminal domain (CTD) of the largest RNA Pol II subunit has a key role in the coupling of transcription with the different maturation steps that are required for the biogenesis of the mature mRNA molecule. The CTD undergoes extensive phosphorylation on Ser2 and Ser5 of the heptapeptides repeats (YSPTSPS). These phosphorylation events are associated to the transition from initiation to elongation. It has been proposed that RNA Pol II pausing at promoter proximal sites is phosphorylated on Ser5 but not on Ser2 (de la Mata et al., 2003; Morris et al., 2005). Several AS factors, including SR proteins bind to the phosphorylated CTD (Das et al., 2007).

Recently, a novel mode of splicing factor recruitment by Argonaute proteins has been uncovered (Ameyar-Zazoua et al., 2012). Argonaute proteins are the catalytic components of the cytoplasmic RNA-inducing silencing (RISC) complex responsible for RNAi silencing. In

the nucleus, they regulate transcription by inducing gene silencing. Interestingly, silencing of AGO1 and AGO2 was found to influence AS of the CD44 gene. Moreover, both AGO proteins physically interact with components of the splicing machinery, suggesting that they may participate in the recruitment of the splicing machinery to chromatin.

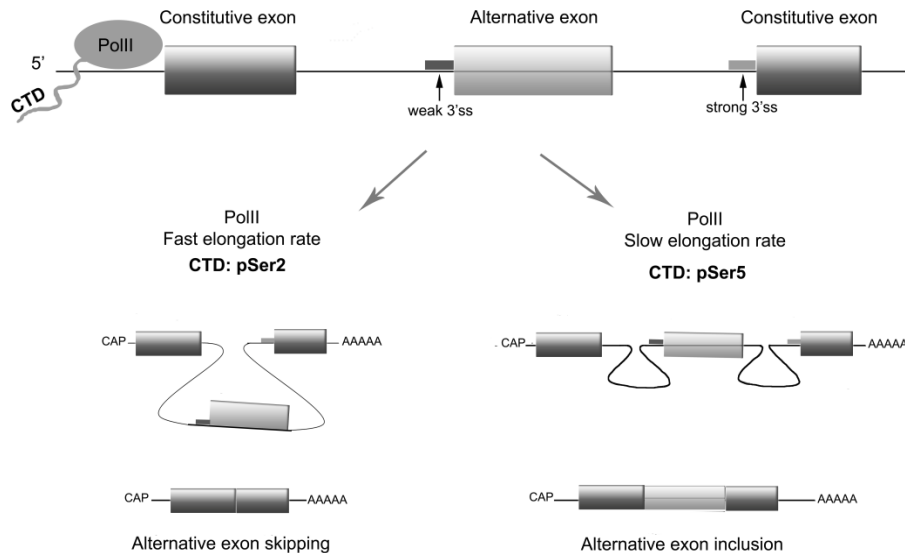


Figure 2. Kinetic coupling between transcription elongation and alternative splicing. Alternative exons are generally characterized by weak splice sites. The inclusion of an alternative exon with a weak 3'ss near a constitutive exon can be modulated by the elongation rate of RNA Pol II. A fast, processive polymerase, which is associated with a CTD phosphorylated in Serine 2 (pSer2), favors the skipping of the alternative exon. In contrast, phosphorylation of Serine 5 (pSer5) is associated with a slow processive Pol II that favor the inclusion of the alternative exon allowing more time for splice site recognition.

Another way to explain the coordination of splicing with transcription is a kinetic coupling between AS and transcription elongation (for review see Srebrow et al., 2006; Allemand et al., 2008, Figure 2). Slowing down the polymerase may favor the use of weak splice sites by delaying the synthesis of downstream splice sites, thus facilitating the recognition of suboptimal exons. The observation that inhibitors of histone deacetylation favors skipping of alternative exons (Nogués et al., 2002), possibly by promoting hyperacetylation of core histones thus facilitating transcription elongation, points towards an involvement of chromatin structure in the elongation-dependent regulation of AS. Further support to this idea has been recently provided by the report that the catalytic subunit Brahma (BRM, *SMARCA2*) of the chromatin remodelling complex SWI/SNF modulates AS by affecting RNA Pol II elongation rate (Batsché et al., 2006, see below). Recently, the kinetic coupling of alternative splicing and RNA Pol II elongation was shown to be influenced by DNA damage (Muñoz et al., 2009). These authors found that UV irradiation inhibits transcription elongation by inducing the hyperphosphorylation of RNA Pol II carboxy-terminal domain. This in turn can influence AS decisions, influencing the choice between cell survival and cell death.

3. NUCLEOSOMES AND ALTERNATIVE SPLICING

Nucleosomes are the basic building blocks of chromatin and consist of 147 base pairs of duplex DNA wrapped around a protein multi-subunit complex called “histone octamer”, which in turn is composed of two copies of each of the four canonical histones proteins H2A, H2B, H3 and H4. The positively charged residues of the histone proteins contact the phosphate backbone of the DNA every 10,4 base pairs, so that the 147 bases stretches of the DNA wrapped around the histone octamer make nearly 14 contacts (Khorasanizadeh, 2004).

During transcription, nucleosomes represent a “roadblock” to gene expression, as their presence may affect the accessibility to the DNA and prevent the polymerase from reading the template strand. Chromatin remodeling complexes are able to overcome these problems, as they displace the nucleosomes starting from the promoter region of a given gene, so that complete transcriptional activation can occur. During elongation, nucleosomes are evicted or remodeled in front of the transcribing polymerase, and are subsequently replaced in the region behind the polymerase (Workman, 2009). The replacement of nucleosomes after the passage of the RNA Pol II is an important step that allows nucleosomes exchange and/or recycling (Kulaeva et al., 2007).

Since 75-90% of genomic DNA is wrapped around nucleosomes (Wu et al., 2009), nucleosome occupancy certainly contributes to the regulation of gene expression. In fact, nucleosomes follow a non-random distribution in the genome, as they mainly cover the coding regions. In contrast, non-coding regions are characterized by a low nucleosome occupancy. These “open” regions are preferentially located at transcription start sites (TSS), which may be depleted of nucleosomes to retain their accessibility for transcription factors binding (Segal et al., 2006; Liu et al., 2011). The differential nucleosome occupancy observed at the level of coding versus non-coding regions possibly depends on specific genomic sequences. As a matter of facts, it has been proposed that coding regions are enriched in nucleosomes because of their relatively high GC content, while non-coding regions, that are characterized by a low GC content, are depleted of nucleosomes (Schwartz et al., 2009). These observations allow the computational prediction of nucleosome binding sites on the genomic sequence, that can be subsequently validated experimentally (Segal et al., 2006; Wu et al., 2009; Liu et al., 2011).

The first evidence of a connection between AS and nucleosome organization dates back to 1991, when Beckmann and colleagues observed that the distance between consecutive 5' and 3' splice sites in the pre-mRNA is very similar to the unit of DNA wrapped around a single nucleosome (Beckmann et al., 1991). Thanks to technological advancements in high-throughput screenings, this initial observation was supported by robust experimental evidence. Schwartz and colleagues found a consistently higher nucleosome occupancy on exons with respect to flanking introns, possibly due to the higher GC-content found in coding regions (Schwartz et al., 2009). Moreover, well-positioned nucleosomes were found in exons with weak splice signals and in isolated exons, suggesting that nucleosomes may facilitate the recognition of these exons. A difference in the levels of nucleosome occupancy was also found between constitutive and alternatively spliced exons. The same observations were independently published by Tilgner and colleagues (Tilgner et al., 2009), who found that a high nucleosome occupancy in genomic regions containing exons with weak splice sites. Moreover, they found a striking connection between nucleosome positioning and exon definition. By analyzing high-throughput data from human and from *C.elegans*, these authors found that high nucleosome

occupancy on specific exons is evolutionarily conserved and correlates with the exon inclusion rate. More specifically, exons which are constitutively included in the mature transcript have high nucleosome occupancy within and upstream of their sequence. Those nucleosomes, that are located within the exons, are preferentially localized at the centre of the coding sequence, and not at the level of the splice sites. Tilgner and colleagues finally propose that the nucleosome-dependent exon definition is independent from transcription, as non-expressed genes, as well as actively expressed genes, show a comparable nucleosome occupancy in their exons. The consideration that nucleosomes define the exons independently from transcription is sustained by the study of Andersson and coauthors (Andersson et al., 2009). In this report, the authors reasoned that the presence of nucleosomes in inactive genes suggests that they participate in the exon-intron architecture of the genome in completely opposite way compared to their role in the regulation of transcription. As a matter of fact, they found significant difference in nucleosome occupancy at the level of actively expressed genes versus silenced genes. In particular, they found that low-expressed genes display a relative depletion of nucleosomes around their TSS, while actively transcribed genes show a peculiar nucleosome pattern that defines the region around their TSS. Nucleosomes located around the TSS are highly mobile and prone to be displaced, in accordance with their role in transcription regulation. In contrast, nucleosomes located in the body of the gene are more resistant to displacement (Weiner et al., 2010).

For what the position occupied by nucleosomes respect to the 5' and 3' splice sites is concerned, the debate is still open, as observations differ. Some reports (see for example Kogan et al., 2005) indicate that nucleosome patterns are evolutionarily conserved and that they mirror the distribution of the splice sites. In this scenario, nucleosomes would participate to the splicing reaction by "covering" and protecting the splice sites from possible mutations. Instead, other reports (see for example Andersson et al., 2009; Tilgner et al., 2009) found a specific enrichment of nucleosomes within the exons, and not at the level of the 5' and 3' splice sites. Despite these differences, there is a general consensus that nucleosomes are recruited on the specific genomic regions based on their sequences.

The correlation between nucleosome positioning and AS has been very recently addressed by Keren-Shaul and colleagues (Keren-Shaul et al., 2013). Based on the evidence that loss of RNA Pol II causes a relaxation of chromatin (Weiner et al., 2010), and that transcription causes massive rearrangements in nucleosomes positioning (Kulaeva et al., 2007), the authors investigated whether the AS reaction could also participate in regulating chromatin structure. They found that overexpression of mutated forms of the U1 snRNA, either with low or strong efficiency in binding to the 5' splice sites present in the pre-mRNA, has an impact on chromatin structure. Specifically, a more efficient binding of U1 snRNA induces an increase in exon inclusion and a concomitant increase in nucleosome occupancy at the level of internal alternative exons. This is paralleled by a local increase in the amount of pausing RNA Pol II at the level of the alternative exon, a feature which is linked to exon inclusion (de la Mata et al., 2003). Even if the mechanism underlying the AS-dependent increase in nucleosome occupancy is still unclear, it appears to be independent from transcription, as both the low-efficient and the strong-efficient form of U1 have the same effects in terms of RNA Pol II recruitment.

But how nucleosomes positioning influences the cotranscriptional AS reaction? The mechanism remains elusive, but some possible explanations have been proposed. First, nucleosomes may affect the RNA Pol II processivity, which in turn may have an impact on exon inclusion (Keren-Shaul et al., 2013). This putative connection may be true, even if several

reports agree that nucleosome positioning on exons is independent from transcription (Tilgner et al., 2009; Keren-Shaul et al., 2013). Second, nucleosomes may influence AS thanks to post-translational modifications (PTMs) present in the tails of the histones that constitute them. The growing body of evidence connecting histone PTMs and AS suggests that this explanation is highly probable, and it is sustained by the observation that the nucleosomes that define the exons in the genome are not only positioned in a predictable way but they also carry specific histone PTMs (Andersson et al., 2009, see below). Third, nucleosomes, either directly or indirectly, may recruit the splicing factors on the splice sites of the nascent pre-mRNA. Even if the splicing reaction can occur *in vitro* without the presence of nucleosomes, the same reaction is more efficient *in vivo* when it is coupled to transcription (Das et al., 2006). This consideration, together with the robust evidence that nucleosomes are positioned more stably on alternative exons with weak splice sites (Tilgner et al., 2009), indicates that nucleosomes are a crucial player in defining the exon-intron architecture in the genome.

4. CHROMATIN REMODELING AND ALTERNATIVE SPLICING

Eukaryotic cells have evolved a number of enzymatic complexes that are able to change the chromatin architecture. These protein complexes, known as chromatin remodelers, are divided in four main families: SWI/SNF, ISWI, CHD and INO80. These families share common features, such as the ability to disrupt the contacts between nucleosomes and DNA, their multi-subunit composition, and the presence of a constitutive and evolutionarily conserved ATPase subunit (Clapier et al., 2009). The ATP-driven chromatin remodelling activity of these complexes regulates the accessibility of the different regions of the chromosome (Hargreaves et al., 2011). The “remodelers” play crucial roles in regulating DNA replication, repair, and recombination, as well as gene expression. In particular, chromatin remodelers are involved in the regulation of transcription, by controlling the accessibility of transcription factors (TFs) at the level of the promoters and by facilitating transcription elongation in the body of the gene. Chromatin remodelling complexes are recruited to specific sites of the chromatin thanks to presence, in their subunits, of protein domains (bromodomains and chromodomains) that recognize specific histone modifications (such as acetylation, methylation and phosphorylation) (Clapier et al., 2009).

As reported by a paper from Muchardt’s lab (Batsché et al., 2006), it has been proposed that chromatin remodelling complexes play also a role in the AS regulation. Specifically, the authors report that Brahma (BRM), one of the two mutually exclusive ATPase subunits of the human SWI/SNF chromatin remodelling complex, is able to modulate the AS reaction by enhancing the inclusion of alternative exons which reside in the body of the genes. In particular, when BRM is present inside the SWI/SNF complex, it interacts with components of the splicing machinery such as U1 and U5 snRNPs. BRM is also able to cooperate with Sam68, a well-known exon inclusion enhancer. This interaction, as well as the shared interaction between Brm, Sam68 and U5, increases the inclusion of alternative exons. Concomitant with the presence of this complex, accumulation of RNA Pol II-pSer5 was observed in the same genomic regions. The pSer5 modification of the CTD is linked to a slow processive form of the polymerase and to exon inclusion (de la Mata et al., 2003). Most importantly, the positive effect played by BRM on exon inclusion seems to be specifically connected to exons with weak splice

sites, which, without a molecular mechanism able to increase their inclusion, would be instead excluded from the mature transcript. In conclusion, this paper suggests a intriguing link between chromatin remodelling complexes and cotranscriptional AS. The observation that BRM plays a role in the exon inclusion process is substantially confirmed by an independent paper, in which Ito and colleagues (Ito et al., 2008) reported that Brm is involved in the inclusion of *TERT* exon 7, by an interaction with p54^{nrb}. Interestingly, modulation of exon 7 inclusion may have an impact on the activity of the telomerase, the protein which is encoded by the *TERT* gene,

Interestingly silencing of different subunits of the SWI/SNF complex in *Drosophila* cells indicate that this complex is not only involved in the AS of internal exons, but also in the choice of different polyadenylation sites. A first report provides evidence that the *Drosophila* SWI/SNF complex is associated with the nascent pre-mRNPs (Tyagi et al., 2009). A more recent paper from the same group then demonstrated that selective knock-out of the *Drosophila* core SWI/SNF subunits (Brm, Snr1 and Mor) affects the alternative processing of a subset of transcripts, changing the relative abundance of the isoforms produced (Waldholm et al., 2011).

5. HISTONE POST-TRANSLATIONAL MODIFICATIONS AND ALTERNATIVE SPLICING

All the four major histone types that are included in the nucleosome have an amino-terminal region that protrudes beyond the nucleosome surface. This “tail” is the target of a wide variety of post-translation modifications (PTMs). In contrast to the DNA methylation, the only chemical modification that occurs on DNA so far identified, histones have at least 100 different PTMs, which include methylation, acetylation, phosphorylation and ubiquitination (Bernstein et al., 2007). Table 1 summarizes the major PTMs affecting transcription and splicing.

From a mechanistic point of view, histones PTMs have an impact on the stability of the interaction between nucleosomes and DNA, modulate the protein-protein interactions involving histones, and/or generate the platform that triggers other subsequential histone PTMs. Acetylation of the histone tails, catalyzed by histone acetylases (HATs), removes the positive charges present on the lysine residues, thereby decreasing the interactions between the tails and the negatively charged phosphate groups present in the DNA. As a consequence of HATs activity, chromatin structure becomes more relaxed and transcription is generally facilitated. On the contrary, de-acetylation of the histone tails, catalyzed by the histone deacetylases (HDACs), results in a more closed and compacted chromatin structure associated with transcriptional silencing. Compared to acetylation, methylation, the second major histone PTM, has a different effect as the addition of a methyl group does not alter the relative charge of lysine and/or arginine residues. For this reason, methylation (catalyzed by histone methylases, HMTs) and demethylation (catalyzed by histone demethylases, HDMs) of histone tails have different outcomes on chromatin compaction depending on their target residue and on the presence of other methyl or acetyl groups in close proximity (Zentner et al., 2013).

The advent of epigenetic studies focused on genome-wide mapping of histone PTMs, has allowed to correlate the presence of these modifications to different gene expression outcomes. It has now become clear that the combination of the modifications targeting the histone tails residing in a specific genomic region creates an “histone code” that can be read by

evolutionarily conserved chromatin-interacting proteins, that contain specific interaction domains. For example, the methylation mark is known to recruit proteins containing the chromodomain module (Eissenberg, 2012) while histone acetylation can be read by proteins containing bromodomains (Filippakopoulos et al., 2012). These protein-protein interactions are strictly regulated by “histone writers” (such as the already mentioned HATs/HDACs and HMTs/HDMs) that modulate both the relative abundance of a specific modification in a given genomic region and the amount of PTMs targeting a single residue.

Compared to acetylation, the patterns of methylation are even more complex, because a single lysine can harbor one (me1), two (me2) or three (me3) methyl groups, and the relative level of methylation of a single lysine may also lead to completely opposite effects. For example, it has been proposed that mono-methylation of lysine 9 of histone 3 (H3K9me1) recruits proteins that facilitate transcription (Barski et al., 2007), while di- (H3K9me2) and tri-methylation (H3K9me3) of the same residue are associated to transcriptional silencing (Barski et al., 2007, Rosenfeld et al., 2009). Specific methylases are dedicated to the task of adding the methyl group to a lysine which already harbors one modifications, while other enzymes are able to add this PTM if two methyl groups are already present, thus enhancing the complexity of the regulation of this process. Moreover, whole-genome mapping of histones modifications has revealed that different histone PTMs map in different regions of the body of a gene and are linked to specific outcomes. Some modifications, like H3K4me3 are enriched at the level of transcription start sites (Kolasinska-zwierz et al., 2009), whereas others, such as H3K36me3 and H3K79me3, are instead present in the body of the gene (Spies et al., 2009); moreover, marks such as H3K27me2, have been mapped at the level of intergenic heterochromatin regions (Kolasinska-zwierz et al., 2009). Table 1 is an attempt to simplify the great complexity of the histone codes that has so far been identified, highlighting their proposed functions.

Regarding the connections between histone’s PTMs and co-transcriptional AS, it has been observed that not only specific histone marks, such as H2BK5me1, H3K27me3 (Andersson et al., 2009) and H3K36me3 (Spies et al., 2009), are enriched in the gene body and at the level of the exons, but that different modifications also lead to diverse splicing outcomes. For example, it has been proposed that H3K36me3, one of the most studied and debated histone modification, is enriched at the level of the genomic regions containing exons, which are constitutively included in the mature transcript, while it is generally less present at the level of alternative exons (Kolasinska-zwierz et al., 2009).

How can the presence of specific histone marks influence the pattern of AS? Two mechanisms have been proposed so far. First, similarly to transcription regulation, histone modifications can provide a platform for the recruitment of specific splicing regulators. For example, it has been shown that H3K36me3 recruits proteins which are able to modulate the splicing outcomes, such as MRG15 (and adaptor protein), PTB (a splicing factor which interacts with MRG15) and SRSF1 (a SR protein which enhances exon inclusion) (Pradeepa et al., 2003). Second, histone modifications can influence RNA Pol II processivity. For example, Schor and colleagues demonstrated that depolarization of neurons decreases the inclusion of the murine *Ncam* exon 18 via local changes in specific histone PTMs which in turn influence the Pol II elongation rate (Schor et al., 2009). Specifically, an increase in the acetylation of H3K9 (either triggered by depolarization or by Trichostatin A, TSA) in the genomic region surrounding the variable *Ncam* exon 18 causes an increase in the chromatin accessibility, which locally allows to the polymerase to proceed faster, resulting in exon skipping.

Table 1. Histone codes and their link with genes expression and alternative splicing

Histone	Site	Modification	Proposed functions	References
H1	Lys26	Me1	Transcriptional silencing	Daujat et al., 2005; Garcia et al., 2004
	Ser27	Pho	Transcriptional activation	Daujat et al., 2005
H2A	Lys5	Ac	Transcriptional activation	Cuddapah et al., 2008
	Ser1	Pho	Transcriptional repression	Zhang et al., 2004
H2B	Lys5	Ac	Transcriptional activation	Karlič et al., 2010
		Me1	Transcriptional activation Enriched in exons	Barski et al., 2007 Schwartz et al., 2009
		Me3	Transcriptional silencing	Rosenfeld et al., 2009
H3	Lys4	Me1	Transcriptional activation	Benevolenskaya et al., 2007
		Me2	Transcriptional activation	Kornblihtt et al., 2013
		Me3	Transcriptional activation Alternative splicing	Spies et al., 2009 Koch et al., 2007; Kornblihtt et al., 2013
	Lys9	Me1	Transcriptional activation	Barski et al., 2007
		Me2	Transcriptional silencing	Rosenfeld et al., 2009; Kornblihtt et al., 2013
		Me3	Transcriptional silencing	Barski et al., 2007
		Ac	Transcriptional activation	Koch et al., 2007; Kornblihtt et al., 2013
	Lys14	Ac	Transcriptional activation	Koch et al., 2007
	Lys27	Me1	Transcriptional activation	Barski et al., 2007
		Me2	Transcriptional silencing Heterochromatin marker	Rosenfeld et al., 2009; Kolasinska-zwierz et al., 2009
		Me3	Transcriptional silencing Enriched in internal exons	Barski et al., 2007; Kornblihtt et al., 2013 Andersson et al., 2009
	Lys 36	Me3	Transcriptional activation Marks constitutively included exons	Schwartz et al., 2009 Kolasinska-zwierz et al., 2009
	Lys 79	Me1	Transcriptional activation	Barski et al., 2007
		Me2	Transcriptional activation	Steger et al., 2008
		Me3	Transcriptional activation Enriched in internal exons	Spies et al., 2009 Andersson et al., 2009
	Ser10	Pho	Transcriptional activation	Lo et al., 2000
H4	Lys20	Me1	Transcriptional activation	Barski et al., 2007
		Me3	Transcriptional activation	Schwartz et al., 2009

Interestingly, this hyperacetylation is paralleled by the increase in the H3K36me3 mark in the all the actively transcribed *Ncam* gene region, indicating a more general change in the chromatin landscape and a possible crosstalk between the histone PTMs. A more recent report by Chen and colleagues demonstrated that the H3K27me2 mark induces an increase in the

elongation rate of the polymerase (Chen et al., 2012). This is possible because H3K27me3 recruits the JMJD3 and KIAA1718 demethylases on a subset of genes. These genes are marked at their promoter-proximal regions with the H3K27me3 and H3K4me3 marks, and are also associated with a promoter-proximal, paused RNA Pol II. The JMJD3/ KIAA1718 complex demethylates the H3K27me3 mark, and this event releases the Pol II, triggering the productive elongation phase. This is possible because the JMJD3 complex recruits elongation factors, such as SPT6, SPT16, CDC73, and SETD2. The silencing of JMJD3 and of KIAA1718 reduces the Pol II elongation rate in the bodies of the monitored genes, a feature previously linked to exon inclusion (de la Mata et al., 2003; Batsché et al., 2006). Interestingly, the H3K27me3 mark has been mapped in the bodies of transcribed genes (Schwartz et al., 2006), suggesting that it may play a role in regulating the intragenic RNA Pol II processivity.

6. INTRAGENIC DNA METHYLATION AND ALTERNATIVE SPLICING

DNA methylation consists in the addition of a methyl group to the C5 position of the cytosine in a cytosine-phosphate-guanine (CpG) dinucleotide by a DNA-methyltransferase enzyme of the *Dnmt* family (Hattori et al., 2004). In vertebrate genomes the frequency of CpG dinucleotides is lower than expected based on random chance. This is due to the intrinsic instability of the methyl-cytosine that can spontaneously deaminate to thymine. For this reason, CpGs are evolutionarily lost over time, resulting in a progressive general depletion of this dinucleotide (Bird et al., 1980). The CpGs display a completely non-uniform distribution along several genomes, and this dinucleotide appears restricted to clusters termed “CpG islands” (CGIs). CGIs are defined as stretches of interspersed DNA sequences, both enriched in cytosine/guanine content and in the presence of several CpG dinucleotides. CGIs are usually 200 base pairs long, display a lower CpG depletion compared to other genome regions, and harbor unmethylated cytosines (Ponger et al., 2002). These clusters of CpG dinucleotides tend to localize in the genomic regions near to the transcription start sites (TSS) of the majority of housekeeping and/or ubiquitously expressed genes, where they constitute a key element defining eukaryotic promoter regions (Gardiner-Garden et al., 1987). A recent report (Vavouri et al., 2010) indicates that CpG-containing promoters have a peculiar transcription-associated chromatin organization, which can be depicted as an ordered and conserved distribution of nucleosomes containing specific histone marks.

CGIs are important elements that epigenetically regulate the expression of eukaryotic genes during differentiation and development. Genome-wide studies revealed that unmethylated CGIs are prominent in undifferentiated cells and in the embryo, and that this methylation-free state is associated with active transcription. During differentiation, some CpG dinucleotides contained in the CGIs acquire the methylation mark in a tissue and cell-specific fashion. This methylated state is associated with silencing of the downstream genes, as the methyl-cytosine directly inhibits the binding of transcription factors and recruits epigenetic modifiers, such as HDACs and Polycomb (PcG) proteins, which are associated with a long-term gene silencing (Deaton et al., 2011).

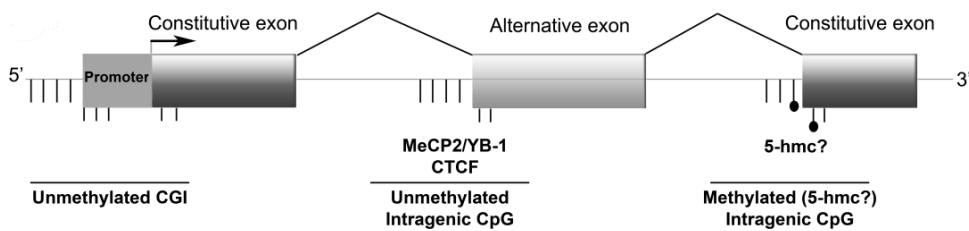
An enrichment of the CpG dinucleotide has also been detected in intragenic regions and their evolutionary conservation suggests that these sequences hold the potential to epigenetically regulate other “layers” of gene expression, such as AS. It has been observed that the content of

the CpG dinucleotide is significantly higher in the introns localized upstream to alternative cassette exons compared to introns preceding constitutive exons or localized downstream to alternative cassette exons. In particular, the CpG frequency is higher in the region flanking the acceptor site of the alternative cassette exons, and is not linked to the relative intron length and/or to the presence of *Alus* sequences (Malousi et al., 2008). An intriguing correlation between the peculiar localization of the CpG dinucleotide at the level of internal cassette exons and the regulation of AS is provided by data published in 2005 by Zoghbi's group (Young et al., 2005). The paper provides evidences that methyl-CpG-binding protein 2 (MeCP2), the protein containing mutations linked to the emergence of the Rett syndrome, is involved in the regulation AS. MeCP2 was previously described to specifically bind the DNA at the level of methylated CpGs (Nan et al., 2001). By protein co-immunoprecipitation assays, Young and collaborators discovered that MeCP2 directly interacts with the Y box-binding protein 1 (YB-1), a protein which mediates the DNA-RNA interactions involved in the regulation of transcription, translation (Khono et al., 2003) and AS (Stickeler et al., 2001). The MeCP2/YB-1 interaction is RNA-dependent, but independent from methylated DNA. This observation suggests that YB-1 is a RNA-dependent MeCP2 binding protein and that the interaction specifically occurs during transcription and it is separated from the previously reported roles of YB-1 in DNA repair and replication. By using splicing reporter minigenes, the authors demonstrated that the MeCP2/YB-1 complex directly modulates the inclusion of internal alternative exons. This observation was then confirmed by a splicing-sensitive, genome-wide survey of the AS events of endogenous genes expressed in the cerebral cortex of the MeCP2 model mice. Interestingly, this mouse model of Rett syndrome exhibits AS alterations relative to the inclusion of cassette exons containing the YB-1-binding ACE domain, such as the genes encoding the NR1 subunit of the NMDA receptor and the *Dlx5* gene. While this paper demonstrated that MeCP2 can contribute to AS regulation, it did not explore the correlation between cytosines methylation and MeCP2-engagement at the level of intragenic CpGs. However, a very recent report established a link between intragenic CpG methylation and regulation of AS. Khare and collaborators (Khare et al., 2013) focused their attention on 5-hydroxymethylcytosine (5-hmC), a derivative of methylated cytosine (5-mC), which is highly abundant in human and mouse brain tissues. Using assays able to discriminate between 5-hmC and 5-mC, the authors detected a striking brain-specific enrichment of 5-hmC in genomic regions containing genes involved in synaptic functions. In this specific class of genes, 5-hmC marks the exonic side of the exon-intron boundaries in a brain-specific fashion and has a direct effect on the splicing outcomes. In facts, the authors found that an increase in the 5-hmC content in the region proximal to the exon and/or within the exon is prominent in constitutively included exons, while the 5-hmC modification is less abundant in exons which are subjected to AS.

From the mechanistic point of view, it has been proposed that the presence of intragenic CpG methylation inhibits the binding of proteins involved in AS regulation and/or induces the formation of peculiar chromatin structures. Shukla and collaborators (Shukla et al., 2011) focused their attention on the *CD45* gene, which is a widely used model gene to study the regulation of AS. During lymphocyte differentiation the regulated inclusion/skipping of *CD45* variable exon V5 gives rise to protein isoforms that can be easily monitored. As most alternative exons, exon V5 has weak splice sites, so that it is usually skipped. The authors reported that the genomic region surrounding exon V5 displays an accumulation of the CCCTC-binding factor (CTCF), a protein that has been described for its roles in insulating inactive genomic regions and promoting long-range interactions between distant genomic regions. The accumulation of

CTCF on the variant V5 exon promotes its inclusion in the nascent transcript by inducing a localized pausing in the RNA Pol II. An inverse correlation between CTCF accumulation and methylation of the cytosines present in exon V5 was instead observed, indicating that methylcytosine inhibit CTCF binding. The DNA methylation patterns change during lymphocyte differentiation, and this dynamic methylation at the level of intragenic sites provide an astonishing mechanistic correlation between exon V5 inclusion and the differentiation process.

Undifferentiated cells



Differentiated cells

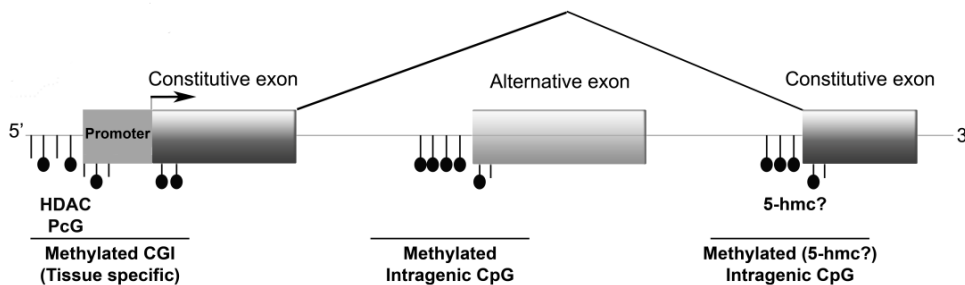


Figure 3. Epigenetic regulation of CpG islands (CGIs) during eukaryotic differentiation and development. Alternative splicing can be modulated during differentiation in a methylation-dependent manner. In undifferentiated cells when intragenic CpG are unmethylated, a MeCP2/YB-1 complex is able to directly modulate inclusion of the alternative exons of CD44, and of CT/CGRP (Young et al., 2005). In differentiated cells, instead a preferential skipping of the alternative exon is observed due to an increase of 5-hydroxymethylcytosine (5-hmC) in the proximal region of constitutive exons, and of methylated intragenic CpGs.

To extend this observation, the authors took advantage of CTCF ChIP-seq data, that showed that the great majority of the binding sites for this protein reside in intragenic regions. Other reports, basing on sequencing data, focused their attention on the correlation between intragenic cytosine methylation and the formation of peculiar chromatin structures (see for example Chodavarapu et al. 2011). Figure 3 is a schematic representation of the current knowledge regarding the intragenic methylation-dependent regulation of AS.

CONCLUSION

It is well-established that chromatin structure plays an important role in the transcriptional control of eukaryotic gene expression, but only recently it has become clear that chromatin organization can also contribute to AS regulation. AS participate in critical biological processes including cell growth and differentiation, cell death, pluripotency, development (Cooper et al., 2009; Kalsotra et al., 2011). The importance of AS is underscored by the fact that 95% of human genes produce alternatively spliced transcripts (Pan et al., 2008; Wang et al., 2008). However, the functional impact of the vast majority of AS events has not been characterized in any way, nor the molecular mechanisms underlying the selection of specific alternative exons are fully understood. These remain major challenges for future research. Moreover, a greater understanding of the structure-to-function relationship – that is, how chromatin organization can influence specific splicing events and what is the biological function of the different mRNA variants – will be required. In this respect, recent studies have revealed a novel layer of complexity. Although this review focuses on the cross-talk between chromatin proteins and splicing factors, it should be mentioned that the emerging class of long non-coding RNAs (lncRNAs) was shown to contribute to epigenetic regulation of gene expression. lncRNA can guide the organization of higher-order ribonucleoprotein complexes thereby modulating the activity of chromatin-modifying complexes (for review Mercer et al., 2013). No doubt that future work will likely uncover new, unexpected mechanisms that connect lncRNAs and chromatin states to regulate AS.

REFERENCES

- Allemand E., Batsché E., Muchardt C. Splicing, transcription, and chromatin: a ménage à trois. *Current Opinion in Genetics & Development*. 2008;18:145-151.
- Ameyar-zazoua M., Rachez C., Souidi M., Robin P., Fritsch L., Young R., Morozova N., Fenouil R., Descostes N., Andrau J., Mathieu J., Hamiche A., Ait-Si-Ali S., Muchardt C., Batsché E., Harel-Bellan A. Argonaute proteins couple chromatin silencing to alternative splicing. *Nature Structural & Molecular Biology*. 2012;19(10):998-1004.
- Andersson R., Enroth S, Rada-iglesias A., Wadelius C., Komorowski J. Nucleosomes are well positioned in exons and carry characteristic histone modifications. *Genome Research*. 2009;19(10):1723-1741.
- Barash Y., Calarco J.A., Gao W., Pan Q., Wang X., Shai O., Blencowe B.J., Frey B.J. Deciphering the splicing code. *Nature*. 2010;465(7294):53-59.
- Barski A., Cuddapah S., Cui K., Roh T.Y., Schones D.E., Wang Z., Wei G., Chepelev I., Zhao K. High-resolution profiling of histone methylations in the human genome. *Cell*. 2007;129(4):823-37.
- Batsché E., Yaniv M., Muchardt C. The human SWI / SNF subunit Brm is a regulator of alternative splicing. *Nature Structural & Molecular Biology*. 2006;13(1):22-29.
- Bauer A.P., Leikam D., Krinner S., Notka F., Ludwig C., La G. The impact of intragenic CpG content on gene expression. *Nucleic Acids Research*. 2010;38(12):3891-3908.
- Beckmann J. S., Trifonov E.N. Splice junctions follow a 205-base ladder. *PNAS*. 1991;88(March):2380-2383.

- Benevolenskaya EV. Histone H3K4 demethylases are essential in development and differentiation. *Biochem Cell Biol.* 2007;85(4):435-43.
- Bird A.P. DNA methylation and the frequency of CpG in animal DNA. *Nucleic Acids Research.* 1980;8(7):1499-1504.
- Bird A. DNA methylation patterns and epigenetic memory. *Genes & Development.* 2002;16:6-21.
- Chen S., Ma J., Wu F., Xiong L., Ma H., Xu W., Ly R., Ly X., Villen J., Gygi S.P., Liu S., Shi Y. The histone H3 Lys 27 demethylase JMJD3 regulates gene expression by impacting transcriptional elongation. *Genes & Development.* 2012;26:1364-1375.
- Chodavarapu R.K., Feng S., Bernatavichute Y.V., Stroud H. Yu Y., Hetzel J., Kuo F., Kim F., Cokus S.J., Casero D., Huijser P., Clark A.T., Krämer U., Merchant S.S., Zhang X., Jacobsen S.E. Relationship between nucleosome positioning and DNA methylation. *Nature.* 2011;466(7304):388-392.
- Clapier C.R., Cairns B.R.. The biology of chromatin remodeling complexes. *Annual Review Biochemistry.* 2009;78:273-304.
- Cooper T.A., Wan L., Dreyfuss G. RNA and disease. *Cell.* 2009;136(4):777-793.
- Cuddapah S., Jothi R., Schones D.E., Roh T.Y., Cui K., Zhao K. Global analysis of the insulator binding protein CTCF in chromatin barrier regions reveals demarcation of active and repressive domains. *Genome Res.* 2009;19(1):24-32.
- Das R., Dufu K., Romney B., Feldt M., Elenko M., Reed R. Functional coupling of RNAP II transcription to spliceosome assembly. *Genes & Development.* 2006;20(9):1100-1109.
- Das R., Yu J., Zhang Z., Gygi M.P., Krainer A.R., Gygi S.P., Reed R. SR proteins function in coupling RNAP II transcription to pre-mRNA splicing. *Molecular Cell.* 2007;867-881.
- Daujat S., Zeissler U., Waldmann T., Happel N., Schneider R. HP1 binds specifically to Lys26-methylated histone H1.4, whereas simultaneous Ser27 phosphorylation blocks HP1 binding. *Journal of Biological Chemistry.* 2005;280:38090-38095.
- David C.J., Manley J.L. Alternative pre-mRNA splicing regulation in cancer : pathways and programs unhinged. *Genes & Development.* 2010;24:2343-2364.
- de La Mata M., Alonso C.R., Fededa J.P., Pelisch F., Cramer P., Bentley D., Kornblihtt A.R. A slow RNA polymerase II affects alternative splicing in vivo. *Molecular Cell.* 2003;12(2):525-532.
- Deaton A.M., Bird A. CpG islands and the regulation of transcription. *Genes & Development.* 2011;25:1010-1022.
- Eissenberg J.C. Structural biology of the chromodomain: form and function. *Gene.* 2012;496(2):69-78.
- Filippakopoulos P., Knapp S. The bromodomain interaction module. *FEBS Letters.* 2012;586(17):2692-2704.
- Gardiner-Garden M., Frommer M. CpG Islands in vertebrate genomes. *Journal of Molecular Biology.* 1987;196(2):261-282.
- Garcia B.A., Busby S.A., Barber C.M., Busby S.A., Barber C.M., Shabanowitz J., Allis C.D., Hunt D.F. Characterization of phosphorylation sites on histone H1 isoforms by tandem mass spectrometry. *Journal of Proteome Research.* 2004;3:1219 - 1227.
- Haines T.R., Rodenhiser D.I., Ainsworth P.J. Allele-specific non-CpG methylation of the Nf1 gene during early mouse development. *Developmental Biology.* 2001;240:585-598.
- Hargreaves D.C., Crabtree G.R. ATP-dependent chromatin remodeling: genetics, genomics and mechanisms. *Cell.* 2011;21(3):396-420.

- Hattori N., Abe T., Hattori N., Suzuki M., Matsuyama T., Yoshida S., Li E., Shiota K. Preference of DNA methyltransferases for CpG islands in mouse embryonic stem cells. *Genome Research*. 2004;14:1733-1740.
- Ito T., Watanabe H., Yamamichi N., Kondo S., Tando T., Haraguchi T., Mizutani T., Sakurai K., Fujita S., Izumi T., Isobe T., Iba H. Brm transactivates the telomerase reverse transcriptase (TERT) gene and. *Biochemistry Journal*. 2008;411(1):201-209.
- Kalsotra A., Cooper T.A. Functional consequences of developmentally regulated alternative splicing. *Nature Review Genetics*. 2012;12(10):715-729.
- Karlič R., Chung H.R., Lasserre J., Vlahovick K., Vingron M. Histone modification levels are predictive for gene expression. *Proc Natl Acad Sci U S A*. 2010 Feb 16;107(7):2926-31.
- Keren-shaul H., Lev-maor G., Ast G. Pre-mRNA splicing is a determinant of nucleosome organization. *Plos One*. 2013;8(1):e53506.
- Khare T., Pai S., Koncencius K., Pal M., Kriukiene E., Liutkeviciute Z., Irimia M., Jia P., Ptak C., Xia M., Tice R., Tochigi M., Morera S., Nazarians A., Belsham D., Wong A.H., Blencowe B.J., Wang S.C., Kapranov P., Kustra R., Labrie V., Klimasauskas S., Petronis A. 5-hmC in the brain is abundant in synaptic genes and shows differences at the exon-intron boundary. *Nature Structural & Molecular Biology*. 2013;19(10):1037-1043.
- Khorasanizadeh S. The nucleosome: from genomic organization to genomic regulation. *Cell*. 2004;116:259-272.
- Kogan S., Trifonov E.N. Gene splice sites correlate with nucleosome positions. *Gene*. 2005;352:57 - 62.
- Koch C.M., Andrews R.M., Flicek P., Dillon S.C., Karaöz U., Clelland G.K., Wilcox S., Beare D.M., Fowler J.C., Couttet P., James K.D., Lefebvre G.C., Bruce A.W., Dovey OM, Ellis PD, Dhami P, Langford CF, Weng Z, Birney E, Carter NP, Vetrie D, Dunham I. The landscape of histone modifications across 1% of the human genome in five human cell lines. *Genome Research*. 2007;17(6):691-707.
- Kohn K., Izumi H., Uchiumi T., Ashizuka M., Kuwano M. The pleiotropic functions of the Y-box-binding protein, YB-1. *BioEssays*. 2003;25:691-698.
- Kolasinska-Zwier P., Down T., Latorre I., Liu T., Liu X.S., Ahringer J. Differential chromatin marking of introns and expressed exons by H3K36me3. *Nature Genetics*. 2009;41(3):376-81.
- Kornbliht A.R., Schor I.E., Alló M., Dujardin G., Petrillo E., Muñoz M.J. Alternative splicing: a pivotal step between eukaryotic transcription and translation. *Nature*. 2013;14:153-166.
- Kulaeva O.I., Gaykalova D., Studitsky V.M. Transcription through chromatin by RNA polymerase II: histone displacement and exchange. *Mutation Research*. 2008;618:116-129.
- Liu H., Duan X., Yu S., Sun X. Analysis of nucleosome positioning determined by DNA helix curvature in the human genome. *BMC Genomics*. 2011;12(1):72.
- Lo W.S., Trievel R.C., Rojas J.R., Duggan L., Hsu J.Y., Allis C.D., Marmorstein R., Berger S.L. Phosphorylation of serine 10 in histone H3 is functionally linked in vitro and in vivo to Gcn5-mediated acetylation at lysine 14. *Molecular Cell*. 2000;5(6):917-26.
- Lorincz M.C., Dickerson D.R., Schmitt M., Groudine M. Intragenic DNA methylation alters chromatin structure and elongation efficiency in mammalian cells. *Nature Structural & Molecular Biology*. 2004;11(11):1068-1075.
- Malousi A., Maglaveras N., Kouidou S. Intronic CpG content and alternative splicing in human genes containing a single cassette exon. *Epigenetics*. 2008;3:2(April):69-73.

- Mercer T.R., Mattick J.S. Structure and function of long noncoding RNAs in epigenetic regulation. *Nature Structural & Molecular Biology*. 2013;20(3):300-307.
- Morris D.P., Michelotti G.A., Schwinn D.A. Evidence that phosphorylation of the RNA Polymerase II carboxyl-terminal repeats is similar in yeast and humans. *Journal of Biological Chemistry*. 2005;280(36):31368-31377.
- Muñoz M.J., Santangelo P.M., Paronetto M.P., de la Mata M., Pelisch F., Boireau S., Glover-Cutter K., Ben-Dov C., Blaustein M., Lozano J.J., Bird G., Bentley D., Bertrand E., Kornblihtt A.R. . DNA damage regulates alternative splicing through inhibition of RNA Polymerase II elongation. *Cell*. 2009;137:708-720.
- Nan X., Bird A. The biological functions of the methyl-CpG-binding protein MeCP2 and its implication in Rett syndrome. *Brain & Development*. 2001;23:32-37.
- Nogués G., Kadener S., Cramer P., Bentley D., Kornblihtt A.R. Transcriptional activators differ in their abilities to control alternative splicing. *Journal of Biological Chemistry*. 2002;277:43110-43114.
- Pan Q., Shai O., Lee L.J., Frey B.J., Blencowe B.J. Deep surveying of alternative splicing complexity in the human transcriptome by high-throughput sequencing. *Nature Genetics*. 2009;40(12):1413-1416.
- Ponger L., Mouchiroud D. CpGProD: identifying CpG islands associated with transcription start sites in large genomic. *Bioinformatics*. 2002;18(4):631-633.
- Pradeepa M.M., Sutherland H.G., Ule J., Grimes G.R., Bickmore W.A. Psp1 / Ldgf p52 binds methylated histone H3K36 and splicing factors and contributes to the regulation of alternative splicing. *PLoS Genetics*. 2012;8(5).
- Proudfoot N.J. Ending the message: poly (A) signals then and now. *Genes & Development*. 2011;25:1770-1782.
- Rosenfeld J.A., Wang Z., Schones D.E., Zhao K., DeSalle R., Zhang M.Q. Determination of enriched histone modifications in non-genic portions of the human genome. *BMC Genomics*. 2009;10(143): doi:10.1186/1471-2164-10-143.
- Saxonov S., Berg P., Brutlag D.L. A genome-wide analysis of CpG dinucleotides in the human genome distinguishes two distinct classes of promoters. *PNAS*. 2006;103(5):1412-1417.
- Schor I.E., Rascovan N., Pelisch F., Allo M., Kornblihtt A.R. Neuronal cell depolarization induces intragenic chromatin modifications affecting NCAM alternative splicing. *PNAS*. 2009;106(11):4325-4330.
- Schor I.E., Kornblihtt A.R. Intragenic chromatin modifications: a new layer in alternative splicing regulation. *Epigenetics*. 2010;5(3):174-179.
- Schwartz Y.B., Kahn T.G., Nix D.A., Li X., Bourgon R., Biggin M., Pirrotta V. Genome-wide analysis of Polycomb targets in *Drosophila melanogaster*. *Nature Genetics*. 2006;38(6):700-705.
- Schwartz S., Meshorer E., Ast G. Chromatin organization marks exon-intron structure. *Nature Structural & Molecular Biology*. 2009;16(9):990-995.
- Segal E., Fondufe-Mittendorf Y., Chen L., Tharstrom A., Field Y., Moore I.K., Wang J.Z., Widom J. A genomic code for nucleosome positioning. *Nature*. 2009;442(7104):772-778.
- Spies N., Nielsen C.B., Padgett R.A., Burge C.B. Biased chromatin signatures around polyadenylation sites and exons. *Molecular Cell*. 2009;36(2):245-254.
- Srebrow A., Kornblihtt A.R. The connection between splicing and cancer. *Journal of Cell Science*. 2006;119:2635-2641.

- Steger D.J., Lefterova M.I., Ying L., Stonestrom A.J., Schupp M., Vakoc A.L., Chen J., Lazar M.A., Blobel G.A., Vakoc C.R. DOT1L/KMT4 recruitment and H3K79 methylation are ubiquitously coupled with gene transcription in mammalian cells. *Molecular and Cellular Biology*. 2008;28(8):2825-2839.
- Tian B., Manley J.L. Alternative cleavage and polyadenylation: the long and short of it. *Trends in Biochemical Sciences*. 2013;38(6):312-320.
- Tilgner H., Nikolaou C., Althammer S., Sammeth M., Beato M., Valrcacel J., Guigo R. Nucleosome positioning as a determinant of exon recognition. *Nature Structural & Molecular Biology*. 2009;16(9):996-1001.
- Tyagi A., Ryme J., Brodin D., Ostlund Farrants A.K., Visa N. SWI / SNF associates with nascent pre-mRNPs and regulates alternative pre-mRNA processing. *Plos Genetics*. 2009;5(5):e1000470.
- Vavouri T., Lehner B. Human genes with CpG island promoters have a distinct transcription-associated chromatin organization. *Genome Biology*. 2012;13(11):R110.
- Wahl M.C., Will C.L., Luhrmann R. The spliceosome: design principles of a dynamic RNP machine. *Cell*. 2009;136:701-718.
- Waldholm J., Wang Z., Brodin D., Tyai A., Yu S., Theopold U., Kristin A., Farrants O. Visa N. SWI / SNF regulates the alternative processing of a specific subset of pre-mRNAs in *Drosophila melanogaster*. *BMC Molecular Biology*. 2011;12(1):46-60.
- Wang E.T., Sandberg R., Luo S., Khrebtkova I., Zhang L., Mayr C., Kingsmore S.F., Schroth G.P., Burge C.B. Alternative isoform regulation in human tissue transcriptomes. *Nature*. 2008;456(7221):470-476.
- Weiner A., Hughes A., Yassour M., Rando O.J., Friedman N. High-resolution nucleosome mapping reveals transcription-dependent promoter packaging. *Genome Research*. 2010;20(1):90-100.
- Workman J.L. Nucleosome displacement in transcription *Genes & Development*. 2006;20:2009-2017.
- Wu Q., Wang J., Yan H. Prediction of nucleosome positions in the yeast genome based on matched mirror position filtering. *Bioinformatics*. 2009;3(10):454-459.
- Young J.I., Hong E.P., Castle J.C., Crespo Barreto J., Bowman A.B., Rose M.F., Kang D., Richman R., Johnson J.M., Berget S., Zoghbi H.Y. Regulation of RNA splicing by the methylation-dependent transcriptional repressor methyl-CpG binding protein 2. *PNAS*. 2005;102(49):17551-17558.
- Zhang Y., Griffin K., Mondal N., Parvin J.D. Phosphorylation of histone H2A inhibits transcription on chromatin templates. *J. Biol Chem*. 2004 May 21;279(21):21866-72.
- Zentner G.E., Henikoff S. Regulation of nucleosome dynamics by histone modifications. *Nature Structural & Molecular Biology*. 2013;20(3):259-266.
- Zhou Z., Fu X. Regulation of splicing by SR proteins and SR protein-specific kinases. *Chromosoma*. 2013;122:191-207.

Chapter 24

ALTERNATIVE RNA SPLICING AND REGULATION OF NITRIC OXIDE SIGNALING

Iraida G. Sharina*

Department of Internal Medicine, Division of Cardiology,
University of Texas Medical School, Houston, TX, US

ABSTRACT

Alternative splicing expands transcriptome diversity and allows cells to meet the requirements of an ever-changing extracellular environment. It has been more than 30 years since nitric oxide (NO), a gaseous free radical, was recognized as a critical physiologic signaling molecule. Since then the list of known NO-directed functions has grown substantially to include regulation of smooth muscle function in vascular and gastrointestinal systems, inhibition of platelet aggregation and adhesion, neurotransmission and neuromodulation, regulation of cellular respiration and cytotoxicity, mitochondrial biogenesis and, immune defense. However, the importance of alternative splicing in regulation of enzymatic components of NO signaling pathway started to emerge only recently. Our understanding of the mechanisms governing this process remains very limited and awaits systematic investigation. In this chapter we will attempt to summarize the available information on alternative splicing of major enzymes mediating canonical NO transduction through the secondary messenger cGMP. We will highlight evidence accumulated from different laboratories that suggest splicing of enzymes in the NO/cGMP pathway, including nitric oxide synthases, heterodimeric soluble guanylylcyclase and cGMP-dependent protein kinase, is very complex and strongly affects NO signaling in response to various environmental cues. Future studies will certainly bring new, exciting insights into the role that alternative splicing plays in NO/cGMP biology.

Keywords: nitric oxide, soluble guanylylcyclase, splicing, regulation

* Corresponding Author's Email: Iraida.G.Sharina@uth.tmc.edu (Iraida G. Sharina, PhD, The University of Texas Medical School at Houston, 141 East Road, Houston, TX-77054; Tel: 713-486-2480; FAX: 713-486-0450).

ABBREVIATIONS

NO,	nitric oxide;
eNOS,	endothelial;
nNOS,	neuronal;
iNOS,	inducible Nitric Oxide Synthases;
sGC,	soluble guanylylcyclase;
cGMP,	cyclic guanosine monophosphate;
cAMP,	cyclic adenosine monophosphate;
TF,	transcription factor;
UTR,	untranslated region;
ORF,	open reading frame; aa, amino acid.

INTRODUCTION

Pre-mRNA splicing is one of the major modes of regulation of gene expression in metazoan organisms. Alternative pre-mRNA processing may generate numerous transcripts from a single protein-coding gene; thus largely increasing the information content and complexity of the genome. Recent studies indicate that more than 80% of human genes undergo alternative splicing [1, 2]. It is estimated that close to 75% of alternative exons introduce changes in the resulting protein products. Similar to transcription, splicing can be regulated both on a global level and in a gene-specific manner by a wide number of cell signaling mechanisms [2, 3]. Alternative splicing enables cells to tailor the expressed proteome in response to environmental stresses and to meet physiological requirements of different tissues. The importance of this process is underlined by a number of human diseases associated with alterations in splicing and splicing regulation [4, 5].

In this review we focus on the role of alternative splicing in regulation of the Nitric Oxide (NO) signaling pathway. We also summarize available information on splicing of major enzymes participating in NO/cGMP signal transduction.

NO/cGMP PATHWAY AND ITS ROLE IN PHYSIOLOGY

It has been more than 30 years since the gaseous free radical nitric oxide was recognized as a critical physiologic signaling molecule.

Biological functions of NO are mediated through several types of molecular interactions, which are traditionally divided into two groups: (1) direct effects of NO, such as reactions with hemoproteins, thiols and superoxide, and (2) indirect effects mediated by an increase in intracellular cyclic GMP level (Figure 1). In this chapter we will focus our attention on the second type, “classical” NO signal transduction mediated by cGMP. The primary source of enzymatic production of NO is via the family of Nitric Oxide Synthases (NOS), which includes endothelial (eNOS), neuronal (nNOS) and inducible (iNOS) isoforms. High output iNOS is expressed primarily in circulating phagocytic cells of immune defense. iNOS is constitutively active and produces high concentrations of NO that functions as a non-specific cytotoxic agent

in anti-microbial and anti-parasitic defense [6]. NO produced by iNOS is not mediated by cGMP, and prolonged activation of iNOS in systemic inflammation results in a precipitous drop of blood pressure in septic shock.

Transient increase in intracellular concentrations of Ca^{2+} is an important regulatory feature in regulation of eNOS and nNOS activity. These enzymes were originally identified in endothelial and neuronal tissues where they produce low concentration fluxes of NO needed to initiate the NO/cGMP signaling cascade. Additional processes, e.g., phosphorylation of eNOS by Akt signaling kinase [7], may modulate their ability to generate NO. Generated NO binds to the heme moiety of soluble guanylyl cyclase (sGC) and activates the enzyme to form the secondary messenger cyclic guanosine monophosphate (cGMP) from guanosine triphosphate (GTP). sGC is the only NO-sensitive guanylylcyclase and represents a unique receptor for this gaseous transmitter. Increased intracellular cGMP concentrations regulate the activity of several cGMP-dependent effectors, including cGMP-binding protein kinases, cGMP-dependent phosphodiesterases, and cyclic nucleotide gated ion channels [8].

Activation of protein kinase G (PKG I α, β) plays a central role in active relaxation of smooth muscle in vascular and gastric tissues. PKG reduces muscle contractile tone through several mechanisms. PKG facilitates Ca^{2+} extrusion to extracellular space and the reuptake of Ca^{2+} to sarcoplasmic reticulum cisterns by activation of Ca^{2+} /ATPase pumps. It also prevents Ca^{2+} influx by inhibiting the activation of Big Potassium Ca^{2+} -activated channels (BK_{Ca}).

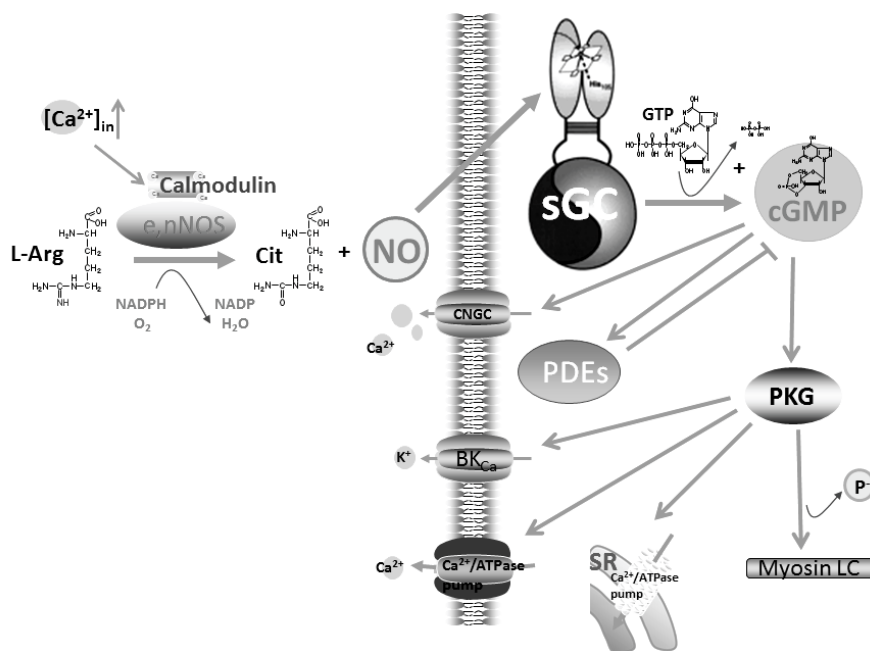


Figure 1. Molecular mechanisms of nitric oxide signaling via cGMP. NO and Citrulline (Cit) are synthesized from L-Arginine (L-Arg) by NO synthases (eNOS, nNOS) located in neuronal, endothelial or other cells. NO synthases are activated by Calmodulin in complex with Ca^{2+} . Produced NO traverses cell membranes and binds and activates soluble guanylylcyclase (sGC). cGMP produced by sGC binds and activates cGMP-dependent protein kinase (PKG), cyclic nucleotide gated ion channels (CNGC) and cGMP phosphodiesterases (PDEs). PKG phosphorylates and activates Ca^{2+} ATPase pumps in sarcoplasmic reticulum (SR) and plasma membrane, and big potassium channels (BK_{Ca}) in plasma membrane. These processes initiate and amplify NO/cGMP signal transduction.

In addition, PKG facilitates dephosphorylation of myosin light chains via stimulation of myosin light chain phosphatase activity, thus directly inducing relaxation of smooth muscle cells (SMC) [9].

cGMP-regulated phosphodiesterases (PDEs) are another important group of cGMP-effector proteins. PDEs play an important role in determining the actual levels of intracellular cyclic nucleotides in virtually all type of tissues by catalyzing the breakdown of intracellular levels of cGMP and cAMP[9].

Cyclic-nucleotide gated channels (CNG) are directly activated by cGMP and constitute a special class of downstream cGMP effectors. Upon cGMP binding, CNG channels nonselectively allow the passage of divalent cations through plasma membrane which mediates sensory functions such as olfaction and photo-reception [10].

Since the discovery of the NO/cGMP signaling pathway the list of known physiological functions mediated by it has grown substantially. The eNOS-sGC axis plays a crucial role in cardiovascular homeostasis. NO is an important endogenous vasodilator that regulates blood vessel tone through activation of sGC. NO/cGMP signaling prevents thrombosis by inhibiting platelet aggregation and leukocytes adhesion, thereby increasing blood fluidity [11, 12]. In addition, NO participates in vascular remodeling through inhibition of vascular SMC proliferation and neointima formation in vascular injury [13]. Therefore, impairment of NO/cGMP signaling is regarded as a hallmark of endothelial dysfunction and contributes to the development of a vast majority of cardiovascular disorders including coronary artery disease, atherosclerosis and hypertension [14]. Activation of the NO/cGMP pathway results in relaxation of smooth muscle of corpus cavernosum resulting in penile erection. Clinical introduction of PDE5 inhibitors, most notably Viagra (sildenafil), which potently and selectively inhibits cGMP hydrolysis, was a landmark for the treatment of male erectile dysfunction [15].

The NO/cGMP pathway promotes angiogenesis by activating proliferation and migration of endothelial cells [16]. It also participates in cellular metabolism by acutely stimulating uptake and oxidation of glucose and fatty acids; inhibiting synthesis of glucose, glycogen and fat in skeletal muscle; and enhancing lipolysis in white adipocytes [17]. In addition, NO/cGMP signaling regulates the permeability of tight junctions facilitating transit across gut and blood-brain barrier and maintaining the homeostasis of the microenvironment in the seminiferous epithelium during spermatogenesis [18].

Neuronal NOS is the isoform responsible for non-adrenergic, non-cholinergic (NANC) inhibitory neurotransmission in the gut. nNOS and sGC deficient animal models exhibit strong gastrointestinal phenotype characterized by severe or fatal gastrointestinal obstruction [19, 20]. Dysfunctional nNOS enzyme has been implicated in several diseases of the GI tract, including oesophageal achalasia, pyloric stenosis, and colonic dysfunction [21].

The canonical NO/cGMP pathway modulates long-term changes of synaptic activity in numerous brain regions. It contributes to distinct forms of learning and memory, such as fear conditioning, motor adaptation, and object recognition. cGMP generated upon NO activation of sGC promotes presynaptic transmitter release through regulation of Ca^{2+} signaling and cytoskeletal rearrangements. In the postsynaptic junction, cGMP modulates trafficking and subunit composition of α -amino-3-hydroxy-5-methyl-4-isoxazolepropionic acid (AMPA) receptors. Phosphorylation of glutamate receptor 1 (GluR1) by PKG increases its surface expression in hippocampal pyramidal cells, while in cerebellar Purkinje cells PKG induces endocytotic removal of AMPARs from postsynaptic membranes. Dysfunction of

NO/cGMP signaling is implicated in the development of neurodegenerative and psychiatric diseases including Morbus, Alzheimer's and schizophrenia [22].

NO/cGMP signaling mediates nociceptive processing of pain signals in the superficial dorsal horn of spinal cord, and is an important contributor to sensitization during inflammatory and neuropathic pain. Selective inhibition of this pathway in spinal cord holds promise for the development of new powerful analgesics [23].

A detailed description of all the functions of the NO/cGMP pathway is out of the scope of this chapter. There is a multitude of excellent scientific reviews on this subject, some of which were referenced above. Here, we have tried to draw the attention of the reader to the notion that NO signaling via cGMP is highly ubiquitous in metazoan physiology and is undeniably important in human health and disease. Preliminary insights indicate that regulation of NO/cGMP pathway is rather complex and occurs at all levels, including transcription, post-transcriptional regulation, translation and protein modification [24-27]. In this chapter we will attempt to summarize the available data regarding splicing of the enzymes which represent the core molecular machinery of the NO signal transduction. We will try to underline the fascinating complexity of NO/cGMP regulation, most of which still awaits discovery and characterization.

NEURONAL NITRIC OXIDE SYNTHASE (nNOS)

Neuronal NOS was the first NO synthase to be purified and cloned from brain tissue. It is most abundantly expressed in central and peripheral nervous tissue, and plays a central role in neurotransmission and neuromodulation [28]. Its gene locus encompasses more than 200 kb and has highly complex structural organization. nNOS splice forms are generated through alternative splicing by exon insertions/deletions and involve usage of multiple promoters [29]. Four nNOS protein isoforms (five including canonical, full size nNOS- α) encoded by multiple splice form transcripts possess different regulatory and catalytic properties [30]. The nNOS- α and nNOS- μ variants contain the postsynaptic density-95/discs large/zone occludens-1 homology (PDZ) domain. PDZ domain allows enzyme to associate with cell membranes, thereby facilitating its activation by calcium fluxes [31]. In addition to defining intracellular localization, the PDZ domain plays an important role in regulating the activity of nNOS isoforms by mediating protein-protein interactions with protein inhibitor of NOS (PIN) [32], the NMDA receptor [33], syntrophin [31], and carboxyl terminal PDZ ligand of NOS (CAPON) [34]. nNOS- α "canonical" protein has highest catalytic activity and is encoded by multiple splice transcripts with alternative UTR regions. nNOS- μ contains a unique 34 aa insertion and possess similar enzymatic properties as nNOS- α isoform. It is predominantly expressed in myocardium and skeletal muscle in rats, and in rat and human penis and urethra [35]. nNOS- β and nNOS- γ isoforms lack the PDZ domain and are localized primarily in the cytoplasm. While catalytic activity of nNOS- β is comparable to the activity of nNOS- α , the activity of nNOS- γ is significantly impaired and represents no more than 3% of nNOS- α 's activity [31]. nNOS- β isoform was demonstrated to mediate penile erection [36]. An additional splice variant called nNOS-2 with a 105 aa deletion was identified in mouse brain and human neuroblastoma cell lines [37, 38].

Precise biological functions of individual nNOS isoforms remain unclear. Expression profiling indicates that some of them participate in the development of various pathologies. For example, nNOS- γ and nNOS- β are upregulated in amyotrophic lateral sclerosis and might play a role in mediating oxidative damage in neuronal tissue [39]. Also, the expression of nNOS- γ splice form was detected in atherosclerotic plaques of apoE knock-out mice, and was suggested to play a role in plaque formation [40]. One of the alternative nNOS- α transcripts is regulated by a hypoxia-inducible promoter, which confers the ability to rapidly upregulate nNOS expression in response to hypoxia [41]. Selective downregulation of nNOS- α and nNOS-2 transcripts indicates that these splice variants may play pharmacologically different roles in modulation of morphine analgesia [42]. Thus, nNOS alternative splicing seems to be subject to diverse expressional regulation and plays a crucial role in modulating nNOS function and activity.

ENDOTHELIAL NITRIC OXIDE SYNTHASE (eNOS)

eNOS is a key regulator of vascular homeostasis and its activity is strongly implicated in maintenance of a healthy vascular phenotype. Until very recently no information was available about alternative splicing of the eNOS gene. However, the dry spell was broken by a very interesting report demonstrating that the human eNOS gene does undergo alternative splicing in intron 13, which results in the expression of three eNOS splice transcripts designated eNOS13A, B and C [43]. Recombinant expression and functional analysis indicate that eNOS13A protein heterodimerizes with canonical eNOS monomer which results in reduction of eNOS activity. Therefore, it might play the role of a dominant negative mutant. Expression of all three eNOS splice variants has been detected in human endothelial cells and in most examined human tissues. High levels of eNOS13A isoform expression was detected in human testis. Additionally, presented results indicate that alternative splicing of eNOS exon 13 is regulated by adjacent A/C-rich intronic splicer-enhancer elements, providing a unique glimpse into the mechanism of this regulation [43]. In a follow up report, the mechanism of eNOS alternative splicing was dissected further, demonstrating that DNA topoisomerase I is directly involved in this process [44]. Inhibition of DNA topo I activity by siRNA treatment abolished the TNF- α -induced increase in expression of inhibitory eNOS splice forms, and prevented the decline of NO biosynthesis in human endothelial cells. This data suggested that alternative splicing of eNOS might be an additional mechanism modulating the rate of NO production in an early response to inflammatory signaling.

NO-INDUCIBLE HETERODIMERIC SOLUBLE GUANYLYL CYCLASE (sGC)

sGC is a heme-containing obligatory heterodimer composed of α and β subunits [45]. Its expression has been detected in all studied tissues, albeit at varying levels and with different subunit compositions [46]. Although sGC was originally purified from cytosolic fraction, it was subsequently found associated with cellular membranes in a variety of cell types [47-53]. Each sGC subunit has two isoforms $\alpha 1/\alpha 2$ and $\beta 1/\beta 2$, encoded by a total of 4 separate genes.

Only sGC heterodimers containing the $\beta 1$ subunit have been shown to possess catalytic activity in vivo [54]. While α subunits ($\alpha 1$ and $\alpha 2$) have similar catalytic properties, they differ in their tissue and subcellular distribution [46, 51, 54]. Several studies have demonstrated that transcripts derived from all sGC genes undergo alternative splicing [55-60].

The $\alpha 1$ sGC gene is found to be expressed in all examined tissues with the highest levels detected in vasculature. To date, $\alpha 1$ splice variants are the most diverse and best characterized group [27]. The variety of human $\alpha 1$ transcripts is consolidated into seven transcript types encoding four different protein isoforms which are named A, B, C and D (NCBI, www.ncbi.nlm.nih.gov). All $\alpha 1$ splice form transcripts can be sub-divided into two groups based on the encoded protein isoforms. The first group includes transcripts 1 to 4, which encode an identical 690 aa protein, isoform A (canonical $\alpha 1$ sGC). These transcripts differ from each other only in their 5'- and 3'-UTR sequences, which are likely to play roles in regulating mRNA stability. The second group includes transcripts 5 to 7 which encode unique polypeptides named isoforms B, C and D. All three $\alpha 1$ sGC protein splice forms have been biochemically characterized and shown to generate heterodimers with $\beta 1$ sGC subunit resulting in diverse functional properties [60, 61]. The N1- $\alpha 1$ variant, Transcript 6, isoform C, encodes $\alpha 1$ sGC protein with extensive deletion in the catalytic domain. It displays properties of a dominant negative mutant and inhibits the activity of the $\alpha 1/\beta 1$ sGC heterodimer. In contrast, the C- $\alpha 1$ sGC splice form, Transcript 5, isoform B, encodes a fully functional protein despite a 240 aa deletion of the N-terminal regulatory domain. sGC heterodimers containing C- $\alpha 1$ sGC are much less sensitive to protein degradation induced by sGC heme oxidizing agent ODQ. C- $\alpha 1$ sGC also partially protects the level of $\beta 1$ sGC, when co-expressed in BE2 cells [60]. The C- $\alpha 1$ sGC splice form is expressed at high levels during differentiation of human embryonic stem cells (hES) [59]. Intracellular distribution of C- $\alpha 1$ isoform varies from canonical $\alpha 1$ sGC [59], and is localized to the more oxidized environment of endoplasmic reticulum [61]. The relative amount of C- $\alpha 1$ transcript was increased by exposure to H_2O_2 in several human cell lines constitutively expressing $\alpha 1/\beta 1$ sGC [62]. The $\alpha 1$ isoform D (Transcript 7) generates a sGC heterodimer with drastically impeded activation by NO and NO-independent sGC activators. Interestingly, our recent studies demonstrate increased non-functional $\alpha 1$ sGC splice forms, $\alpha 1$ -IsoC and IsoD, and decreased functional oxidation-resistant C- $\alpha 1$, $\alpha 1$ -IsoB, variant expression in diseased aortic tissue collected from patients undergoing aortic repair surgery (Sharina *et al*, submitted). These results suggest that $\alpha 1$ sGC splicing likely plays an important role in regulation of sGC activity, and could be one of the mechanisms contributing to impaired NO/cGMP signaling in vascular disease.

The $\alpha 2$ sGC transcripts have more restricted patterns of expression and are highly abundant in neuronal tissues [46]. The $\alpha 2i$ splice variant demonstrates properties of a dominant negative mutant when co-expressed with canonical subunits [55]. Recently, the existence of a splice variant homologous to the human $\alpha 2i$ has been detected in rat pituitary and liver tissues [57]. In that report, semi-quantitative RT-PCR demonstrated that $\alpha 2i$, but not $\alpha 2$ sGC transcript, is upregulated in response to estrogen in adult rat pituitary gland. This interesting observation suggests that the splicing of the $\alpha 2$ sGC subunit might be regulated by sex hormones in rodents.

The ubiquitously expressed $\beta 1$ sGC is indispensable for sGC activity [19, 46, 63]. Heterogeneity of $\beta 1$ sGC transcripts was first demonstrated by Chhajlani *et al.* who showed the presence of two mRNA species in human lung tissue [58]. In addition to the canonical transcript, they observed a shorter mRNA (NCBI GenBank Accession AF020340,) predicted to encode a non-functional peptide containing 33 aa deletion in the proximity of a catalytic

domain. Presently several $\beta 1$ sGC splice form transcripts exist in the GeneBank database. Their detailed description can be found in the review [27]. Sequence analysis indicates that a number of unique $\beta 1$ polypeptides with multiple N- and C-terminal deletions and insertions are encoded by these splice transcripts. No functional characterization of $\beta 1$ sGC splice forms has been performed so far. Changes in protein sequence of the $\beta 1$ sGC encoded by these splice variants are likely to affect the regulatory properties of sGC heterodimers that would include them. Human $\beta 1$ sGC splice forms have been isolated from a variety of different tissues including lung, fetal brain, placenta, and tongue tumor, suggesting complex tissue specific splicing of the $\beta 1$ sGC gene.

To date, the biological role of the $\beta 2$ sGC subunit gene remains a mystery [55, 64]. The coding region for the $\beta 2$ subunit lacks the methionine codon, and the expression of $\beta 2$ protein has yet to be detected *in vivo*. According to several publications, the recombinant $\alpha 1/\beta 2$ heterodimer possesses very low, if any, enzymatic activity and was suggested to act as a dominant negative isoform to other subunits [65, 66]. It is the only known sGC subunit which is reported to possess activity as a homodimer [67]; this resembles particulate guanylyl cyclases. Similar to other sGC subunits, the $\beta 2$ pre-mRNA undergoes alternative splicing. Several $\beta 2$ splice variants have been described in the literature. Two alternative transcripts encode the N-terminally truncated splice isoforms corresponding to the splice variants originally described by Behrends' group (Acc. Ns CR615161 and AK296746). The β_{2SV} sGC, which is generated by the skipping of exons encoding the N-terminal heme-binding domain of the protein, was isolated from human corpus cavernosum [68]. Based on deletion size and location, it was suggested that sGC enzymes containing this β_{2SV} splice form should be NO-insensitive. Another splice form encoding an N-terminally truncated $\beta 2$ sGC protein was isolated from human gastric carcinoma [56]. In a recent report, a new splice variant encoding an ORF with additional amino acids at the N-terminus was described [69]. Interestingly, the same report demonstrated that the 5'UTRs of the two alternative $\beta 2$ sGC transcripts ('canonical' and new splice form) possess an active internal ribosome entry site (IRES). Moreover, the activity of the 5'UTR from the alternative transcript was much higher than that of the 'canonical' one, supporting the concept that alternative splicing plays a role in modulation of $\beta 2$ sGC expression.

cGMP-DEPENDENT PROTEIN KINASE G (PKG I)

The biological importance of PKG in transducing NO signaling was first appreciated in the promotion of vascular smooth muscle relaxation and inhibition of platelet aggregation. Numerous advances in methodology and generation of PKG deficient mice extended the roles of PKGI to regulation of gastrointestinal motility, endothelial permeability, erectile function, cardiac protection, and modulation of proliferation of SMC [9]. Two PKG enzymes (PKGI and PKGII) are encoded by two separate genes in mammals. We will focus on PKGI since it is the main cGMP-dependent kinase which transduces the NO signal. There are two well described splice variants of PKGI, designated as PKGI α and PKGI β . These splice proteins differ by a unique stretch of 100 amino acids in the N-terminal regulatory domains which share only 36% homology. This difference exerts a profound influence on PKGI α and PKGI β cGMP binding affinity, selectivity of enzymes to cyclic nucleotides analogs, specificity towards protein

substrates, subcellular localization, and state of activation [9]. The leucine zipper encoded in the N-terminal regulatory domain appears to be responsible for selective protein–protein interactions of PKGI isoforms. For example, only PKGI α , but not PKGI β , associates with antidepressant-sensitive serotonin transporter (SERT) to modulate serotonin uptake [70]. Interaction with the leucine zippers of PKGI α dimers with myosin light chain phosphatase (MLCP) is required for MLCP phosphorylation and activation [71]. The leucine zipper of PKGI α mediates binding and activation of Regulator of G-protein Signaling-2 (RGS2), facilitating vascular smooth muscle cell relaxation [72]. One α -helix located in the leucine zipper of PKGI β mediates specific binding and phosphorylation of two protein substrates, transcription factor II-I (TFII-I) and IP₃ receptor-associated PKG substrate (IRAG) [73]. Association with these proteins also determines PKGI β intracellular localization.

PKGI α binds cGMP with almost 10 fold greater affinity than PKGI β , which probably reflects different physiological roles for these isoforms [74]. It has been proposed that increased expression of PKGI β , with its lower affinity for cGMP, is responsible for the blunted vasodilation response to NO signaling in an angiotensin-hypertensive rat model [75].

Both PKGI isoforms function as homodimers and are found in cytosolic and membrane fractions, segregating into specific cellular compartments in particular cell types. PKGI β is the predominant isoform in platelets, where it is found almost entirely in membrane-bound form; PKGI α is the major isoform expressed in pulmonary vasculature [76]. PKGI α has been reported to protect neuronal cells from apoptosis by phosphorylation of Bad, an apoptosis-regulating BCL-2 family member [77]. One of the most intriguing features of PKGI α is its ability to be activated by oxidative stress, such as exposure to H₂O₂, which induces inter-subunit disulfide bonds and alters enzyme affinity for substrates [78]. Smooth muscle cells expressing redox-insensitive enzyme demonstrated no detectable phosphorylation of myosin light chain (MLC), suggesting that this mechanism plays an important role in regulation of PKGI α function.

cGMP PHOSPHODIESTERASES (PDEs)

PDE's hydrolyze the cyclic phosphate ring in cGMP and cAMP molecules. The products of this reaction, 5'-GMP or 5'-AMP, are inactive as second messengers in cyclic nucleotide-dependent signaling pathways. Therefore, PDEs serve as feedback regulators and terminators of NO/cGMP signaling, and play a crucial role in determining the intracellular cGMP levels. The expression and hydrolytic activities of PDEs, found virtually in all cell types and in almost all subcellular compartments, are subject to complex and dynamic regulatory processes. Mammalian PDEs are represented by 11 families of enzymes derived from 21 genes. Of them, 8 (PDE1 to 3, PDE5 and 6, and PDE9 to 11) are cGMP-hydrolyzing [9]. All PDE genes undergo alternative splicing, generating more than 50 protein isoforms with mostly unknown biological roles. PDEs have highly homologous catalytic domains, but differ in their diverse regulatory N-terminal and extreme C-terminal regions, which function as modulators of activity and protein-protein interactions. Three PDE families, PDE 5, 6, and 9 have 100-fold higher substrate affinities towards cGMP over cAMP and are considered cGMP-specific PDEs [79]. Here we focus our attention on cGMP-specific members of the PDE family. Information on other PDEs can be found in an excellent review by Bender and Beavo [80].

PDE5 is best known as the molecular target for several very successful drugs used to treat erectile dysfunction, such as Sildenafil (Viagra). The protein structure of canonical PDE5 is characterized by the presence of two high affinity binding sites for cGMP in its N-terminal regulatory domain named GAF-A and GFF-B. Three splice variants of PDE5, PDE5A1 to 3, all with alterations in N-terminal domain encoded by alternative first exons, have been identified so far. Transcription of PDE5A2 occurs from an alternative promoter and is regulated by cAMP and cGMP via AP2 and SP1 elements [81]. PDE5 isoforms also demonstrate differential tissue-specific expression. PDE5A1 and A2 splice forms have been detected in multiple tissues while PDE5A3 seems to be restricted only to vascular smooth muscle [82].

PDE6 family members are expressed in the photoreceptor outer segments of mammalian retina, where they play a critical role in the conversion of light signals into photo responses. No functional splice variants for PDE6 genes have been identified so far. However, exon 14 skipping in PDE6B gene was found to be responsible for the drastic decrease in PDE6B protein expression in an atypical retinal degeneration 3 (*atrd3*) mouse model of recessive retinitis pigmentosa [83]. This report draws attention to the importance of faithful mRNA splicing and the deleterious effects of its malfunction in genetic diseases.

PDE9 has the highest affinity for cGMP and is found co-localized with sGC and nNOS in the brain, strongly indicating that it participates in regulation of NO signaling in neuronal tissue [84]. The PDE9A gene locus maps to the region involved in several neurologic diseases including bi-polar disorder [85]. A PDE9 inhibitor is presently under pre-clinical investigation as a possible treatment for Alzheimer's disease [86]. PDE9A pre-mRNA undergoes extremely complex alternative splicing with at least 20 different splice forms identified so far [87]. All variants are alternatively spliced to produce unique changes in their N-terminal protein sequence. Only PDE9A1 and PDE9A5 protein isoforms have been characterized so far. Both isoforms have a high affinity for cGMP, similar K_m values, and 2 fold differences in V_{max} values. They also demonstrate different subcellular localization: PDE9A1 is found preferentially in the nucleus, while PDE9A5 is found in cytosol [88]. The biological roles of other PDE9A splice forms remain to be determined.

CYCLIC NUCLEOTIDE GATED ION CHANNELS (CNGCs)

cGMP produced by NO-activated sGC may directly activate CNG channels. CNG channels are nonselective cation channels; they do not discriminate between alkali ions and allow the passage of divalent ions, including Ca^{2+} . Activation of CNG channels causes membrane depolarization and facilitates excitation of neurons [89]. CNGCs serve as molecular sensors responding to changes in intracellular cGMP concentrations in visual and olfactory systems. In addition, they are found to play roles in neurotransduction and neuromodulation in other brain regions such as hippocampus, the striatum, the hypothalamus, and the locus coeruleus.

CNGCs function as heterotetrameric complexes consisting of two or three different types of subunits. The CNG family is loosely divided into two types of subunits, called A and B subunits. When expressed alone in heterologous systems, A subunits form functional channels, whereas B subunits have been shown to function only as activity modulators of homomeric α subunit channels. They are encoded by six different genes, four for A subunits (A1 to A4) and two for B subunits (B1 and B3). Multiple splice variants have been identified for the CNGA3

and the CNGB1 subunits from various tissues and organisms [90]. Splice forms of A3 gene encode protein isoforms with different N-terminal domains. However, as with many other examples in the NO/cGMP signaling pathway, specific functional roles of splice isoforms remain unexplored. Two splice variants are known for CNGB1, B1a and B1b. These splice variants differ in the glutamic acid-rich part of their N-terminal domains. These domains are responsible for protein-protein interaction and demonstrate tissue-specific expression in rods, photoreceptors, and within the olfactory system [91]. The expression of B1 splice forms was also detected in testis where they demonstrate different localization and sensitivity to cyclic nucleotides. It has been suggested that they play a part in controlling the flagella bending waves in sperm [92].

CONCLUSION

The last 20 years of research in splicing regulation of NO/cGMP signaling has progressed from virtually non-existent to a fast developing and very complex field. Data from multiple laboratories strongly indicate that splicing of NO/cGMP pathway enzymes is a convoluted process modulated by multiple environmental and cellular factors. Emerging evidence suggest that this regulation is crucially important to support NO signaling in normal and pathological conditions.

The diversity of the expressed transcripts allows for a dynamic spatio-temporal regulation of the NO/cGMP pathway in response to multiple physiological signals. Most of the enzymes in NO/cGMP signaling, including NOS, sGC, PKG, CNGs and PDEs, function as multimers, and are often composed of two types of different subunits. Incorporation of splice isoforms into functional multimers creates the potential to generate a great number of unique hetero- and homo-dimers with diverse enzymatic and regulatory properties. The assortment of available alternatively spliced transcripts or splice protein variants may tailor the properties of the enzyme(s) in the pathway to better suit the molecular environment of the cell. This consideration alone is a compelling argument for the importance of alternative splicing in functional regulation of the NO/cGMP pathway.

Alternative splicing fundamentally contributes to NO/cGMP enzyme expression not only by giving rise to alternative protein isoforms, but also by producing transcripts with different regulatory sequences in non-translated regions. These new sequences may affect the stability and translocation of mRNA, as well as its translation. Alternative mRNA transcripts that differ in their 3' and/or 5' UTR regions are likely to play a critical role in regulating mRNA steady-state levels through mechanisms of RNA stabilization and/or inhibitory RNA targeting.

Studies of the regulation of NO/cGMP enzymes through alternative splicing still remain in initial stages of development. Currently the components of splicing machinery involved in the regulation of sGC splicing remain almost completely unknown. Recently our group gained insight into the regulatory mechanism of H₂O₂ induction of the expression of oxidation resistant C- α 1 sGC splice form [62]. Our studies show that H₂O₂ exposure selectively decreases the expression of PTBP1 and hnRNP A2/B1 splicing factors. The same factors were identified by *in silico* analysis as potential regulators controlling the splicing of C- α 1 sGC. Another interesting discovery was made regarding the crucial role of DNA topoisomerase I in regulation of eNOS alternative splicing in response to TNF- α (see eNOS section above, [44]). However,

no systematic investigation of splicing regulation of NO/cGMP enzymes has been done so far. Increasing numbers of reports demonstrate that alternative splicing of different enzymes in the pathway may be altered by similar regulatory factors under inflammatory and oxidative conditions [39-41, 44, 62, 75, 93]. A coordinated regulation of splicing of different enzymatic components of NO/cGMP signaling would facilitate cellular adaptation to oxidative stress and/or the development of diverse pathologies associated with it. In the future, a major challenge will be to define the mechanisms and splicing regulators participating in such coordinated regulation.

Although the biological roles for the majority of splice variants of NO/cGMP enzymes are poorly understood, unraveling them will likely be essential to our understanding of nitric oxide biology. Advances in the understanding of the molecular mechanisms will facilitate the development of new therapeutic strategies allowing targeted manipulation of the pathway through the regulation of splicing of selected components.

REFERENCES

- [1] Ben-Dov, C., et al., Genome-wide analysis of alternative pre-mRNA splicing. *J. Biol. Chem.*, 2008. 283(3): p. 1229-33.
- [2] Blencowe, B.J., Alternative splicing: new insights from global analyses. *Cell*, 2006. 126(1): p. 37-47.
- [3] Shin, C. and J.L. Manley, Cell signalling and the control of pre-mRNA splicing. *Nat. Rev. Mol. Cell Biol.*, 2004. 5(9): p. 727-38.
- [4] Tazi, J., N. Bakkour, and S. Stamm, Alternative splicing and disease. *Biochim. Biophys. Acta.*, 2009. 1792(1): p. 14-26.
- [5] Wang, G.S. and T.A. Cooper, Splicing in disease: disruption of the splicing code and the decoding machinery. *Nat. Rev. Genet.*, 2007. 8(10): p. 749-61.
- [6] MacMicking, J., Q.W. Xie, and C. Nathan, Nitric oxide and macrophage function. *Annu. Rev. Immunol.*, 1997. 15: p. 323-50.
- [7] Luo, Z., et al., Acute modulation of endothelial Akt/PKB activity alters nitric oxide-dependent vasomotor activity in vivo. *J. Clin. Invest.*, 2000. 106(4): p. 493-9.
- [8] Waldman, S.A. and F. Murad, Cyclic GMP synthesis and function. *Pharmacol. Rev.*, 1987. 39(3): p. 163-96.
- [9] Francis, S.H., et al., cGMP-dependent protein kinases and cGMP phosphodiesterases in nitric oxide and cGMP action. *Pharmacol. Rev.*, 2010. 62(3): p. 525-63.
- [10] Lucas, K.A., et al., Guanylylcyclases and signaling by cyclic GMP. *Pharmacol. Rev.*, 2000. 52(3): p. 375-414.
- [11] Dangel, O., et al., Nitric oxide-sensitive guanylylcyclase is the only nitric oxide receptor mediating platelet inhibition. *J. Thromb. Haemost.*, 2010. 8(6): p. 1343-52.
- [12] Radomski, M.W., R.M. Palmer, and S. Moncada, Endogenous nitric oxide inhibits human platelet adhesion to vascular endothelium. *Lancet*, 1987. 2(8567): p. 1057-8.
- [13] Jeremy, J.Y., et al., Nitric oxide and the proliferation of vascular smooth muscle cells. *Cardiovasc. Res.*, 1999. 43(3): p. 580-94.
- [14] Pacher, P., J.S. Beckman, and L. Liaudet, Nitric oxide and peroxynitrite in health and disease. *Physiol. Rev.*, 2007. 87(1): p. 315-424.

- [15] Sandner, P., D. Neuser, and E. Bischoff, Erectile dysfunction and lower urinary tract. *Handb. Exp. Pharmacol.*, 2009(191): p. 507-31.
- [16] Pyriochou, A., et al., Soluble guanylylcyclase activation promotes angiogenesis. *J. Pharmacol. Exp. Ther.*, 2006. 319(2): p. 663-71.
- [17] Dai, Z., et al., Nitric oxide and energy metabolism in mammals. *Biofactors*, 2013.
- [18] Lee, N.P. and C.Y. Cheng, Nitric oxide and cyclic nucleotides: their roles in junction dynamics and spermatogenesis. *Adv. Exp. Med. Biol.*, 2008. 636: p. 172-85.
- [19] Friebe, A., et al., Fatal gastrointestinal obstruction and hypertension in mice lacking nitric oxide-sensitive guanylylcyclase. *Proc. Natl. Acad. Sci. USA*, 2007. 104(18): p. 7699-704.
- [20] Kim, C.D., R.K. Goyal, and H. Mashimo, Neuronal NOS provides nitrenergic inhibitory neurotransmitter in mouse lower esophageal sphincter. *Am. J. Physiol.*, 1999. 277(2 Pt 1): p. G280-4.
- [21] Lundberg, J.O. and E. Weitzberg, Biology of nitrogen oxides in the gastrointestinal tract. *Gut*, 2013. 62(4): p. 616-29.
- [22] Kleppisch, T. and R. Feil, cGMPsignalling in the mammalian brain: role in synaptic plasticity and behaviour. *Handb. Exp. Pharmacol.*, 2009(191): p. 549-79.
- [23] Schmidtko, A., I. Tegeder, and G. Geisslinger, No NO, no pain? The role of nitric oxide and cGMP in spinal pain processing. *Trends Neurosci*, 2009. 32(6): p. 339-46.
- [24] Andreopoulos, S. and A. Papapetropoulos, Molecular aspects of soluble guanylylcyclase regulation. *Gen. Pharmacol.*, 2000. 34(3): p. 147-57.
- [25] Hobbs, A.J., Soluble guanylatecyclase: the forgotten sibling. *Trends Pharmacol. Sci.*, 1997. 18(12): p. 484-91.
- [26] Pyriochou, A. and A. Papapetropoulos, Soluble guanylylcyclase: more secrets revealed. *Cell Signal.*, 2005. 17(4): p. 407-13.
- [27] Sharina, I.G., et al., RNA splicing in regulation of nitric oxide receptor soluble guanylylcyclase. *Nitric. Oxide*, 2011. 25(3): p. 265-74.
- [28] Garthwaite, J., Concepts of neural nitric oxide-mediated transmission. *Eur. J. Neurosci.*, 2008. 27(11): p. 2783-802.
- [29] Boissel, J.P., P.M. Schwarz, and U. Forstermann, Neuronal-type NO synthase: transcript diversity and expressional regulation. *Nitric. Oxide*, 1998. 2(5): p. 337-49.
- [30] Alderton, W.K., C.E. Cooper, and R.G. Knowles, Nitric oxide synthases: structure, function and inhibition. *Biochem. J.*, 2001. 357(Pt 3): p. 593-615.
- [31] Brenman, J.E., et al., Interaction of nitric oxide synthase with the postsynaptic density protein PSD-95 and alpha1-syntrophin mediated by PDZ domains. *Cell*, 1996. 84(5): p. 757-67.
- [32] Hemmens, B., et al., The protein inhibitor of neuronal nitric oxide synthase (PIN): characterization of its action on pure nitric oxide synthases. *FEBS Lett.*, 1998. 430(3): p. 397-400.
- [33] Christopherson, K.S., et al., PSD-95 assembles a ternary complex with the N-methyl-D-aspartic acid receptor and a bivalent neuronal NO synthase PDZ domain. *J. Biol. Chem.*, 1999. 274(39): p. 27467-73.
- [34] Jaffrey, S.R., et al., CAPON: a protein associated with neuronal nitric oxide synthase that regulates its interactions with PSD95. *Neuron*, 1998. 20(1): p. 115-24.
- [35] Silvagno, F., H. Xia, and D.S. Bredt, Neuronal nitric-oxide synthase-mu, an alternatively spliced isoform expressed in differentiated skeletal muscle. *J. Biol. Chem.*, 1996. 271(19): p. 11204-8.

- [36] Hurt, K.J., et al., Alternatively spliced neuronal nitric oxide synthase mediates penile erection. *Proc. Natl. Acad. Sci. USA*, 2006. 103(9): p. 3440-3.
- [37] Fujisawa, H., et al., Expression of two types of nitric oxide synthase mRNA in human neuroblastoma cell lines. *J. Neurochem.*, 1994. 63(1): p. 140-5.
- [38] Ogura, T., et al., Structural diversity of neuronal nitric oxide synthase mRNA in the nervous system. *Biochem. Biophys. Res. Commun.*, 1993. 193(3): p. 1014-22.
- [39] Catania, M.V., et al., Increased expression of neuronal nitric oxide synthase spliced variants in reactive astrocytes of amyotrophic lateral sclerosis human spinal cord. *J. Neurosci.*, 2001. 21(11): p. RC148.
- [40] Schodel, J., et al., Expression of neuronal nitric oxide synthase splice variants in atherosclerotic plaques of apoE knockout mice. *Atherosclerosis*, 2009. 206(2): p. 383-9.
- [41] Ward, M.E., et al., Hypoxia induces a functionally significant and translationally efficient neuronal NO synthase mRNA variant. *J. Clin. Invest.*, 2005. 115(11): p. 3128-39.
- [42] Kolesnikov, Y.A., et al., Functionally differentiating two neuronal nitric oxide synthase isoforms through antisense mapping: evidence for opposing NO actions on morphine analgesia and tolerance. *Proc. Natl. Acad. Sci. USA*, 1997. 94(15): p. 8220-5.
- [43] Lorenz, M., et al., Alternative splicing in intron 13 of the human eNOS gene: a potential mechanism for regulating eNOS activity. *FASEB J.*, 2007. 21(7): p. 1556-64.
- [44] Eisenreich, A., et al., Effects of the Cdc2-like kinase-family and DNA topoisomerase I on the alternative splicing of eNOS in TNF-alpha-stimulated human endothelial cells. *Biol. Chem.*, 2008. 389(10): p. 1333-8.
- [45] Kamisaki, Y., et al., Soluble guanylatecyclase from rat lung exists as a heterodimer. *J. Biol. Chem.*, 1986. 261(16): p. 7236-41.
- [46] Mergia, E., et al., Major occurrence of the new alpha2beta1 isoform of NO-sensitive guanylylcyclase in brain. *Cell Signal.*, 2003. 15(2): p. 189-95.
- [47] Bellingham, M. and T.J. Evans, The alpha2beta1 isoform of guanylylcyclase mediates plasma membrane localized nitric oxide signalling. *Cell Signal.*, 2007. 19(10): p. 2183-93.
- [48] Feussner, M., et al., Association of soluble guanylatecyclase with the sarcolemma of mammalian skeletal muscle fibers. *Acta. Histochem.*, 2001. 103(3): p. 265-77.
- [49] Lin, L.H. and W.T. Talman, Soluble guanylatecyclase and neuronal nitric oxide synthase colocalize in rat nucleus tractussolitarii. *J. Chem. Neuroanat.*, 2005. 29(2): p. 127-36.
- [50] Pifarre, P., et al., NO-sensitive guanylylcyclase beta1 subunit is peripherally associated to chromosomes during mitosis. Novel role in chromatin condensation and cell cycle progression. *Int. J. Biochem. Cell Biol.*, 2009. 41(8-9): p. 1719-30.
- [51] Russwurm, M., N. Wittau, and D. Koesling, Guanylylcyclase/PSD-95 interaction: targeting of the nitric oxide-sensitive alpha2beta1 guanylylcyclase to synaptic membranes. *J. Biol. Chem.*, 2001. 276(48): p. 44647-52.
- [52] Venema, R.C., et al., Novel complexes of guanylatecyclase with heat shock protein 90 and nitric oxide synthase. *Am. J. Physiol. Heart Circ. Physiol.*, 2003. 285(2): p. H669-78.
- [53] Zabel, U., et al., Calcium-dependent membrane association sensitizes soluble guanylylcyclase to nitric oxide. *Nat. Cell Biol.*, 2002. 4(4): p. 307-11.
- [54] Russwurm, M., et al., Functional properties of a naturally occurring isoform of soluble guanylylcyclase. *Biochem. J.*, 1998. 335 (Pt 1): p. 125-30.

- [55] Behrends, S., et al., A variant of the alpha 2 subunit of soluble guanylylcyclase contains an insert homologous to a region within adenylyl cyclases and functions as a dominant negative protein. *J. Biol. Chem.*, 1995. 270(36): p. 21109-13.
- [56] Behrends, S. and K. Vehse, The beta(2) subunit of soluble guanylylcyclase contains a human-specific frameshift and is expressed in gastric carcinoma. *Biochem. Biophys. Res. Commun.*, 2000. 271(1): p. 64-9.
- [57] Cabilla, J.P., et al., Nitric oxide sensitive-guanylylcyclase subunit expression changes during estrous cycle in anterior pituitary glands. *Am. J. Physiol. Endocrinol. Metab.*, 2009. 296(4): p. E731-7.
- [58] Chhajlani, V., et al., Heterogeneity in human soluble guanylatecyclase due to alternative splicing. *FEBS Lett.*, 1991. 290(1-2): p. 157-8.
- [59] Sharin, V.G., et al., Nitric Oxide Receptor Soluble GuanylylCyclase Undergoes Splicing Regulation in Differentiating Human Embryonic Cells. *Stem Cells Dev.*, 2010.
- [60] Sharina, I.G., et al., Alpha1 soluble guanylylcyclase (sGC) splice forms as potential regulators of human sGC activity. *J. Biol. Chem.*, 2008. 283(22): p. 15104-13.
- [61] Kraehling, J.R., et al., The Amino-Terminus of Nitric Oxide Sensitive Guanylyl Cyclasealpha(1) Does Not Affect Dimerization but Influences Subcellular Localization. *PLoS One*, 2011. 6(9): p. e25772.
- [62] Cote, G.J., et al., Hydrogen peroxide alters splicing of soluble guanylylcyclase and selectively modulates expression of specific splicing regulators in human cancer cells. *PLoS One*, 2012. In Press.
- [63] Groneberg, D., et al., Smooth muscle-specific deletion of nitric oxide-sensitive guanylylcyclase is sufficient to induce hypertension in mice. *Circulation*, 2010. 121(3): p. 401-9.
- [64] Yuen, P.S., L.R. Potter, and D.L. Garbers, A new form of guanylylcyclase is preferentially expressed in rat kidney. *Biochemistry*, 1990. 29(49): p. 10872-8.
- [65] Gibb, B.J., V. Wykes, and J. Garthwaite, Properties of NO-activated guanylylcyclases expressed in cells. *Br. J. Pharmacol.*, 2003. 139(5): p. 1032-40.
- [66] Gupta, G., et al., The beta2 subunit inhibits stimulation of the alpha1/beta1 form of soluble guanylylcyclase by nitric oxide. Potential relevance to regulation of blood pressure. *J. Clin. Invest.*, 1997. 100(6): p. 1488-92.
- [67] Koglin, M., et al., Nitric oxide activates the beta 2 subunit of soluble guanylylcyclase in the absence of a second subunit. *J. Biol. Chem.*, 2001. 276(33): p. 30737-43.
- [68] Behrends, S., et al., Expression of nitric oxide-sensitive guanylylcyclase subunits in human corpus cavernosum. *Biochem. Pharmacol.*, 2000. 59(6): p. 713-7.
- [69] Vazquez-Padron, R.I., et al., An internal ribosome entry site mediates the initiation of soluble guanylylcyclase beta2 mRNA translation. *FEBS J.*, 2008. 275(14): p. 3598-607.
- [70] Steiner, J.A., et al., cGMP-dependent protein kinase Ialpha associates with the antidepressant-sensitive serotonin transporter and dictates rapid modulation of serotonin uptake. *Mol. Brain*, 2009. 2: p. 26.
- [71] Lee, E., et al., Interactions between the leucine-zipper motif of cGMP-dependent protein kinase and the C-terminal region of the targeting subunit of myosin light chain phosphatase. *J. Mol. Biol.*, 2007. 373(5): p. 1198-212.
- [72] Tang, K.M., et al., Regulator of G-protein signaling-2 mediates vascular smooth muscle relaxation and blood pressure. *Nat. Med.*, 2003. 9(12): p. 1506-12.

- [73] Casteel, D.E., G.R. Boss, and R.B. Pilz, Identification of the interface between cGMP-dependent protein kinase I β and its interaction partners TFII-I and IRAG reveals a common interaction motif. *J. Biol. Chem.*, 2005. 280(46): p. 38211-8.
- [74] Wolfe, L., J.D. Corbin, and S.H. Francis, Characterization of a novel isozyme of cGMP-dependent protein kinase from bovine aorta. *J. Biol. Chem.*, 1989. 264(13): p. 7734-41.
- [75] Gerzanich, V., et al., Alternative splicing of cGMP-dependent protein kinase I in angiotensin-hypertension: novel mechanism for nitrate tolerance in vascular smooth muscle. *Circ. Res.*, 2003. 93(9): p. 805-12.
- [76] Eigenthaler, M., et al., Concentration and regulation of cyclic nucleotides, cyclic-nucleotide-dependent protein kinases and one of their major substrates in human platelets. Estimating the rate of cAMP-regulated and cGMP-regulated protein phosphorylation in intact cells. *Eur. J. Biochem.*, 1992. 205(2): p. 471-81.
- [77] Johlfs, M.G. and R.R. Fiscus, Protein kinase G type-I α phosphorylates the apoptosis-regulating protein Bad at serine 155 and protects against apoptosis in N1E-115 cells. *Neurochem. Int.*, 2010. 56(4): p. 546-53.
- [78] Burgoyne, J.R., et al., Cysteine redox sensor in PKG1 α enables oxidant-induced activation. *Science*, 2007. 317(5843): p. 1393-7.
- [79] Bender, A.T. and J.A. Beavo, PDE1B2 regulates cGMP and a subset of the phenotypic characteristics acquired upon macrophage differentiation from a monocyte. *Proc. Natl. Acad. Sci. USA*, 2006. 103(2): p. 460-5.
- [80] Bender, A.T. and J.A. Beavo, Cyclic nucleotide phosphodiesterases: molecular regulation to clinical use. *Pharmacol. Rev.*, 2006. 58(3): p. 488-520.
- [81] Lin, C.S., et al., Regulation of human PDE5A2 intronic promoter by cAMP and cGMP: identification of a critical Sp1-binding site. *Biochem. Biophys. Res. Commun.*, 2001. 280(3): p. 693-9.
- [82] Lin, C.S., et al., Human PDE5A gene encodes three PDE5 isoforms from two alternate promoters. *Int. J. Impot. Res.*, 2002. 14(1): p. 15-24.
- [83] Muradov, H., et al., Atypical retinal degeneration 3 in mice is caused by defective PDE6B pre-mRNA splicing. *Vision Res.*, 2012. 57: p. 1-8.
- [84] Andreeva, S.G., et al., Expression of cGMP-specific phosphodiesterase 9A mRNA in the rat brain. *J. Neurosci.*, 2001. 21(22): p. 9068-76.
- [85] Guipponi, M., et al., Identification and characterization of a novel cyclic nucleotide phosphodiesterase gene (PDE9A) that maps to 21q22.3: alternative splicing of mRNA transcripts, genomic structure and sequence. *Hum. Genet.*, 1998. 103(4): p. 386-92.
- [86] Wunder, F., et al., Characterization of the first potent and selective PDE9 inhibitor using a cGMP reporter cell line. *Mol. Pharmacol.*, 2005. 68(6): p. 1775-81.
- [87] Rentero, C., A. Monfort, and P. Puigdomenech, Identification and distribution of different mRNA variants produced by differential splicing in the human phosphodiesterase 9A gene. *Biochem. Biophys. Res. Commun.*, 2003. 301(3): p. 686-92.
- [88] Wang, P., et al., Identification and characterization of a new human type 9 cGMP-specific phosphodiesterase splice variant (PDE9A5). Differential tissue distribution and subcellular localization of PDE9A variants. *Gene*, 2003. 314: p. 15-27.
- [89] Feil, R. and T. Kleppisch, NO/cGMP-dependent modulation of synaptic transmission. *Handb. Exp. Pharmacol.*, 2008(184): p. 529-60.
- [90] Kaupp, U.B. and R. Seifert, Cyclic nucleotide-gated ion channels. *Physiol. Rev.*, 2002. 82(3): p. 769-824.

- [91] Bonigk, W., et al., The native rat olfactory cyclic nucleotide-gated channel is composed of three distinct subunits. *J. Neurosci.*, 1999. 19(13): p. 5332-47.
- [92] Wiesner, B., et al., Cyclic nucleotide-gated channels on the flagellum control Ca^{2+} entry into sperm. *J. Cell Biol.*, 1998. 142(2): p. 473-84.
- [93] Richter, W., S.L. Jin, and M. Conti, Splice variants of the cyclic nucleotide phosphodiesterase PDE4D are differentially expressed and regulated in rat tissue. *Biochem. J.*, 2005. 388(Pt 3): p. 803-11.

Chapter 25

ALTERNATIVE SPLICING BY ANALYZING A HUMAN MRNA DIVERSITY USING DATA OF FLJ HUMAN cDNAs

Takao Isogai^{1,3,*} and Ai Wakamatsu^{2,3}

¹Fukushima Medical University, Fukushima, Japan

²Japan Biological Informatics Consortium, Tokyo, Japan

³Graduate School of Pharmaceutical Sciences, The University of Tokyo
Tokyo, Japan

ABSTRACT

Genes could produce multiple protein-coding transcripts by alternative splicing (AS). It was known that AS is related to several diseases in generating biological and functional diversity. So, analysis of the mRNA diversity of genes would be important for understanding gene function. Previously, we obtained 1.46 million human full-length cDNAs (FLJ cDNA) and sequenced their 5'-ends. We selected approximately 55 thousand cDNAs from FLJ cDNA and sequenced completely. Our FLJ cDNAs were constructed by an optimized oligo-capping method. Thus, by using 5'-EST data, a lot of valuable information was obtained regarding the diversity of the transcription start site (TSS) and amino acid sequences at the N-terminal ends of proteins. We found that alternative TSSs were utilized for tissue-specific expression. Using this data, we constructed FLJ Human cDNA Database ver. 3.2, <http://flj.lifesciencedb.jp>. But, a lot of AS-related information still remains to be extracted from our 1.4 million cDNA resources. From a huge number of human sequence information, we selected only the reliable cDNAs for the analysis of the mRNA diversity. And, we developed probes which can analyze the mRNA diversity. As a result, by comparing our constructed 5,784 pairs of independent probes, we are able to detect the expression profiles of splicing patterns in 3,413 genes. Moreover, using these probes, we analyzed the mRNA diversity of genes after inducing neuronal differentiation in human NT2 teratocarcinoma cells using all-trans retinoic acid (RA). Analyses of NT2 cells identified 452 RA-responsive genes. The mRNA diversity analysis revealed that the

* Corresponding Author's Email: tisogai@fmu.ac.jp (Tel: +81-24-547-1562; Fax: +81-24-547-1124).

rate of genes that showed AS in their N-terminus, internal region, C-terminus, is almost the same, respectively.

INTRODUCTION

We think that it is important to understand the variety of mRNA expressed from each gene for the gene functional analysis. Therefore we pay attention to it and analyze this. Because the human gene number was estimated to be only approximately 20,000 genes [1], it was thought that expressing different proteins from one gene frequently caused various life phenomena in humans. As for this variety, many genes change a splice pattern, and it is said that it is realized by alternative splicing (AS) producing multiple mature mRNA by one immature mRNA [2]. AS is a system where different kinds of proteins are often expressed from one gene. It is said that the gene expression program at this RNA stage is controlled for time space in the living organisms, and it is important for creating the complicated various forms and functions in those by this controlled program [3-8]. It is known that it is related to various diseases and that the proteins, where the biological property and function are different, are produced from one gene by AS [9]. When abnormality occurs in AS, which should be originally exact, an abnormal protein, which lacked in a function, and/or a cytotoxic abnormal protein, is produced and may be related to the disease onset [10-11]. Therefore at first it is necessary to analyze the variety of mRNA by each gene to study the influence that the change of the gene expression level gives to a gene function. However, there are many points that must be solved about the variety of the mRNA, including the function of every produced protein-coding transcript, by analyzing and identifying the various mRNA produced by one gene, the mRNA change from one gene by the environment factor and/or the controlling mechanism of producing the variety of mRNA transcribed from one gene.

Therefore first we must question what type of protein-coding transcript was produced by AS from one gene, so we identified various mRNA produced by a gene. Then, we decided to examine the influence that variety of mRNA gave to a gene function by AS to elucidate an unknown gene function and the gene expression control system.

IDENTIFICATION OF TRANSCRIPTION PRODUCTS PRODUCED BY VARIETY OF THE HUMAN mRNA

We thought that identification of various mRNA produced by one gene was important to the gene functional analysis. All human genome sequencing was determined by the human genome-sequencing project [1], but all gene regions transcribed to mRNAs, which encoded proteins, were not necessarily identified. Furthermore, the number of the mRNA, which can theoretically be produced by AS, is enormous because the gene is often comprised of plural exons and the combination is complicated. Therefore, it is very difficult to predict it only from the genome DNA sequence information by *in silico*. It is necessary to analyze sequences of a great number of mRNAs, possibly brought about *in vivo* by AS, and to identify them to elucidate a question of what type of protein-coding transcript is produced by AS of one gene.

Therefore we acquired approximately 1,460,000 of human full length cDNA (FLJ cDNA) from the CAP site of mRNA, which were constructed from approximately 100 kinds of human tissues and cells by the oligo-capping method, and analyzed all 5'-EST (expressed sequence tag), with average sequence lengths of approximately 500 bases [12-14]. We selected approximately 55,000 of FLJ cDNA, which acquired the information for the 5'-EST sequences, and performed full length sequencing [12-15]. Then we built a data set of human cDNA sequences, which included the FLJ cDNA, and performed mapping and clustering with the human genome sequence. The human genome sequence used 18 UCSC hg NCBI Build 36.1. Approximately 55,000 full length sequenced FLJ cDNA sequences and approximately 1,460,000 5'-EST sequences of FLJ cDNAs were used for analysis. We also used approximately 52,000 sequences of the human full length sequenced cDNAs registered with a public database (DB) at the same time, approximately 30,000 sequences of human RefSeq (NCBI Reference Sequences; <http://www.ncbi.nlm.nih.gov/RefSeq/>), approximately 48,000 sequences of human Ensembl sequence (Ensembl, human gene transcripts; <http://www.ensembl.org/index.html>) for analysis. Furthermore, we used the human EST sequence information registered with the public DB and analyzed it. As a result, we built the FLJ Human cDNA Database, <http://flj.lifesciencedb.jp> [15].

Using all sequences mapped to the same chromosomal region, we selected only reliable full length cDNA sequences for a protein coding gene prediction. Using the full-length cDNA selected above, we evaluated each genome locus with a manual for every chromosomal region one by one, whether it was an encoded protein or not. As a result, among 23,241 genes, which we predicted were on the human genome, the number of the genes, which we predicted to be a reliable gene, was 16,754. About the analysis result, including the identification of the number of the genes, we reported it to Wakamatsu, A. et al. [15].

We thought that it was necessary to acquire the expression profile about every transcription product, protein-coding transcript, which encoded human protein, for analyzing the influence that the change of these gene expression levels gave to a gene function. A pattern and an expression level of the mRNAs in each gene were fairly different. Therefore, for examining a gene function and the correlation of the expression level, we selected genes by the overall expression level on every gene region.

We analyzed the variety of the mRNA using approximately 1,460,000 5'-ESTs of FLJ cDNA, which we acquired, mainly until now. We were able to acquire considerable knowledge about the varieties of the transcription start site (TSS) and the amino acid sequence of the N-terminal side, because it was the data which arrested the TSS that manufactured it by the improved oligo-capping method precisely, and those 5'-ESTs were sequenced to have length of approximately 500 bases [12-17]. There is the first exon variation (FEV), which is one type of alternative N-terminus [Figure 2-A] caused by the variation of TSS [18]. We found the genes, which FEV showed the expression pattern that was tissue-specific, to a lot so far [15]. We thought that those results brought about a different protein-coding transcript using suitable TSS of each gene under a certain tissue condition and the situation, and there was the gene that might give a functional variation [19-21]. It is very important to elucidate a gene function by analyzing the correlation between such a function and the change of a splice pattern.

However, it was difficult to precisely predict AS, except for the N-terminal side, and those expression frequencies to have the characteristics that the ESTs of the FLJ cDNAs, which we used for analysis, which mainly specialized in 5'-sides of mRNAs until now. We still did not sufficiently analyze the variety of mRNA in the downstream region more than 500 bases from

the TSS. Therefore we analyzed AS, except for the N-terminal side, to elucidate the relationship between those changes and gene functions as described below. We thought that it was the suitable gene set, which we chose this time, for analyzing a change of AS in all CDS coding regions, because the gene set included a change of CDS as a result of classification in a change of N-terminal, internal and C-terminal sides of CDS. Using these genes, we decided to examine the relations between the change of the expression level of every AS and the gene function.

Then, we decided to exclude the gene, which was high in expression level of every region of it, from an analysis candidate. We thought the excluded gene, which had a high expression level constitutively expressed in the cell and/or quite frequently already understood the gene function by previous advanced analysis. At first we predicted the expression level about every gene locus from the total number of human cDNAs (human full length sequenced cDNAs in public DB, human RefSeq, human Ensembl) and human ESTs mapped with respect to the gene locus. We defined the total number of cDNAs and ESTs mapped with respect to the same gene locus as cluster size and made a histogram. Then we selected a gene for further analysis with a cluster size less than 500. By this selection, 14,382 genes were included in it, which occupied 85.8% of the 16,754 genes that we predicted to be reliable genes [Figure 1].

Using the number of FLJ 5'-ESTs mapped with respect to the same gene locus, we decided to sort the further analysis candidate genes. We acquired full length cDNAs from approximately 100 kinds of human tissues and cells to perform the functional analyses of the human gene. We constructed approximately 1,460,000 full-length cDNA resources said to be FLJ cDNA and performed a sequence analysis of 5'-ESTs, which averaged length was approximately 500bp. These resources and sequence analysis information would be a useful tool in a gene function analysis. Because most materials used for construction of FLJ cDNAs are commercially available for approximately 100 kinds of human tissues and cells, we were able to obtain RNA easily. Therefore we decided to sort the gene where we already had resources and information. Then we defined the number of 5'-ESTs derived from FLJ cDNAs mapped onto the human genome with respect to the same gene locus as the FLJ 5'-EST size and then made a histogram. Using this result, we decided to select genes with FLJ 5'-EST size more than five. By this selection, 9,907 genes were included in it, which occupied 59.1% of 16,754 genes that we predicted to be reliable [Figure 1].

We predicted the CDS of the protein-coding transcript produced from the same gene locus and classified the CDS according to the result. Then we sorted the classification result for the purpose. In the gene functional analysis, we thought that it was particularly important to analyze the variety of the protein-coding transcript rather than the variety of mRNA. Therefore, we excluded CDSs, which a variation by AS existed in the noncoding region (UTR) or was predicted to be completely matched in CDS, from an analysis object about the gene. As a result, the number of genes meeting the condition that their protein-coding transcripts produce CDS of different plural kinds, is 9,323 genes. These genes occupied 55.6% of 16,754 genes that we predicted to be reliable [Figure 1].

Furthermore, we sorted the genes based on the number of protein-coding transcripts produced from the same gene locus. In AS, the mature mRNA is produced by changing the part and the combination of exons of the immature mRNA. In this complicated process splicing eliminates an unnecessary part, then the different plural number class of mature mRNAs is produced. In other words the number of mRNA, which can be produced theoretically by AS, is enormous, but the kinds are different in every gene. We thought that it would be difficult to analyze exhaustive gene expression, corresponding to the combination of a too complicated

protein-coding transcript. Therefore we predicted complexity of AS of the genome region and sorted out the results from the number of protein-coding transcripts produced from the same gene locus. Then we selected the gene with the number of predicted protein-coding transcripts less than ten. The gene to meet was among 6,297 genes, and held a condition 37.6% of 16,754 genes that we predicted to be reliable [Figure 1].

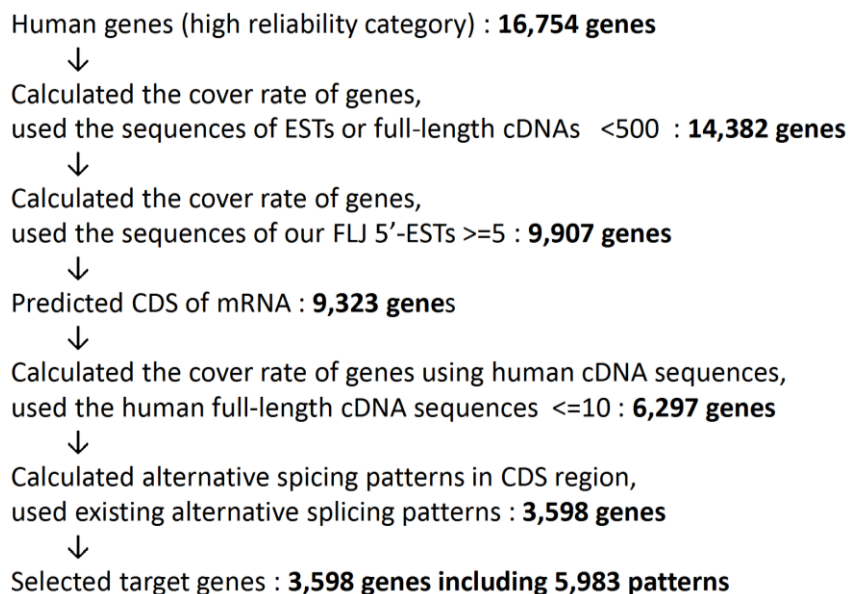


Figure 1. Procedure for the identification of candidate genes. Outline of our gene-characterization method from human full-length cDNAs and ESTs mapped to the human genome. EST, expressed sequence tag; CDS, coding region. “Human genes (high reliability category)”, 16,754 genes, were shown in Wakamatsu, A. et al. [15].

We obtained variation results produced in CDS by similarity searching for all protein-coding transcripts produced from the same gene locus using Align, similarity analysis software by Blastn and Blastp, and then manually evaluated the results as to what kind of CDS by AS was produced from that. After that each gene was classified by the splice pattern that we should have paid attention to. In this evaluation, low reliability patterns predicted in CDS were eliminated. These include patterns in which the immature mRNA and intron, was included. We also eliminated unreliable patterns caused by substitution and point mutation. By this evaluation, we selected 3,598 genes, which occupied 21.5% of 16,754 genes that we predicted to be reliably suitable for variety analysis of the mRNAs. In 3,598 genes, 5,983 patterns were found as a splice pattern, where we should have paid attention [Figure 1].

Then we classified the five kinds of splice patterns we found [Figure 2-A] by paying attention to the splice kind and the CDS region that thereby changed. As a result, the first pattern, “1: Alternative N-terminus”, was 1,408 patterns [Figure 2-B]. In this pattern, there is TSS as shown in Figure 2-A on exon, which is different from TSS of the known full-length cDNA, and is equivalent but alternative to the N-terminal side with the translation start point by the splicing patterns caused by different TSS. The fifth pattern, “5: Alternative C-terminus”, a C-terminal side is alternative and to change CDS, was 1,030 patterns [Figure. 2-B]. About another pattern, “2, 3 and 4: Alternative internal region”, that both N terminal and C-terminal

sides did not change but the internal region was alternative [Figure 2-A], had 3,545 patterns [Figure 2-B]. The pattern, “Alternative internal region”, was classified to three kinds [Figure 2-A], Type 1, 2 and 3, with different types of CDS regions. We classified that type 1 was an N terminal of CDS changed, type 3 was a C-terminus of CDS changed, and type 2 was a pattern except for type 1 and 3 [Figure 2-A]. As a result, we were able to classify in 677, 2,236, 632 patterns, respectively [Figure 2-B].

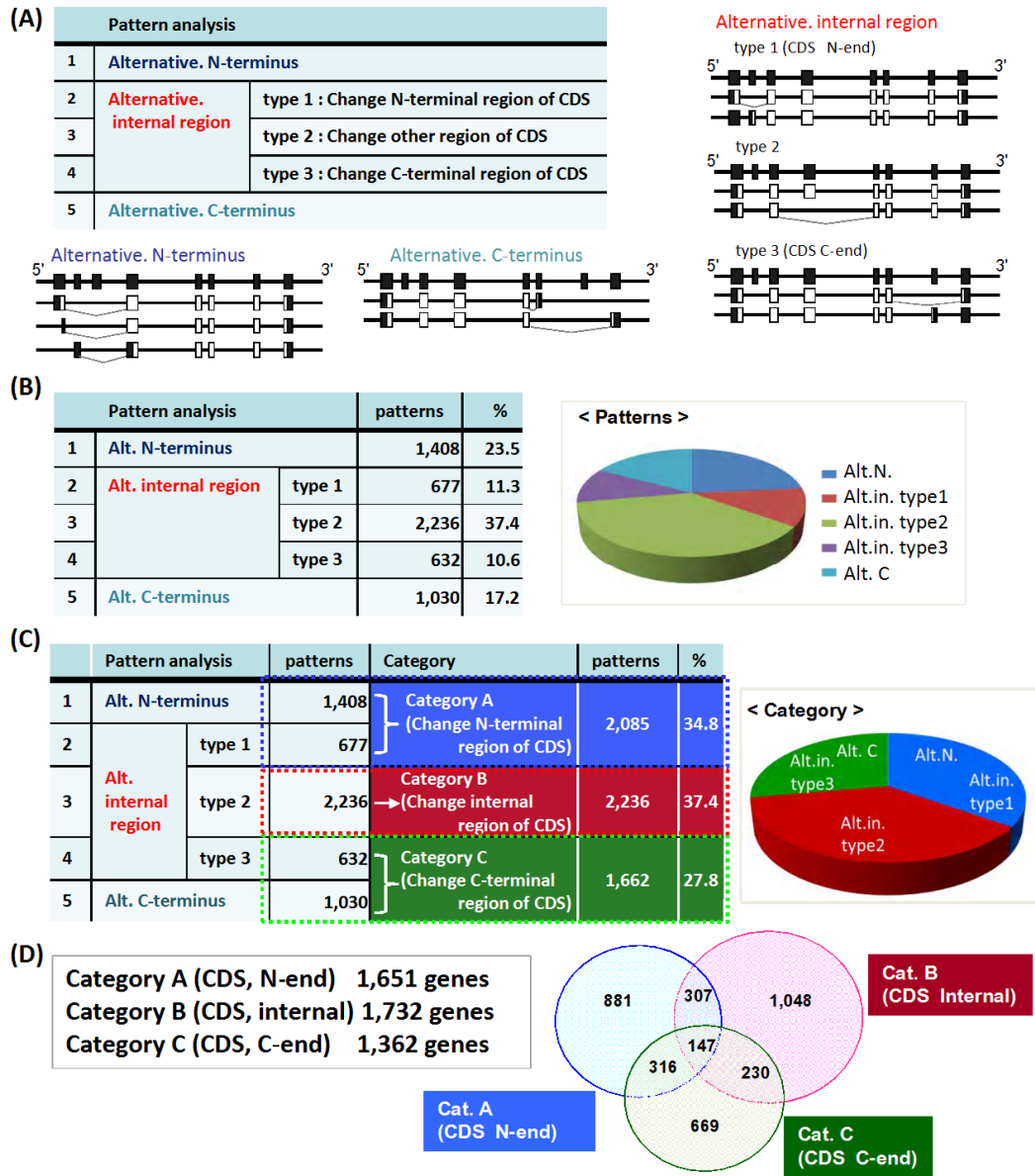


Figure 2. Classifications of selected 3,598 genes based on mRNA diversities. (A) Types of alternative splicing of protein-coding transcripts. Boxes, exons; black lines, introns; black boxes, coding regions; white boxes, untranslated regions. (B) Classifications of 5,983 patterns based on types of alternative splicing. (C) Categorization of 5,983 patterns by regions of alternative splicing. (D) Classifications of selected 3,598 genes by results of categorization.

From a gene, a multiple protein-coding transcript was produced by AS, but we thought that a protein-coding transcript, which was derived by stimulation and changed an expression level, was important to elucidate what region of the gene was often expressed by AS. For example, a possibility is when AS happens a lot to an N terminal end, that the gene is regulated by transcription. In addition, both ends do not change, but an expression level fluctuation is related to the presence of the gene functional domain, which is caused by having the exon or not in the internal region. In this case we think that the possibility that having the exon or not plays an important role for the change of the gene function. Therefore for analysis of AS data as to what kind of exon region of the mRNA was expressed and affected the change of the expression level, we decided to classify five kinds of patterns in three categories by paying attention to the region that had a change in CDS. Category A, change N-terminal region of CDS, which was the first category, combined two kinds of patterns, “1: Alternative N-terminus” and type 1 of “2: Alternative internal region” [Figure. 2-C]. The pattern classified in this group was 2,085 patterns [Figure. 2-C]. Category B, change internal region of CDS, which was the second category, consisted only of one kind of pattern, type 2 of “3: Alternative internal region” [Figure. 2-C]. The pattern classified in this group was 2,236 patterns [Figure. 2-C]. Category C, change C-terminal region of CDS, which was the third category, combined two kinds of patterns, type 3 of “4: Alternative internal region” and “5: Alternative C-terminus” [Figure. 2-C]. The pattern classified in this group was 1,662 patterns [Figure. 2-C].

Finally, with the three categories that we classified, we analyzed the AS pattern number to be about 3,598 genes to clarify what category deserved attention. Depending on a gene, plural AS happened, and the gene sorted in all three categories was exposed. As a result, Category A, “change N-terminal region of CDS; CDS, N-end”, category B, “change internal region of CDS; CDS, internal”, and category C, “change C-terminal region of CDS; CDS, C-end”, were calculated to be 1,651, 1,732 and 1,362 genes, respectively [Figure 2-C&D].

CONSTRUCTION OF THE CUSTOM DNA MICROARRAY FOR ANALYSIS OF MRNA VARIATION

We decided to perform a new construction of a custom DNA microarray for analyzing the expression pattern of every splice pattern, which could be analyzed to compare 5,983 splice patterns of 3,598 genes. We used a Roche NimbleGen multiplex array [4-plex format (4x72K)]. On this array, the probe length was 60 bases. Therefore, using 3,598 genes, we tried to identify a specific sequence region of every protein-coding transcript, which was necessary for a probe design. But a small amount of genes was not able to find the specific sequence region that was necessary for the design of the probe because the sequence length of the exon had a short. As a result, 3,413 genes were selected from 3,598 genes as the analysis candidates for exhaustive gene expression analysis of AS regions. To the 3,413 genes, we constructed the probes which assumed the specific target regions for the expression analysis for every AS pattern. Furthermore, we designed the probe to assume a common region shared in all protein-coding transcripts to be able to analyze an overall expression level of the gene. As a result, we identified 14,676 sites of specific sequence regions based on 8,672 cDNAs, which were selected as representative and alternative splicing variants from the consisted cDNAs of 3,413 genes, to analyze the expression frequency, and designed 48,388 probes in total in the regions

by basically using three kinds of probes for each site. We extracted the probe pair, which could compare AS patterns by the probes, which were specifically manufactured in AS patterns from there. As a result, from about 3,413 genes, we understood that we could compare the expression frequency of 5,784 AS patterns using probes of 5,784 pairs. Then we analyzed the AS pattern number of about 3,413 genes in the same way as the above-mentioned gene sets, 3,598 genes. As a result, Category A, “change N-terminal region of CDS; CDS, N-end”, category B, “change internal region of CDS; CDS, internal” and category C, “change C-terminal region of CDS; CDS, C-end”, were calculated to 1,520, 1,692 and 1,346 genes, respectively.

ANALYSIS OF ALTERNATIVE SPLICING USING NT2 CELLS INDUCED NEURONAL DIFFERENTIATION BY RETINOIC ACID

Using the custom DNA microarray, which we built, we decided to analyze the relations between a variety of the mRNA and the gene function. The genes were identified so that the expression levels were fluctuated by all-trans retinoic acid (RA) treatment using NT2 pluripotential human embryonal carcinoma cell, known as the NT2/D1 cell. This cell line is known so that neuronal differentiation is induced by treatment with RA [22]. Under the RA levels, which are able to cause neuronal differentiation, we understand that cell growth is not inhibited on this instance [22]. Furthermore, we are able to identify that all of them come from environmental factors, if the cultured cells change by RA induction and also by a variety of the mRNA changes [23-24]. Each stage of these cultured cells is one group with the same genetic background. AS is said to be important to the mammalian nervous system, when an important change occurs concerning neuronal differentiation such as cell differentiation, axon induction, neurite formation is regulated by AS [25].

We analyzed the change of the expression level for 35 days using NT2 cells after RA induction by the custom DNA microarray, and gathered analysis results of six time point samples, 0, 1, 2, 7, 14 and 35 days, which were listed below them sequentially with 0-day, 1-day, 2-day, 7-day, 14-day and 35-day, respectively [Figure 3-A]. Two kinds of independent experiments were performed. We prepared and repeated the schedule to subculture the samples of four kinds, 0-day, 1-day, 2-day and 7-day, as the first analysis (test 1) [Figure 3-B]. Two kinds of independent experiments were performed. We prepared and repeated the schedule to subculture the four kinds of samples, 0-day, 7-day, 14-day and 35-day, as the second analysis (test 2) [Figure 3-B]. From reproducible evaluation (results not shown) by the experiment, we prepared the analytical control to use a 0-day control sample of the first analysis (test1 0-day).

As a result, the expression level of a total of 2,171 sites of specific sequence regions, two of test2 0-day, 33 of test1 1-day, 85 of test1 2-day, 337 of test1 7-day, 357 of test2 7-day, 452 of test2 14-day and 905 of test2 35-day, satisfied our setting threshold 4-fold, compared with the control sample, test1 0-day, respectively. Then we merged those sites of specific sequence regions and as a result, 1,147 sites were picked up. In those we eliminated sites after evaluation then used another control sample, test2 0-day. Then we performed a cluster analysis for a selected 1,145 sites, as a result, the probes were clustered in order of RA induction time points, and two 0-day and two 7-day samples were each clustered with the same induction time points [Figure 3-C]. We evaluated genes of the selected 1,145 sites of specific sequence regions by analyzing the relationship between genes and probes designed on sites. As a result, we clarified

that they came from 583 genes. Because 2,513 sites of specific sequence regions, including sites on a common region, shared in all protein-coding transcripts were designed on those genes, as a result the expression level of 1,145 sites of specific sequence regions was changed by RA induction, but the remaining 1,368 sites were not changed by it.

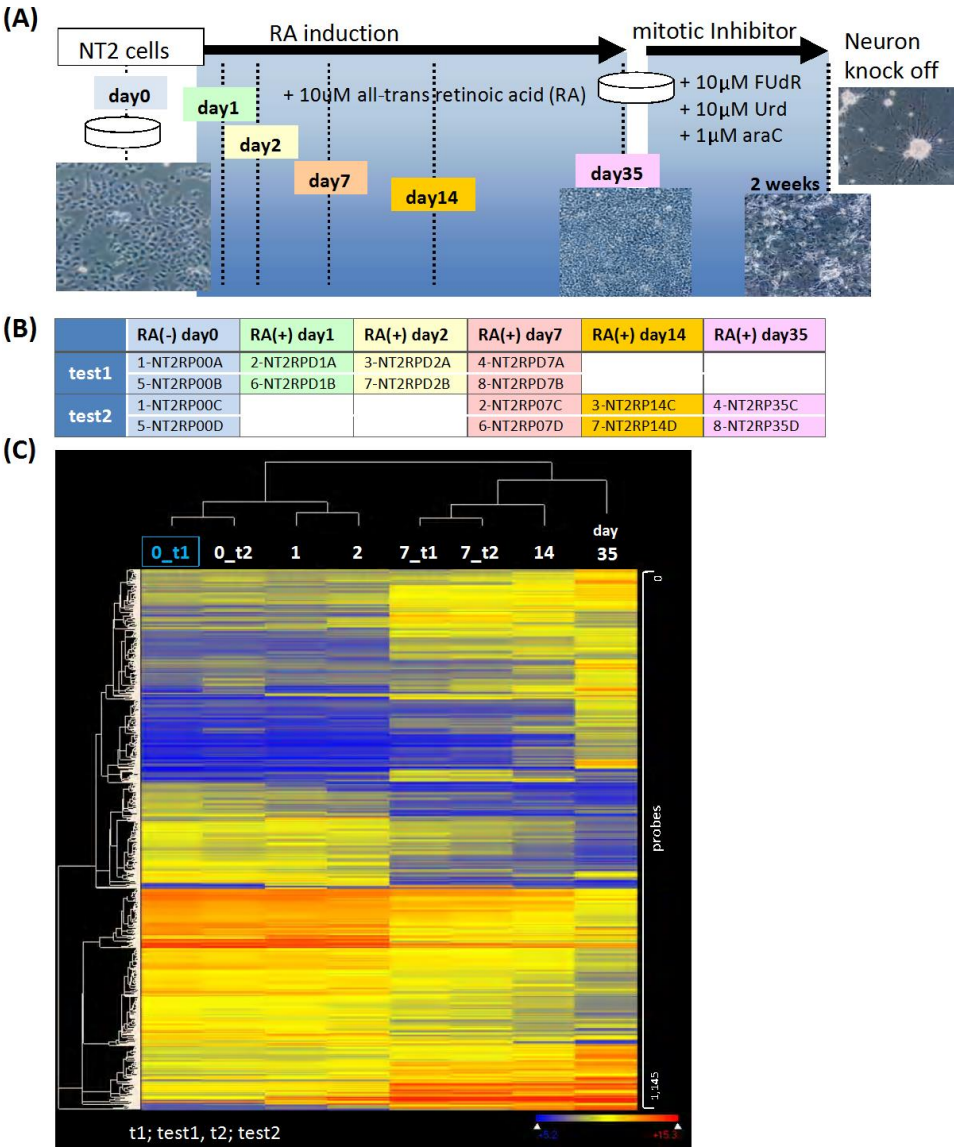


Figure 3. Comparison of gene expression profiles of control and RA-induced NT2 cells by DNA microarray analysis. (A) Method for RA induction for NT2 cells. In response to all-trans retinoic acid (RA), NT2 cells differentiate towards a neuronal lineage. FUDR, 5-fluoro-2'-deoxyuridine; Urd, uridine; araC, cytosine beta-D-arabinofuranoside. NT2 cells shared the same genetic background; analyze changes in the mRNA diversity of genes following changes in an environmental condition (RA induction). (B) List of total RNAs used in this DNA microarray study. (C) Gene expression profiles of RA-treated cell data after filtering out the control cell data (0_t1). Columns and rows indicate RA time point samples and probes, respectively. Probes and samples are aligned in the order defined by the results of the clustering analysis. The color bar represents the grades of the relative expression levels: increase, red; decrease, blue.

In the 583 genes, we searched pairs of sites of specific sequence regions where we could compare the AS pattern, as a result, 452 genes were extracted. In the 452 genes, 1,003 sites of specific sequence regions showed the expression level change by RA induction, and 1,051 sites did not show it. Regarding the extracted 452 genes, we could compare the expression frequency of 857 AS patterns using 857 pairs of sites of specific sequence regions. Then we paid attention to each gene comparison contrast pair and analyzed it to see whether AS changed an expression level. As a result, in 614 pairs of the sites, the expression level of the pair of both or either changed, but the expression level of both pairs did not change into 243 remaining pairs. An expression level changed in about 614 pairs. We performed a comparison analysis to see whether the expression level was due to timing, and judged whether the expression profile was the same or different. As a result, the profiles of the expression level were different in about 480 pairs of sites, but the profiles were the same in about 134 of the remaining pairs.

We selected about 25 genes from those. We designed the specific DNA primers which could recognize protein-coding transcripts for comparing the AS patterns like custom DNA microarrays individually, and analyzed the expression profiles by the RA induction by real-time PCR using NT2 cells. As a result, because the technique was different, some genes showed that the degree of the change of the expression level by RA induction was different, however, all genes showed that the expression profiles of them by that were agreeable. The results of five genes, HOXA3, CRISPLD1, HOXA1, CYP2S1, and MTA3, were shown in Figure 4. We described below the analysis result examples, analysis 1-5, respectively. The results of another 20 genes, CDH6, CYP26B1, DENND5B, DMKN, DYSF, FAM65B, FGD4, FNDC5, GSTO2, HNF1B, LMCD1, LSAMP, MAPKAPK2, NCAM2, NEFM, RBP1, SEMA3C, SKAP2, ST3GAL5, and TFEC, were not shown here.

Analysis 1, HOXA3: The HOXA3 gene is one of the homeobox genes and it is said that it is a DNA-binding transcription factor [26-27]. The HOXA3 gene produced a protein-coding transcript having two kinds of different CDSs, NM_030661.3 and NM_153632.1. The two kinds of transcription products shared a homeobox domain, but were able to classify in Category A (CDS, N-end) because N-terminal regions are different [Figure 4-A]. We got a result of expression profiles by both real-time PCR and the custom DNA microarray by RA induction that an expression level was changed for two kinds of transcription products, and the rate of climb of two kinds of expression levels were different [Figure 4-A].

Analysis 2, CRISPLD1: CRISPLD1, cysteine-rich secretory protein LCCL domain containing 1, plays a role in NSCLP, nonsyndromic cleft lip plus palate, through the interaction with CRISPLD2 and folate pathway genes [28]. The CRISPLD1 gene produced protein-coding transcripts having two kinds of different CDSs, FLJ57290 and NM_031461.3. We classified this case in Category A (CDS, N-end) because two kinds of transcription products choose different TSS among different exons said to be FEV [Figure 4-B]. We got a result that an expression profile was as different as for an expression level changing as for two kinds of transcription products by both real-time PCR and the custom DNA microarray by RA induction, and the two kinds of expression profiles were different [Figure 4-B].

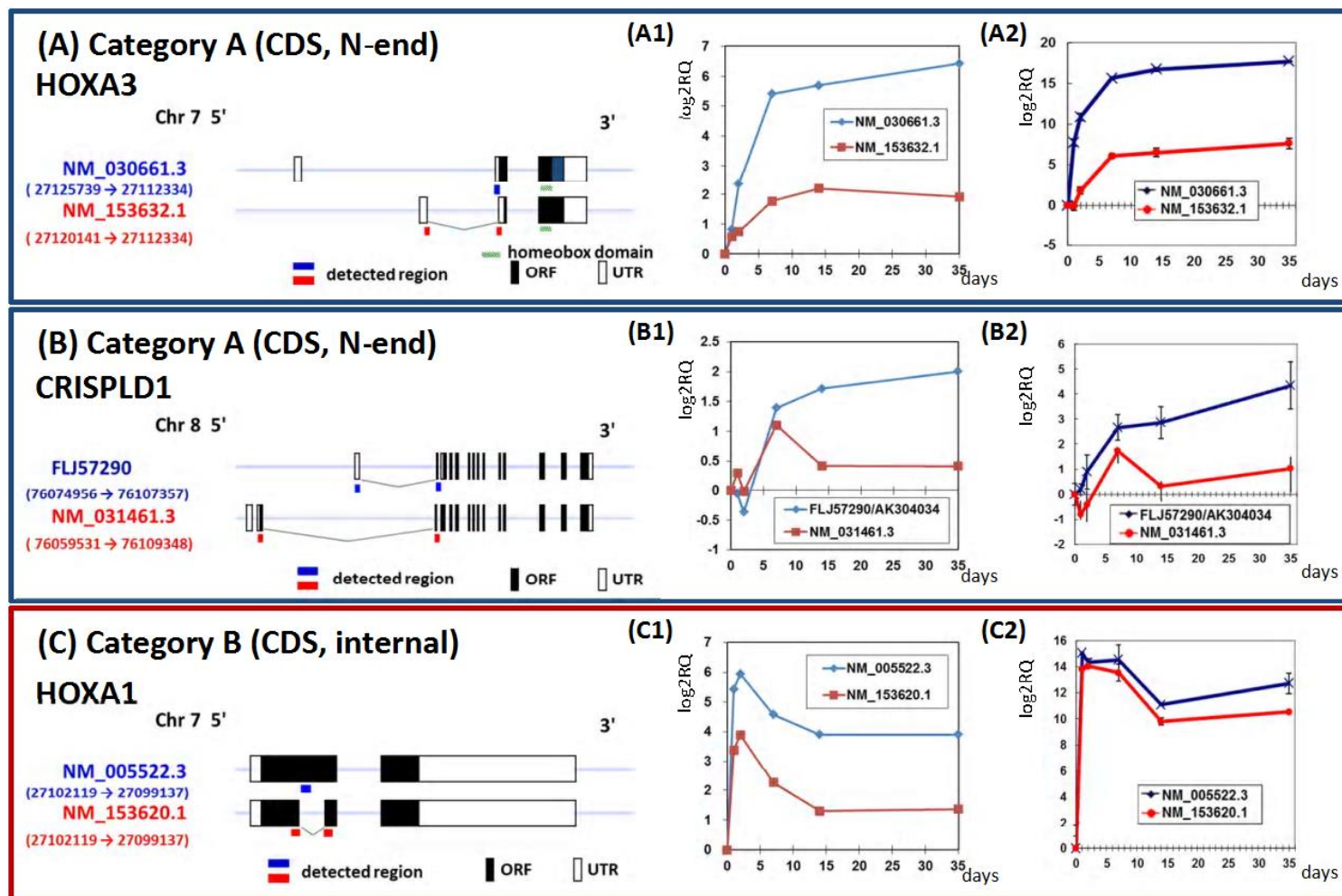


Figure 4. (Continued).

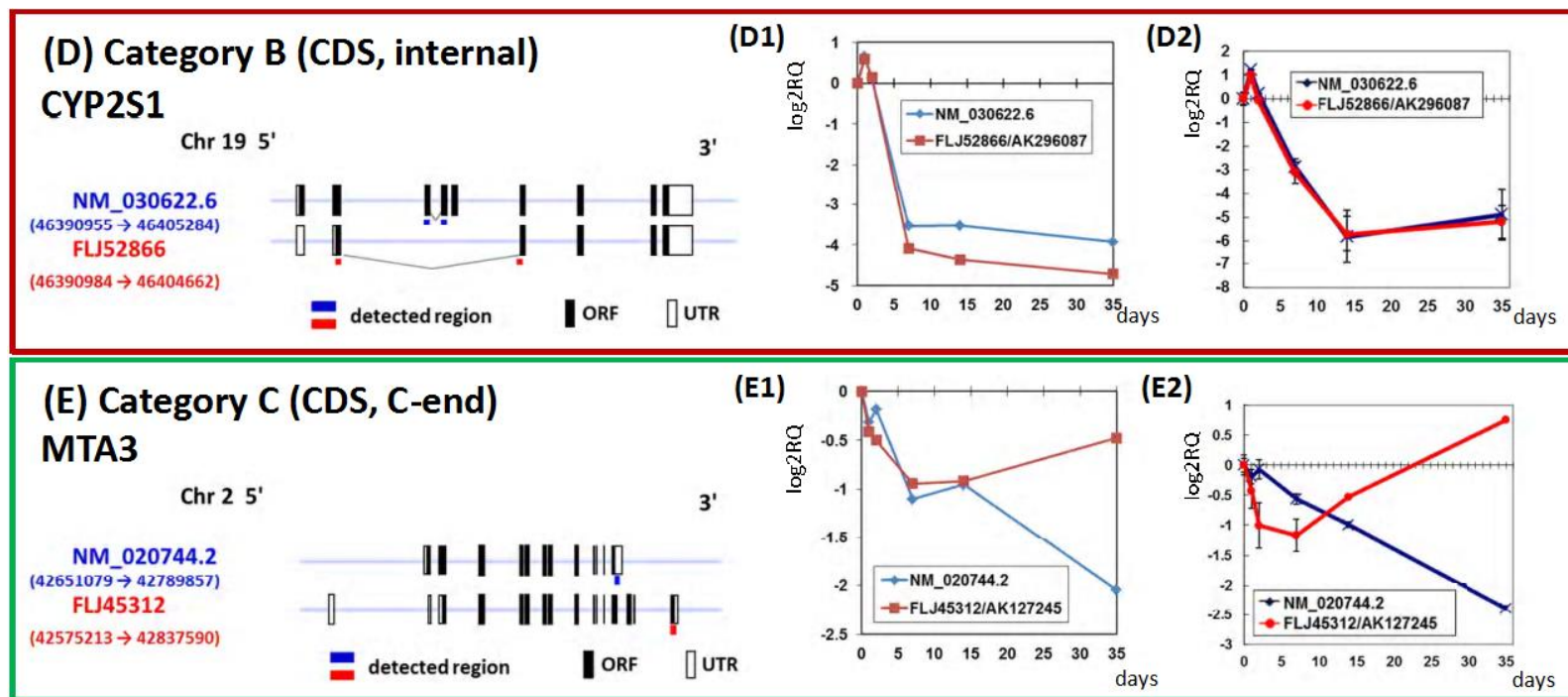


Figure 4. Quantitative analysis of expression of identified five RA-responsive genes by real-time PCR. Expression levels of two transcripts produced by AS from each gene were analyzed by DNA microarray (A1, B1, C1, D1 and E1) or by real-time PCR (A2, B2, C2, D2 and E2), and were represented in log₂ base. Name of the RA-responsive genes: (A) HOXA3, (B) CRISPLD1, (C) HOXA1, (D) CYP2S1, (E) MTA3. Boxes, exons; purple lines, introns; black boxes, coding regions; white boxes, untranslated regions; red or blue bars, detected regions; and numbers given in parentheses, genomic alignment positions. The real-time PCR data were normalized with respect to that of the human GAPDH.

Analysis 3, HOXA1: HOXA1 gene is a DNA-binding transcription factor in one of the homeobox genes [29-30]. The HOXA1 gene produced protein-coding transcripts having two kinds of different CDSs, NM_005522.3 and NM_153620.1. As for two kinds of transcription products, only NM_005522.3 was shown to have a homeobox domain by the difference of the splice pattern. We classified this case in Category B (CDS, internal) because the internal regions were different [Figure 4-C]. We got a result that an expression profile was as different as for an expression level changing as for two kinds of transcription products by both real-time PCR and the custom DNA microarray by RA induction, and the two kinds of expression profiles were different [Figure 4-C]. But the degree of the change was slightly different by experiment technique.

Analysis 4, CYP2S1: CYP2S1 encodes a member of the cytochrome P450 superfamily of enzymes [31]. The cytochrome P450 proteins are monooxygenases which catalyze many reactions involved in drug metabolism and synthesis of cholesterol, steroids and other lipids. This protein localizes to the endoplasmic reticulum [31]. The CYP2S1 gene produced protein-coding transcripts having two kinds of different CDSs, FLJ52866 and NM_030622.6. We classified this case in Category B (CDS, internal) because internal regions were different [Figure 4-D]. We got a result that an expression profile was as different as for an expression level changing as for two kinds of transcription products by both real-time PCR and the custom DNA microarray by RA induction, but the two kinds of expression profiles were the same [Figure 4-D].

Analysis 5, MTA3: Metastasis-associated protein (MTA3) is a constituent of the Mi-2/nucleosome remodeling and deacetylase (NuRD) protein complex that regulates gene expression by altering the chromatin structure and is able to facilitate cohesion loading onto DNA. The biological function of MTA3 within the NuRD complex is unknown [32]. The MTA3 gene produced protein-coding transcripts having two kinds of different CDSs, FLJ45312 and NM_020744.2. We classified this case in Category C (CDS, C-end) because the C-terminal regions were different [Figure 4-E]. We got a result that an expression profile was as different as for an expression level changing as for two kinds of transcription products by both real-time PCR and the custom DNA microarray by RA induction, and the two kinds of expression profiles were different [Figure 4-E].

Regarding the genes we selected in this way, we understood that we could cover data about the mRNA variations of genes by AS using the custom DNA microarray since probes were designed at sites of specific sequence regions. If data about mRNA variations of genes by AS could be accumulated using this approach, we thought that we would be able to elucidate the mechanism that an expression level of multiple mRNAs, produced from the same gene, was changed by response to various factors and then the gene functioned depending on the response to those. We also thought that it was important for new function elucidation.

ANALYSIS OF RELATIONSHIP BETWEEN mRNA VARIATION BY ALTERNATIVE SPLICING AND THE GENE FUNCTION

Finally, for analyzing the relationship between the variety of mRNA and the gene function, about 452 genes responsive to RA, which were identified as having expression level changed

protein-coding transcripts by the expression analysis, were analyzed as to what kind of region was changed by AS. 857 AS patterns were built in 452 genes responsive to RA [Figure 5-A].

(A)

Category	Patterns (3,598 genes)	Patterns (452 RA induced genes)	Expression level by RA induction		Expression profile by RA induction	
A (CDS, N-end)	2,085	273 (13.1%)	change	155 (7.4 %)	differ	134 (6.4%)
			not change	118 (5.7 %)	same	21 (1.0%)
B (CDS, internal)	2,236	339 (15.1%)	change	259 (11.6%)	differ	178 (8.0%)
			not change	80 (3.6%)	same	81 (3.6%)
C (CDS, C-end)	1,662	245 (14.7%)	change	200 (12.0%)	differ	168 (10.1%)
			not change	45 (2.7%)	same	32 (1.9%)
total	5,983	857 (14.3%)				

(B)

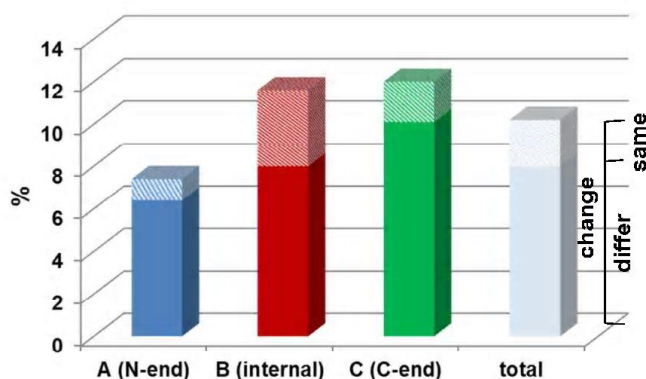


Figure 5. Quantitative evaluation of expression of selected 3,598 genes by DNA microarray analysis using NT2 cells. Evaluation of expression profile of 452 RA induced genes by DNA microarray analysis.

Then we analyzed the AS patterns of 452 genes responsive to RA induction according to a category. AS pattern number of Category A (CDS, N-end), Category B (CDS, internal) and Category C (CDS, C-end) was 273 (13.1%), 339 (15.1%) and 245 (14.7%), respectively [Figure 5-A]. From this result, the distribution of 857 AS patterns of 452 responsive genes selected by the expression change by RA induction did not differ much with that of 5,983 patterns of 3,598 genes for the analysis candidates [Figure 5-A].

After that analysis, we divided the AS patterns into those that did not show a change and those that showed an expression level change by RA induction [Figure 5-A]. As a result, in Category A (CDS, N-end) where the N-terminal side of CDS region changed, the AS patterns that the expression levels changed were 155 (7.4%) [Figure 5-A]. Then we compared the expression profiles of the 155 AS patterns where the expression levels changed by RA induction; as a result 134 (6.4%) showed different profiles [Figure 5-A]. In Category B (CDS, internal) where the internal region of CDS region changed except for the N-terminal and C-

terminal regions, the AS patterns that the expression levels changed were 259 (11.6%) [Figure 5-B]. Then we compared the expression profiles of the 259 AS patterns where the expression levels were changed by RA induction, and as a result 178 (8.0%) showed different profiles [Figure 5-B]. In Category C (CDS, C-end) where the C-terminal side of CDS region changed, the AS patterns that the expression levels changed were 200 (12.0%) [Figure 5-A]. Then we compared the expression profiles of the 200 AS patterns where the expression levels were changed by RA induction, and as a result 168 (10.1%) showed different profiles [Figure 5-B]. By the distribution of the AS patterns of 614 pairs where the expression levels changed by RA induction from this result, the rate of AS patterns classified in the category of Category A (CDS, N-end) was fewer than that of another two categories [Figure 5-B]. In the evaluation of every gene, there were fewer genes distributed over Category A (CDS, N-end) with the genes that the expression levels changed than the other two categories [Figure 5-A].

From our previous studies regarding AS in the N-terminal region [15, 18-21, 23, 24], we thought that there was only a strong correlation between the mRNA variation of the N-terminal region of protein-coding transcript by AS and the gene function according to the gene expression pattern. However, from the results described above, a lot of mRNA variation by AS in both the internal and C-terminal regions of protein-coding transcripts also existed, and some of those were correlated with an expression level change, which sometimes induced a functional alteration of the gene, such as by RA induction [Figure 5-B]. From these results, we would have to use human full-length cDNA resources and the sequence analysis information, which we obtained during previous research, for further sequence analysis of both the internal and C-terminal regions of protein-coding transcripts using next generation DNA sequencing system. Further analysis is necessary regarding AS for more detailed gene function analysis using full-length cDNAs.

CONCLUSION

One of the most important prerequisites for understanding the functions of genes is to analyze mRNA diversity. At least 70% of all human genes produce multiple transcripts by AS [3, 8]. Genes could produce multiple protein-coding transcripts by AS for generating biological and functional diversity.

We sequenced approximately 55 thousand human full-length cDNAs and approximately 1.45 million 5'-ESTs from human cDNA libraries constructed by the oligo-capping method for gene functional analyses [12-15]. The studies indicated that a single gene could utilize AS for tissue-specific transcription [15]. We also constructed the FLJ Human cDNA Database [15], <http://flj.lifesciencedb.jp>, which would be useful in analyzing the diversity of protein-coding transcripts. Furthermore, we found a close relationship between the predicted function of a gene and its tissue-specific expression, implying that the gene was transcribed into an appropriate transcript according to the needs and circumstances [15]. However, we were unable to distinguish whether the observed expression pattern was associated with the genetic or the environmental risk factors, primarily because the tissues and cell lines used in our analyses were not derived from one person.

Therefore, we used human NT2 cells for analyzing mRNA diversity [23-24]. This cell line is a good model for the embryonic development as well as for tumor cell differentiation. In

response to RA, NT2 cells differentiate towards a neuronal lineage with associated loss of cell growth and tumorigenicity [22]. Because these cells shared the same genetic risk factor, we were able to analyze changes in the mRNA diversity of genes as a result of RA induction, an environmental risk factor [23-24]. We believe the mRNA diversity and gene function are correlated [23-24, 33], and therefore, it is imperative to understand this relationship in order to identify the genes that are specifically involved in neurodifferentiation. This point of view could be expanded to elucidate other phenomena.

Accumulating more data on the mRNA diversity of genes using approaches similar to what we have described above will not only reveal the underlying mechanism by which genes produce specific protein coding transcripts in response to various factors, but will also contribute to the discovery of new gene functions. Moreover, identification of the risk factors responsible for the mRNA diversity would be an important step for gene function analyses and identifying candidate genes as novel targets for the development of new drugs and therapies.

REFERENCES

- [1] International Human Genome Sequencing Consortium. (2004). Finishing the euchromatic sequence of the human genome. *Nature*, 431, 931-45.
- [2] Black, D. L. (2003). Mechanisms of alternative pre-messenger RNA splicing. *Annu. Rev. Biochem.*, 72, 291-336.
- [3] Lander, E. S. et al. (2001). Initial sequencing and analysis of the human genome. *Nature*, 409, 860-921.
- [4] Lopez, A. J. (1998). Alternative splicing of pre-mRNA: developmental consequences and mechanisms of regulation. *Annu. Rev. Genet.*, 32, 279-305.
- [5] Black, D. L. (2000). Protein diversity from alternative splicing: a challenge for bioinformatics and post-genome biology. *Cell*, 103, 367-70.
- [6] Modrek, B. & Lee, C. (2002). A genomic view of alternative splicing. *Nat. Genet.*, 30, 13-9.
- [7] Stamm, S. (2002). Signals and their transduction pathways regulating alternative splicing: a new dimension of the human genome. *Hum. Mol. Genet.*, 11, 2409-16.
- [8] Wang, E. T. et al. (2008). Alternative isoform regulation in human tissue transcriptomes. *Nature*, 456, 470-6.
- [9] Maniatis, T. & Tasic, B. (2002). Alternative pre-mRNA splicing and proteome expansion in metazoans. *Nature*, 418, 236-43.
- [10] Monani, U. R. et al. (1999). A single nucleotide difference that alters splicing patterns distinguishes the SMA gene SMN1 from the copy gene SMN2. *Hum. Mol. Genet.*, 8, 1177-83.
- [11] Bracco, L. & Kearsey, J. (2003). The relevance of alternative RNA splicing to pharmacogenomics. *Trends Biotechnol.*, 21, 346-53.
- [12] Ota, T. et al. (2004). Complete sequencing and characterization of 21,243 full-length human cDNAs. *Nat. Genet.*, 36, 40-5.
- [13] Otsuki, T. et al. (2005). Signal sequence and keyword trap in silico for selection of full-length human cDNAs encoding secretion or membrane proteins from oligo-capped cDNA libraries. *DNA Res.*, 12, 117-26.

- [14] Kimura, K. et al. (2006). Diversification of transcriptional modulation: large-scale identification and characterization of putative alternative promoters of human genes. *Genome Res.*, 16, 55-65.
- [15] Wakamatsu, A. et al. (2009). Identification and functional analyses of 11,769 full-length human cDNAs focused on alternative splicing. *DNA Res.*, 16, 371-383.
- [16] Suzuki, Y. & Sugano, S. (2003). Construction of a full-length enriched and a 5'-end enriched cDNA library using the oligo-capping method. *Methods Mol. Biol.*, 221, 73-91.
- [17] Suzuki, Y. et al. (1997). Construction and characterization of a full length-enriched and a 5'-end-enriched cDNA library. *Gene*, 200, 149-56.
- [18] Zhang, T. et al. (2004). Multiple variable first exons: a mechanism for cell- and tissue-specific gene regulation. *Genome Res.*, 14, 79-89.
- [19] Landry, J. R. et al. (2003). Complex controls: the role of alternative promoters in mammalian genomes. *Trends Genet.*, 19, 640-8.
- [20] Wu, Q. & Maniatis, T. (2000). Large exons encoding multiple ectodomains are a characteristic feature of protocadherin genes. *Proc. Natl. Acad. Sci. USA*, 97, 3124-9.
- [21] Strassburg, C. P. et al. (1997). Differential expression of the UGT1A locus in human liver, biliary, and gastric tissue: identification of UGT1A7 and UGT1A10 transcripts in extrahepatic tissue. *Mol. Pharmacol.*, 52, 212-20.
- [22] Andrews, P. W. (1984). Retinoic acid induces neuronal differentiation of a cloned human embryonal carcinoma cell line in vitro. *Dev. Biol.*, 103, 285-93.
- Wakamatsu, A. et al. (2010). Alternative splicing of genes during neuronal differentiation of NT2 pluripotent human embryonal carcinoma cells. *FEBS Lett.*, 584, 4041-4047.
- [23] Isogai, T. & Wakamatsu, A. (2012). mRNA diversity of genes by alternative splicing during neuronal differentiation of NT2 pluripotent human embryonal carcinoma cells by all-trans retinoic acid. Editors: Cheng, L-H. & Ito, Y. "*Retinoic Acid: Structure, Mechanisms and Roles in Disease*". 53-86, Nova Science Publishers, Inc. Hauppauge, NY, USA.
- [24] Li, Q. et al. (2007). Neuronal regulation of alternative pre-mRNA splicing. *Nat. Rev. Neurosci.*, 8, 819-31.
- [25] Greer, J. M. et al. (2000). Maintenance of functional equivalence during paralogous Hox gene evolution. *Nature*, 403, 661-5.
- [26] Mace, K. A. et al. (2005). HOXA3 induces cell migration in endothelial and epithelial cells promoting angiogenesis and wound repair. *J. Cell. Sci.*, 118, 2567-77.
- [27] Chiquet, B. T., et al. (2011). Nonsyndromic cleft lip and palate: CRISPLD genes and the folate gene pathway connection. *Birth Defects Res A Clin Mol Teratol.*, 91, 44-9.
- [28] Fernandez, C. C. & Gudas, L. J. (2009). The truncated Hoxa1 protein interacts with Hoxa1 and Pbx1 in stem cells. *J. Cell. Biochem.*, 106, 427-443.
- [29] Chariot, A. et al. (1995). Retinoic acid induces three newly cloned HOXA1 transcripts in MCF7 breast cancer cells. *Biochem. Biophys. Res. Commun.*, 215, 713-720.
- [30] Rylander, T. et al. (2001). Identification and tissue distribution of the novel human cytochrome P450 2S1 (CYP2S1). *Biochem Biophys Res Commun.*, 281, 529-35.
- [31] Kwintkiewicz, J. et al. (2012). Metastasis-associated protein 3 (MTA3) regulates G2/M progression in proliferating mouse granulosa cells. *Biol Reprod.*, 86, 1-8.
- [32] Sato, N. et al. (2001). Increased production of beta-amyloid and vulnerability to endoplasmic reticulum stress by an aberrant spliced form of presenilin 2. *J. Biol. Chem.*, 276, 2108-2114.

Chapter 26

ALTERNATIVE SPLICING IN HUMAN IMMUNE SYSTEM AND AUTOIMMUNE DISEASES

Xiang Guo

Translational Sciences, MedImmune, LLC, Gaithersburg, MD, US

ABSTRACT

Alternative splicing of pre-messenger RNA is nearly universal, involving more than 90% of human genes. It's an essential step in gene expression and responsible for much of the proteome diversity in mammalian genomes. The immune system requires a great diversity of functional proteins and immune cells need to respond to various foreign invasions rapidly. Alternative splicing provides one more layer of regulation that is essential for the function of human immune system. Many immune genes have been found to undergo alternative splicing, which plays an important role in the regulation of immune cell activation and function. Dysregulated splicing has been shown to be involved in various immune disorders, such as systemic lupus erythematosus (SLE) and rheumatic arthritis (RA). It may be a direct cause of the disease, or a modifier of disease susceptibility and severity. Further understanding of how alternative splicing may be used as a general mechanism in immune response is essential for our research in the pathophysiology of autoimmune diseases and development of new therapeutics for those diseases. This chapter provides an updated review about alternative splicing in human immune system as well as the relationship between dysregulated splicing and autoimmune diseases, particularly SLE and RA.

The immune system is a complex network of leukocytes, tissues, and organs that was designed to defend the body against foreign invasions. A sufficient diversity and flexibility is required for a functional immune system to ensure rapid adaption and response to changing environment in a quick and precise way. The great diversity of immune system can partly be achieved by gene regulation. While there have been many reports showing the crucial role of transcriptional regulation in the immune system, much less is known about another mechanism of gene regulation, namely alternative splicing.

Alternative splicing of per-messenger RNA is nearly universal, involving more than 90% of human genes [1]. It's an essential step in gene expression and responsible for much of the proteome diversity in mammalian genomes. Distinct mRNA transcripts can be generated from a single gene locus by including or excluding exons during post-transcriptional processing. Then, variant transcripts may produce proteins with different functional domains in different cell types and cell states. Alternative splicing can also regulate gene expression levels by eliminating microRNA binding site or through nonsense mediated decay.

Dysregulated splicing has been demonstrated in various diseases including autoimmune disorders [2]. It may be a direct cause of the disease or a modifier of disease susceptibility and severity. Investigation of the roles alternative splicing may play in autoimmune diseases is essential for the understanding of disease pathophysiology and development of new medicine for those diseases. This chapter summarizes the current knowledge about splicing in human immune system and autoimmune diseases.

ALTERNATIVE SPLICING AND IMMUNE CELL FUNCTION

Alternative splicing is regulated in a cell specific manner to control the immune response. It helps to shape the course and outcome of lymphocyte activation, which adds another layer of complexity for innate and adaptive cellular immunity. Multiple genome-wide analyses have demonstrated the crucial role of alternative splicing in modulating the activation threshold and maintaining homeostasis of lymphocytes, particularly T cells [3-7].

T cells respond to antigen challenge and express a plethora of proteins to initiate a broad network of signaling cascade, which then results in changes in cell migration, cytokine secretion, and other effector functions. Both transcriptional and splicing regulations have been shown to influence T-cell biology, but they appear to do so through distinct functional pathways. In a recent RNA-Seq study using PMA stimulation of human T-cell line to mimic mature TCR signaling, a similar percentage of genes exhibited splicing changes as transcriptional changes but no significant overlap was observed between two sets of genes. Those signal-responsive splicing events are involved in mediating critical T-cell effector functions, such as cell signaling, development, and trafficking. The vast majority of these genes are also regulated in a similar pattern in response to antigen challenge of primary CD4⁺ or CD8⁺ T cells [3].

The short-term changes that happen after TCR activation involves alternatively spliced genes that encode immediate T-cell effector functions. Intracellular protein kinases (PTKs) are first activated after antigen recognition by the TCR. Two different isoforms of the SRC-family PTK FYN have different efficiency at mediating TCR-induced calcium mobilization [8]. Virally infected T cells show increased expression of the low efficiency isoform, FYNB, which may be beneficial to the virus by decreasing T cell activation [9]. Cell-surface adhesion molecules are also crucial to early events in T-cell activation. CD44 is a cell surface glycoprotein involved in lymphocyte activation and homing. It has 10 variable cassette exons, which can produce a large number of CD44 variant forms. The smallest isoform without all variable exons are mainly expressed in naïve T cells, while multiple longer isoforms are expressed in activated T cells [10, 11]. Antibodies specific for those longer CD44 variants can

block *in vivo* activation of T cells, which demonstrates the crucial role of CD44 splicing for T-cell function [10].

T cell activation requires both TCR stimulation and CD28 co-stimulation. While TCR stimulation has been shown to cause both transcriptional and splicing changes of a large number of genes, very few genes demonstrate CD28 specific changes in gene transcription. In contrast, CD28 co-stimulation leads to profound changes in alternative splicing in naïve T cell upon their activation, with nearly 8 times as many genes being regulated through alternative splicing as gene transcription [4].

Once a T cell has carried out its required effector functions, prolonged T cell activation has to be blocked to prevent hyper-proliferation or hyperactivity of the immune response. Many genes that only display alternative splicing following prolonged activation encode proteins important for homeostasis or development of immunologic memory. CD45 is a transmembrane protein tyrosine phosphatase that is essential for TCR signaling. The smaller isoform CD45R0 shows weaker response to TCR ligation than the larger isoform CD45 RA. The change from CD45RA to CD45R0 after T-cell activation functions as a feedback mechanism to reduce prolonged TCR signaling [12]. Cytotoxic T-lymphocyte antigen 4 (CTLA-4) is a protein receptor that inhibits the CD28-dependent co-stimulatory signal. The stimulation of T cells induces the inclusion of a cassette exon that encodes the transmembrane domain of CTLA-4, so the expressed CTLA-4 protein is membrane-bound to prevent hyperstimulation of T cells [13]. Other splicing events involved in late activation of T cells include heterogeneous nuclear ribonucleoprotein D0 (AUF1) and lymphoid enhancer-binding factor 1 (LEF1) [14].

Regulatory T cells (Treg) can suppress proliferation and effector function of CD4⁺ T helper cells. They are critical for the negative regulation of immune response and the prevention of autoimmunity. Multiple differentially expressed splice variants have been detected between resting and activated cells, such as interleukin receptor 1 associated kinase (IRAK1), SAM68, and IL2RA [6]. The truncated isoform IRAK1c lacks kinase activity and it functions as a dominant negative to suppress NF- κ B activation [15]. The specific expression of IRAK1c in activated Treg cells may regulate TCR signaling and the subsequent inflammatory response.

Splicing regulation is also evident in the function of other immune cells. Significant splicing changes have been detected in CD19⁺ B cells activated by anti-CD40/IL2/IL10, although the total number of alternatively spliced genes is less than that detected in CD2⁺ T cells activated by anti-CD3/CD28 antibodies [5]. Human neutrophils express several splice variants of functional genes such as Fc α receptor [16], caspase-8 [17], glucocorticoid receptor [18], phospholipase D [19], glutaredoxin 1 [20], GM-CSF receptor α [21], and pantetheinase family proteins [22]. Adenosine is an endogenous inhibitor of neutrophil functions via binding to four subtypes of adenosine receptors: A1, A_{2A}, A_{2B}, and A3. It also delays neutrophil apoptosis, which may be involved in rheumatoid arthritis (RA) pathophysiology [23]. The A_{2A} receptor mRNA is up-regulated upon lipopolysaccharide (LPS) stimulation of human neutrophil, in which the differential splicing of 5'-UTR plays an important regulatory role. A_{2A} receptor transcripts with long 5'-UTRs are mainly expressed in resting neutrophil, whereas transcripts with short 5'-UTRs predominate in stimulated neutrophils which displayed higher translation efficiency than the longer variants [24]. Dendritic cells (DCs) are critical initiators of innate immunity and play a key role in controlling the magnitude and quality of adaptive immune responses. A recent study has demonstrated widespread alternative splicing events and splicing factor transcriptional signatures induced by an *E. coli* challenge to human DCs. Novel

isoforms have been validated for interferon regulatory factor 4 (IRF4), cyclin-dependent kinase inhibitor 3 (CDKN3), and 6 other genes in DCs following exposure to a pathogen [25].

ALTERNATIVE SPLICING AND CYTOKINE RESPONSE

Immune cells communicate with one another by releasing and responding to chemical messengers called cytokines. Splice variants of many cytokines have been shown to be functional antagonists of their corresponding wild-type cytokines. Cellular location may also be different for splice variants of the same cytokine, which can be soluble or membrane bound. Alternative splicing of cytokines, cytokine receptors, and signal transduction molecules has been shown to be an important regulatory mechanism for the human immune system.

IL-2, IL-4, IL-7, IL-9, IL-15, and IL-21 are a family of structurally related cytokines whose type I cytokine receptors share the identical IL-2 receptor γ chain (CD132). These γ c-dependent cytokines are involved in survival, proliferation and effector function of activated T cells. Two IL-2 variants that omit either exon 2 or exon 3 display an antagonistic activity and inhibit wild-type IL-2 binding to IL-2 receptors, which may be involved in acute rejection of transplantation [26, 27]. The IL-4 variant without the second exon (IL-4 δ 2) also exhibits an antagonistic activity to the wild type IL-4. Increased level of IL-4 δ 2 has been found in patients with systemic sclerosis and atopic asthma, which may be used as a diagnostic marker for these diseases [28, 29]. The α subunit of IL-2 receptor (CD25) may undergo dysregulated splicing in adult T-cell leukemia (ATL) patients to generate a short peptide without the conventional transmembrane domain [30]. The soluble isoform of IL-4 receptor α is coded by exons 3-8 that lacks the exons for the trans-membrane and intracellular regions. Sequence variations around exon 8 seem to be responsible for the selective splicing of IL-4 receptor α in allergic asthma patients [31]. Splice variants of IL-7, IL-7 receptor, IL-9 receptor, IL-15, IL-15 receptor, and IL-21 have all been identified and their functional relevance in human immune system remains to be further elucidated [32-37].

IL-6 has diverse functions in the immune system and it is implicated in a wide variety of inflammation-associated disease states. The splice variant of IL-6 without exon 4 acts as a dominant negative form of full-length IL-6. It seems to compete with native IL-6 for the interaction of IL-6R α (CD126), thus block IL-6R β (CD130) mediated signaling [38]. Another IL-6 variant produced by renal cell carcinoma also displayed IL-6 antagonist properties [39]. In addition, alternative splicing gives rise to soluble form of IL-6 receptor, which has been shown to be increased in various pathological conditions including rheumatoid arthritis and viral diseases [40].

Similar to IL-6 receptor, the receptors for granulocyte macrophage colony-stimulating factor (GM-CSF), IL-3, and IL-5 are type I cytokine receptors consisting of a ligand-specific α subunit and a common β subunit (CD131). Alternative splicing of GM-CSF receptor α subunit (GMR α) produces several soluble proteins, including one variant with a 102 nucleotide insertion derived from an Alu DNA repeat element [21] and another with exclusion of the exon encoding the transmembrane domain [41]. These soluble proteins act as functional antagonists of GM-CSF *in vitro*, but their biological roles *in vivo* are not yet clear [42]. IL-5 receptor α subunit (IL-5R α) has a soluble isoform-specific exon whose transcription exerts an antagonistic effect on eosinophil proliferation and differentiation. The balance of soluble and membrane-

bound IL-5R α regulated by splicing may play important role in IL-5 related diseases such as asthma and eosinophilic syndromes [43].

The class 2 cytokines consist of IL-10, IL-19, IL-20, IL-22, IL-24, IL-26, interferons, and interferon-like molecules. IL-24 is also known as melanoma differentiation-associated gene 7 (mda-7) since it is discovered as a tumor suppressor protein. Isoforms of IL-24 have differing levels of effectiveness in inducing apoptosis of cancer cells [44], and the loss of one isoform MDA-7S has been associated with metastatic melanoma [45]. IL-10 variant was also observed in cancer patients and its expression has been related to favorable response to chemotherapy [46]. The receptors for class 2 cytokines are similar to type I cytokine receptors except they do not possess the signature sequence WSXWS that is characteristic of type I receptors. It has been suggested that splicing of IL-20 receptor may be associated with the exacerbation of lupus nephritis [47].

Interleukin 1 receptor antagonist (IL-1Ra) competes with IL-1 α and IL-1 β for the binding sites of type I IL-1 receptor (IL-1RI), and acts as one of the control mechanisms of IL-1 signaling. Type II IL-1 receptor (IL-1RII) is a decoy receptor that inhibits the activity of its ligands – IL-1 α , IL-1 β , and IL-1RI. Cell-specific production of IL-1Ra splice variants has been demonstrated in human neutrophil and monocyte. While both LPS-stimulated neutrophils and PBMC synthesize soluble IL-1Ra and intracellular IL-1RaII (16 KDa), only PBMC transcribe and translate intracellular IL-1RI (18 KDa) [48]. Alternative splicing of IL-1RII generates soluble and membrane-bound forms of IL-1RII whose differential induction has been correlated with clinical glucocorticoid responsiveness in patients with autoimmune inner ear disease [49].

In addition to the production of cytokines and cytokine receptors, various spliced variants of intracellular molecules are involved in the signal transduction from a cytokine receptor to downstream targets. JAK/STAT pathways mediate signaling from a long list of membrane receptors including interferon, interleukin, hematopoietic and growth factor receptors. Signal transducers and activators of transcription (STAT) proteins are the products of at least seven genes, and additional isoforms of these gene products result from alternative splicing and post-translational processing. The resulting family of transcription factors generates a wide potential for activation and regulation of target genes and gene combinations with far-reaching consequences for development, cell growth and homeostasis [50, 51]. Another example is related to alternative splicing of IL-1R-associated kinase I (IRAK1) and myeloid differentiation primary-response gene 88 (MyD88). The smaller isoform of IRAK1 lacks kinase activity, which may function as part of a negative-feedback loop in which IRAK1 function is decreased in response to IL-1 stimulation [52, 53]. Full-length MyD88 protein recruits IRAK4 to activate IRAK1 in response to IL-1 binding, while the short isoform of MyD88 inhibits IL-1 signaling [53].

DYSREGULATED SPLICING IN AUTOIMMUNE DISEASES

Splicing regulation is critical for the normal functioning of human immune system, and dysregulated splicing has been shown to play an important role in various immune disorders. Many mutations or epigenetic changes may alter alternative splicing of disease-associated genes thus exerting their deleterious effects on immune response. Aberrant splicing may be a direct cause of the disease, or a modifier of disease susceptibility and severity. Those splice

variants may also be responsible for the non-responsiveness or drug resistance of patient subsets to treatments of autoimmune diseases [54].

Glucocorticoids regulate a variety of biological processes on virtually all physiological organ systems, and they have been used extensively as potent immunosuppressive agents in the treatments of many autoimmune and inflammatory disorders. However, insensitivity or resistance to glucocorticoids is often observed in patients taking these hormones. Such drug resistance may be partially explained by differential expression of two splice forms of glucocorticoid receptor (GR) [55]. The classic receptor variant GR α mediates most of the known functions of glucocorticoids, while the other variant GR β exerts a dominant negative effect upon the GR α -induced transcriptional activity. The latter isoform GR β has been shown to be highly expressed in glucocorticoids-resistant patients, which may block GR α -induced transcriptional activity causing the resistance to hormone treatment in those patients [56]. Proinflammatory cytokines such as IFN- α may be responsible for the splicing regulation of GR in lymphocytes or neutrophils from patients with autoimmune conditions [57-59]. In addition, one SNP in the 3'-UTR of GR β mRNA seems to increase its stability thus causing increased expression GR β protein. This SNP has been associated with increased risk of rheumatoid arthritis (RA) and systemic lupus erythematosus (SLE), suggesting the direct involvement of GR variant in the pathogenesis of rheumatic diseases [60].

SLE is a complex and heterogeneous autoimmune disease with a strong genetic component. Its pathophysiology involves a number of immunological pathways, such as the type I interferon pathway, B-cell signaling, cytokine production, and neutrophil function. Multiple susceptibility loci are already identified in genes participated in the function of those pathways, and some of them have been proposed to affect splicing.

Interferon regulatory factor 5 (IRF5) is a transcription factor that induces the transcription of interferon- α . There are more than 10 distinct isoforms of IRF5 with cell-specific expression, regulation, and function [61]. SLE patients express an IRF5 transcript variant signature that is distinct from healthy donors and associated with the IRF5-SLE H2 risk haplotype [62, 63]. This haplotype includes the T-allele of SNP rs2004640 that introduces a new donor splice site in the first exon [64], and the A-allele of SNP rs10954213 that provides an alternate polyA⁺ signal shortening the length of the 3'-UTR and enhancing IRF-5 stability [65].

The hallmark of SLE is the elevated production of auto-antibodies by B cells which leads to the formation of immune complexes causing tissue damage and chronic inflammation. B-cell scaffold protein with ankyrin repeats (BANK1) has been reported to be a susceptibility gene for SLE in Europeans, European-Americans, and Asians [66]. The non-synonymous SNP rs10516487 in BANK1 influences splicing efficiency by creating an exonic splicing enhancer site, and generates protein isoforms with differential self-association properties. The BANK1 variant with an increased self-assembly ability is highly produced in SLE patients, which could result in the more sustained BCR-mediated signaling in SLE. A T-cell marker CD5 was found to be expressed in B cells 30 years ago, and the CD5-positive B cell (B1) population has been considered to play a paradoxical role in SLE pathophysiology since then. The membrane expression of CD5 is regulated by alternative splicing of exon 1A and exon 1B. The full-length protein variant generated by exon 1A raises the threshold of BCR thus limits the response of auto-reactive B cells, while the truncated variant encoded by exon 1B remains in the cytoplasm. The truncated form is overexpressed in SLE patients, which leads to a decrease of CD5 cell surface expression and possible exacerbation of disease activity [67]. The pre-B-cell leukemia homeobox 1 (Pbx1) gene has a central role during development and

organogenesis. A novel splice isoform of Pbx1 has recently been found to express more frequently in the CD4⁺ T cells from lupus patients than from healthy controls. The presence of this variant correlates with an increased central memory T cell population, suggesting Pbx1 to be a novel lupus susceptibility gene that regulates T cell activation and tolerance [68].

Other aberrant splicing events identified in SLE include T-cell receptor ξ chain [69], CD72 [70], CD44 [71, 72], CTLA-4 [73], complement receptor 2 [74], Ras guanyl-nucleotide releasing protein 1 [75], leukocyte immunoglobulin-like receptor A2 [76], phospholipid scramblase 1 (PLSCR1) [77], and signalling lymphocyte activation molecule (SLAM) family receptors CS1 (CD319) and 2B4 (CD244) [78]. These examples identified by small-scale studies strongly suggested the importance of alternative splicing in SLE pathogenesis. A comprehensive evaluation of splicing regulation in SLE remains to be done, which shall greatly help us in understanding and management of this systemic autoimmune disease.

RA is a chronic inflammatory disorder that affects the body's joint and the extra-articular tissues of the skin, heart, and muscles. Increased flow of white blood cells into the synovium leads to inflammation which causes acute swelling and the destruction of cartilage and bone tissue in RA patients. Multiple pro-inflammatory cytokines, chemokines and growth factors play a fundamental role in its pathogenesis.

TNF- α is one of the most potent pro-inflammatory cytokines in RA, and anti-TNF α therapy has successfully controlled symptoms of many RA patients. However, about 25% to 30% of RA patients treated with TNF antagonists do not have any significant clinical response. Membrane-bound TNF receptors (TNFR1 and TNFR2) mediate TNF- α signaling, and their soluble forms may neutralize TNF- α in the circulation. A novel spliced variant of soluble TNFR2 (DS-TNFR2) has recently been identified to antagonize TNF- α action. RA patients with high levels of DS-TNFR2 had a prolonged therapeutic response to anti-TNF therapy and less radiographic progression than patients with normal levels [79].

Another potential biomarker of response to TNF- α antagonists is the soluble form of IL-7 receptor (sIL-7R) generated by alternative splicing of the full-length transcript. Transcripts for both membrane-bound (rIL-7R) and sIL-7R are over-expressed in synovial tissues of RA patients, and they are up-regulated by the addition of TNF- α and other pro-inflammatory cytokines in fibroblast-like synovial cells. However, mIL-7R protein was not detected on cell surface while sIL-7R levels were induced by those cytokines, indicating that sIL-7R is a marker of fibroblast. Strikingly, high baseline sIL-7R serum levels are significantly associated with poor response to TNF- α antagonist. It may qualify as a new biomarker of response to therapy in RA [80].

Similar to SLE, a large number of individual splicing events have been identified in RA by small scale studies while there lacks a comprehensive survey of splicing regulation in RA [81, 82]. Interestingly, multiple genes have been found to undergo aberrant splicing regulation in different autoimmune diseases. For example, increased CD44v6 expression on SLE T cells and RA monocytes has been shown to correlate with SLE and RA disease activity respectively. In contrast, CD44v3 expression was decreased on RA monocytes but increased on SLE T cells [72, 83]. Investigation of splicing regulation may reveal shared features and unique markers for different autoimmune diseases. In addition, therapeutic strategies may be developed to manipulate splicing for the treatment of SLE, RA, and other immune disorders [84].

SUMMARY

Further understanding of how alternative splicing may be used as a general mechanism in immune response is essential for our research in the pathophysiology of autoimmune diseases and development of new therapeutics for those diseases. Dysregulated variants may be developed as diagnosis, prognosis, pharmacodynamics, and predictive biomarkers for immune diseases. More importantly, certain isoforms of known disease targets may have unique epitopes within their extracellular region which would allow the development of specific monoclonal antibodies. Antisense technology may also be used to mediate alternative splicing for the treatment of different diseases [84].

The advent of high throughput RNA sequencing (RNA-seq) provides an unprecedented opportunity for the study of alternative splicing. It allows accurate quantitative measurements for alternative splicing levels and discovery of novel isoforms [85]. The ongoing improvements in technology and increased application of RNA-seq in immunology research shall greatly expand our knowledge in human immune systems and autoimmune diseases.

REFERENCES

- [1] Wang, ET; Sandberg, R; Luo, S; Khrebtkova, I; Zhang, L; Mayr, C; et al. Alternative isoform regulation in human tissue transcriptomes. *Nature.*, 2008 Nov 27; 456(7221), 470-476.
- [2] Wang, GS; Cooper, TA. Splicing in disease, disruption of the splicing code and the decoding machinery. *Nat Rev Genet.*, 2007 Oct; 8(10), 749-761.
- [3] Martinez, NM; Pan, Q; Cole, BS; Yarosh, CA; Babcock, GA; Heyd, F; et al. Alternative splicing networks regulated by signaling in human T cells. *RNA.*, 2012 May; 18(5), 1029-1040.
- [4] Butte, MJ; Lee, SJ; Jesneck, J; Keir, ME; Haining, WN; Sharpe, AH. CD28 costimulation regulates genome-wide effects on alternative splicing. *PLoS One.*, 2012; 7(6), e40032.
- [5] Grigoryev, YA; Kurian, SM; Nakorchevskiy, AA; Burke, JP; Campbell, D; Head, SR; et al. Genome-wide analysis of immune activation in human T and B cells reveals distinct classes of alternatively spliced genes. *PLoS One.* 2009; 4(11), e7906.
- [6] Birzele, F; Fauti, T; Stahl, H; Lenter, MC; Simon, E; Knebel, D; et al. Next-generation insights into regulatory T cells, expression profiling and FoxP3 occupancy in Human. *Nucleic Acids Res.*, 2011 Jul 4.
- [7] Tondeur, S; Pangault, C; Le Carrou, T; Lannay, Y; Benmahdi, R; Cubizolle, A; et al. Expression map of the human exome in CD34+ cells and blood cells, increased alternative splicing in cell motility and immune response genes. *PLoS One.*, 5(2), e8990.
- [8] Davidson, D; Viallet, J; Veillette, A. Unique catalytic properties dictate the enhanced function of p59fynT, the hemopoietic cell-specific isoform of the Fyn tyrosine protein kinase, in T cells. *Mol Cell Biol.*, 1994 Jul; 14(7), 4554-4564.
- [9] Weil, R; Levraud, JP; Dodon, MD; Bessia, C; Hazan, U; Kourilsky, P; et al. Altered expression of tyrosine kinases of the Src and Syk families in human T-cell leukemia virus type 1-infected T-cell lines. *J Virol.*, 1999 May; 73(5), 3709-3717.

- [10] Arch, R; Wirth, K; Hofmann, M; Ponta, H; Matzku, S; Herrlich, P; et al. Participation in normal immune responses of a metastasis-inducing splice variant of CD44. *Science.*, 1992 Jul 31; 257(5070), 682-685.
- [11] Konig, H; Ponta, H; Herrlich, P. Coupling of signal transduction to alternative pre-mRNA splicing by a composite splice regulator. *EMBO J.*, 1998 May 15; 17(10), 2904-2913.
- [12] Xu, Z; Weiss, A. Negative regulation of CD45 by differential homodimerization of the alternatively spliced isoforms. *Nature Immunology.*, 2002 Aug; 3(8), 764-771.
- [13] Magistrelli, G; Jeannin, P; Herbault, N; de Coignac, AB; Gauchat, JF; Bonnefoy, JY; et al. A soluble form of CTLA-4 generated by alternative splicing is expressed by nonstimulated human T cells. *European Journal of Immunology.*, 1999 Nov; 29(11), 3596-3602.
- [14] Ip, JY; Tong, A; Pan, Q; Topp, JD; Blencowe, BJ; Lynch, KW. Global analysis of alternative splicing during T-cell activation. *Rna-a Publication of the Rna Society.*, 2007 Apr; 13(4), 563-572.
- [15] Rao, N; Nguyen, S; Ngo, K; Fung-Leung, WP. A novel splice variant of interleukin-1 receptor (IL-1R)-associated kinase 1 plays a negative regulatory role in toll/IL-1R-induced inflammatory signaling. *Molecular and Cellular Biology.*, 2005 Aug; 25(15), 6521-6532.
- [16] Pleass, RJ; Andrews, PD; Kerr, MA; Woof, JM. Alternative splicing of the human IgA Fc receptor CD89 in neutrophils and eosinophils. *Biochemical Journal.*, 1996 Sep 15; 318, 771-777.
- [17] Eckhart, L; Henry, M; Santos-Beneit, AM; Schmitz, I; Krueger, A; Fischer, H; et al. Alternative splicing of caspase-8 mRNA during differentiation of human leukocytes. *Biochemical and Biophysical Research Communications.*, 2001 Dec 14; 289(4), 777-781.
- [18] Strickland, I; Kisich, K; Hauk, PJ; Vottero, A; Chrousos, GP; Klemm, DJ; et al. High constitutive glucocorticoid receptor beta in human neutrophils enables them to reduce their spontaneous rate of cell death in response to corticosteroids. *Journal of Experimental Medicine.*, 2001 Mar 5; 193(5), 585-593.
- [19] Di Fulvio, M; Gomez-Cambronero, J. Phospholipase D (PLD) gene expression in human neutrophils and HL-60 differentiation. *Journal of Leukocyte Biology.*, 2005 Jun; 77(6), 999-1007.
- [20] Park, JB; Levine, M. Cloning, sequencing, and characterization of alternatively spliced glutaredoxin 1 cDNA and its genomic gene - Chromosomal localization, mRNA stability, and origin of pseudogenes. *Journal of Biological Chemistry.*, 2005 Mar 18; 280(11), 10427-10434.
- [21] Pelley, JL; Nicholls, CD; Beattie, TL; Brown, CB. Discovery and characterization of a novel splice variant of the GM-CSF receptor alpha subunit. *Experimental Hematology.* 2007 Oct; 35(10), 1483-1494.
- [22] Nitto, T; Inoue, T; Node, K. Alternative spliced variants in the pantetheinase family of genes expressed in human neutrophils. *Gene.*, 2008 Dec 15; 426(1-2), 57-64.
- [23] Ottonello, L; Cutolo, M; Frumento, G; Arduino, N; Bertolotto, M; Mancini, M; et al. Synovial fluid from patients with rheumatoid arthritis inhibits neutrophil apoptosis, role of adenosine and proinflammatory cytokines. *Rheumatology.*, 2002 Nov; 41(11), 1249-1260.
- [24] Kreth, S; Ledderose, C; Kaufmann, I; Groeger, G; Thiel, M. Differential expression of 5'-UTR splice variants of the adenosine A2A receptor gene in human granulocytes,

- identification, characterization, and functional impact on activation. *FASEB J.*, 2008 Sep; 22(9), 3276-3286.
- [25] Rodrigues, R; Grosso, AR; Moita, L. Genome-Wide Analysis of Alternative Splicing during Dendritic Cell Response to a Bacterial Challenge. *PLoS One.*, 2013; 8(4), e61975.
- [26] Tsytisikov, VN; Yurovsky, VV; Atamas, SP; Alms, WJ; White, B. Identification and characterization of two alternative splice variants of human interleukin-2. *Journal of Biological Chemistry.*, 1996 Sep 20; 271(38), 23055-23060.
- [27] Janezic, AL; Bukilica, M; Jones, GV; Khanna, R; Morris-Stiff, G; Jurewicz, WA. Alternatively spliced variants of IL-2 mRNA in sequential transplant kidney core needle biopsies. *Transplantation Proceedings.*, 2001 Feb-Mar; 33(1-2), 383-386.
- [28] Sakkas, LI; Tourtellotte, C; Berney, S; Myers, AR; Platsoucas, CD. Increased levels of alternatively spliced interleukin 4 (IL-4delta2) transcripts in peripheral blood mononuclear cells from patients with systemic sclerosis. *Clin Diagn Lab Immunol.*, 1999 Sep; 6(5), 660-664.
- [29] Seah, GT; Gao, PS; Hopkin, JM; Rook, GA. Interleukin-4 and its alternatively spliced variant (IL-4delta2) in patients with atopic asthma. *Am J Respir Crit Care Med.*, 2001 Sep 15; 164(6), 1016-1018.
- [30] Horiuchi, S; Koyanagi, Y; Tanaka, Y; Waki, M; Matsumoto, A; Zhou, YW; et al. Altered interleukin-2 receptor alpha-chain is expressed in human T-cell leukaemia virus type-I-infected T-cell lines and human peripheral blood mononuclear cells of adult T-cell leukaemia patients through an alternative splicing mechanism. *Immunology.*, 1997 May; 91(1), 28-34.
- [31] Bergin, AM; Balder, B; Kishore, S; Sward, K; Hahn-Zoric, M; Lowhagen, O; et al. Common variations in the IL4R gene affect splicing and influence natural expression of the soluble isoform. *Hum Mutat.*, 2006 Oct; 27(10), 990-998.
- [32] Meazza, R; Gaggero, A; Neglia, F; Basso, S; Sforzini, S; Pereno, R; et al. Expression of two interleukin-15 mRNA isoforms in human tumors does not correlate with secretion, role of different signal peptides. *Eur J Immunol.*, 1997 May; 27(5), 1049-1054.
- [33] Vudattu, NK; Magalhaes, I; Hoehn, H; Pan, D; Maeurer, MJ. Expression analysis and functional activity of interleukin-7 splice variants. *Genes Immun.*, 2009 Mar; 10(2), 132-140.
- [34] Goodwin, RG; Friend, D; Ziegler, SF; Jerzy, R; Falk, BA; Gimpel, S; et al. Cloning of the human and murine interleukin-7 receptors, demonstration of a soluble form and homology to a new receptor superfamily. *Cell.*, 1990 Mar 23; 60(6), 941-951.
- [35] Grasso, L; Huang, M; Sullivan, CD; Messler, CJ; Kiser, MB; Dragwa, CR; et al. Molecular analysis of human interleukin-9 receptor transcripts in peripheral blood mononuclear cells. Identification of a splice variant encoding for a nonfunctional cell surface receptor. *J Biol Chem.*, 1998 Sep 11; 273(37), 24016-24024.
- [36] Dubois, S; Magrangeas, F; Lehours, P; Raher, S; Bernard, J; Boisteau, O; et al. Natural splicing of exon 2 of human interleukin-15 receptor alpha-chain mRNA results in a shortened form with a distinct pattern of expression. *J Biol Chem.*, 1999 Sep 17; 274(38), 26978-26984.
- [37] Rahman, M; Nara, H; Onoda, T; Araki, A; Li, J; Hoshino, T; et al. Cloning and characterization of an isoform of interleukin-21. *FEBS Lett.*, 2007 Aug 21; 581(21), 4001-4009.

- [38] Bihl, MP; Heinemann, K; Rudiger, JJ; Eickelberg, O; Perruchoud, AP; Tamm, M; et al. Identification of a novel IL-6 isoform binding to the endogenous IL-6 receptor. *Am J Respir Cell Mol Biol.*, 2002 Jul; 27(1), 48-56.
- [39] Alberti, L; Bachelot, T; Duc, A; Biota, C; Blay, JY. A spliced isoform of interleukin 6 mRNA produced by renal cell carcinoma encodes for an interleukin 6 inhibitor. *Cancer Res.*, 2005 Jan 1; 65(1), 2-5.
- [40] Horiuchi, S; Ampofo, W; Koyanagi, Y; Yamashita, A; Waki, M; Matsumoto, A; et al. High-level production of alternatively spliced soluble interleukin-6 receptor in serum of patients with adult T-cell leukaemia/HTLV-I-associated myelopathy. *Immunology.*, 1998 Nov; 95(3), 360-369.
- [41] Ashworth, A; Kraft, A. Cloning of a potentially soluble receptor for human GM-CSF. *Nucleic Acids Res.*, 1990 Dec 11; 18(23), 7178.
- [42] Brown, CB; Beaudry, P; Laing, TD; Shoemaker, S; Kaushansky, K. In vitro characterization of the human recombinant soluble granulocyte-macrophage colony-stimulating factor receptor. *Blood.*, 1995 Mar 15; 85(6), 1488-1495.
- [43] Tavernier, J; Van der Heyden, J; Verhee, A; Brusselle, G; Van Ostade, X; Vandekerckhove, J; et al. Interleukin 5 regulates the isoform expression of its own receptor alpha-subunit. *Blood.*, 2000 Mar 1; 95(5), 1600-1607.
- [44] Whitaker, EL; Filippov, V; Filippova, M; Guerrero-Juarez, CF; Duerksen-Hughes, PJ. Splice variants of mda-7/IL-24 differentially affect survival and induce apoptosis in U2OS cells. *Cytokine.*, 2011 Nov; 56(2), 272-281.
- [45] Allen, M; Pratscher, B; Roka, F; Krepler, C; Wacheck, V; Schofer, C; et al. Loss of novel mda-7 splice variant (mda-7s) expression is associated with metastatic melanoma. *J Invest Dermatol.*, 2004 Sep; 123(3), 583-588.
- [46] Wu, S; Gessner, R; Taube, T; von Stackelberg, A; Henze, G; Seeger, K. Expression of interleukin-10 splicing variants is a positive prognostic feature in relapsed childhood acute lymphoblastic leukemia. *J Clin Oncol.*, 2005 May 1; 23(13), 3038-3042.
- [47] Wei, CC; Chang, MS. A novel transcript of mouse interleukin-20 receptor acts on glomerular mesangial cells as an aggravating factor in lupus nephritis. *Genes Immun.*, 2008 Dec; 9(8), 668-679.
- [48] Malyak, M; Smith, MF, Jr; Abel, AA; Hance, KR; Arend, WP. The differential production of three forms of IL-1 receptor antagonist by human neutrophils and monocytes. *J Immunol.*, 1998 Aug 15; 161(4), 2004-2010.
- [49] Vambutas, A; DeVoti, J; Goldofsky, E; Gordon, M; Lesser, M; Bonagura, V. Alternate splicing of interleukin-1 receptor type II (IL1R2) in vitro correlates with clinical glucocorticoid responsiveness in patients with AIED. *PLoS One.*, 2009; 4(4), e5293.
- [50] Levy, DE; Darnell, JE, Jr. Stats, transcriptional control and biological impact. *Nat Rev Mol Cell Biol.*, 2002 Sep; 3(9), 651-662.
- [51] Watanabe, Y; Ikegawa, M; Naruse, Y; Tanaka, M. A novel splicing variant form suppresses the activity of full-length signal transducer and activator of transcription 5A. *FEBS J.*, 2009 Nov; 276(21), 6312-6323.
- [52] Jensen, LE; Whitehead, AS. IRAK1b, a novel alternative splice variant of interleukin-1 receptor-associated kinase (IRAK), mediates interleukin-1 signaling and has prolonged stability. *J Biol Chem.*, 2001 Aug 3; 276(31), 29037-29044.
- [53] Burns, K; Janssens, S; Brissoni, B; Olivos, N; Beyaert, R; Tschopp, J. Inhibition of interleukin 1 receptor/Toll-like receptor signaling through the alternatively spliced, short

- form of MyD88 is due to its failure to recruit IRAK-4. *J Exp Med.*, 2003 Jan 20; 197(2), 263-268.
- [54] Evsyukova, I; Somarelli, JA; Gregory, SG; Garcia-Blanco, MA. Alternative splicing in multiple sclerosis and other autoimmune diseases. *RNA Biol.*, 2010 Jul-Aug; 7(4), 462-473.
- [55] Kino, T; Charmandari, E; Chrousos, GP. Glucocorticoid receptor, implications for rheumatic diseases. *Clin Exp Rheumatol.*, Sep-Oct; 29(5 Suppl 68), S32-41.
- [56] Piotrowski, P; Burzynski, M; Lianeri, M; Mostowska, M; Wudarski, M; Chwalinska-Sadowska, H; et al. Glucocorticoid receptor beta splice variant expression in patients with high and low activity of systemic lupus erythematosus. *Folia Histochem Cytobiol.*, 2007; 45(4), 339-342.
- [57] Strickland, I; Kisich, K; Hauk, PJ; Vottero, A; Chrousos, GP; Klemm, DJ; et al. High constitutive glucocorticoid receptor beta in human neutrophils enables them to reduce their spontaneous rate of cell death in response to corticosteroids. *J Exp Med.*, 2001 Mar 5; 193(5), 585-593.
- [58] Tliba, O; Damera, G; Banerjee, A; Gu, S; Baidouri, H; Keslacy, S; et al. Cytokines induce an early steroid resistance in airway smooth muscle cells, novel role of interferon regulatory factor-1. *Am J Respir Cell Mol Biol.*, 2008 Apr; 38(4), 463-472.
- [59] Xu, Q; Leung, DY; Kisich, KO. Serine-arginine-rich protein p30 directs alternative splicing of glucocorticoid receptor pre-mRNA to glucocorticoid receptor beta in neutrophils. *J Biol Chem.*, 2003 Jul 18; 278(29), 27112-27118.
- [60] Derijk, RH; Schaaf, MJ; Turner, G; Datson, NA; Vreugdenhil, E; Cidlowski, J; et al. A human glucocorticoid receptor gene variant that increases the stability of the glucocorticoid receptor beta-isoform mRNA is associated with rheumatoid arthritis. *J Rheumatol.*, 2001 Nov; 28(11), 2383-2388.
- [61] Mancl, ME; Hu, G; Sangster-Guity, N; Olshalsky, SL; Hoops, K; Fitzgerald-Bocarsly, P; et al. Two discrete promoters regulate the alternatively spliced human interferon regulatory factor-5 isoforms. Multiple isoforms with distinct cell type-specific expression, localization, regulation, and function. *J Biol Chem.*, 2005 Jun 3; 280(22), 21078-21090.
- [62] Stone, RC; Du, P; Feng, D; Dhawan, K; Ronnblom, L; Eloranta, ML; et al. RNA-Seq for enrichment and analysis of IRF5 transcript expression in SLE. *PLoS One.*, 2013; 8(1), e54487.
- [63] Feng, D; Stone, RC; Eloranta, ML; Sangster-Guity, N; Nordmark, G; Sigurdsson, S; et al. Genetic variants and disease-associated factors contribute to enhanced interferon regulatory factor 5 expression in blood cells of patients with systemic lupus erythematosus. *Arthritis Rheum.*, 2010 Feb; 62(2), 562-573.
- [64] Graham, RR; Kozyrev, SV; Baechler, EC; Reddy, MV; Plenge, RM; Bauer, JW; et al. A common haplotype of interferon regulatory factor 5 (IRF5) regulates splicing and expression and is associated with increased risk of systemic lupus erythematosus. *Nat Genet.*, 2006 May; 38(5), 550-555.
- [65] Graham, RR; Kyogoku, C; Sigurdsson, S; Vlasova, IA; Davies, LR; Baechler, EC; et al. Three functional variants of IFN regulatory factor 5 (IRF5) define risk and protective haplotypes for human lupus. *Proc Natl Acad Sci U S A.*, 2007 Apr 17; 104(16), 6758-6763.

- [66] Kozyrev, SV; Abelson, AK; Wojcik, J; Zaghlool, A; Linga Reddy, MV; Sanchez, E; et al. Functional variants in the B-cell gene BANK1 are associated with systemic lupus erythematosus. *Nat Genet.*, 2008 Feb; 40(2), 211-216.
- [67] Youinou, P; Renaudineau, Y. CD5 expression in B cells from patients with systemic lupus erythematosus. *Crit Rev Immunol.*, 31(1), 31-42.
- [68] Cuda, CM; Li, S; Liang, S; Yin, Y; Potula, HH; Xu, Z; et al. Pre-B cell leukemia homeobox 1 is associated with lupus susceptibility in mice and humans. *J Immunol.*, Jan 15; 188(2), 604-614.
- [69] Tsuzaka, K; Setoyama, Y; Yoshimoto, K; Shiraishi, K; Suzuki, K; Abe, T; et al. A splice variant of the TCR zeta mRNA lacking exon 7 leads to the down-regulation of TCR zeta, the TCR/CD3 complex, and IL-2 production in systemic lupus erythematosus T cells. *J Immunol.*, 2005 Mar 15; 174(6), 3518-3525.
- [70] Hitomi, Y; Tsuchiya, N; Kawasaki, A; Ohashi, J; Suzuki, T; Kyogoku, C; et al. CD72 polymorphisms associated with alternative splicing modify susceptibility to human systemic lupus erythematosus through epistatic interaction with FCGR2B. *Hum Mol Genet.*, 2004 Dec 1; 13(23), 2907-2917.
- [71] Crispin, JC; Keenan, BT; Finnell, MD; Bermas, BL; Schur, P; Massarotti, E; et al. Expression of CD44 variant isoforms CD44v3 and CD44v6 is increased on T cells from patients with systemic lupus erythematosus and is correlated with disease activity. *Arthritis Rheum.*, May; 62(5), 1431-1437.
- [72] Crispin, JC; Keenan, BT; Finnell, MD; Bermas, BL; Schur, P; Massarotti, E; et al. Expression of CD44 variant isoforms CD44v3 and CD44v6 is increased on T cells from patients with systemic lupus erythematosus and is correlated with disease activity. *Arthritis Rheum.*, 2010 May; 62(5), 1431-1437.
- [73] Ichinose, K; Zhang, Z; Koga, T; Juang, YT; Kis-Toth, K; Sharpe, AH; et al. Brief Report, Increased expression of a short splice variant of CTLA-4 exacerbates lupus in MRL/lpr mice. *Arthritis Rheum.*, Mar; 65(3), 764-769.
- [74] Douglas, KB; Windels, DC; Zhao, J; Gadeliya, AV; Wu, H; Kaufman, KM; et al. Complement receptor 2 polymorphisms associated with systemic lupus erythematosus modulate alternative splicing. *Genes Immun.*, 2009 Jul; 10(5), 457-469.
- [75] Yasuda, S; Stevens, RL; Terada, T; Takeda, M; Hashimoto, T; Fukae, J; et al. Defective expression of Ras guanyl nucleotide-releasing protein 1 in a subset of patients with systemic lupus erythematosus. *J Immunol.*, 2007 Oct 1; 179(7), 4890-4900.
- [76] Mamegano, K; Kuroki, K; Miyashita, R; Kusaoi, M; Kobayashi, S; Matsuta, K; et al. Association of LILRA2 (ILT1, LIR7) splice site polymorphism with systemic lupus erythematosus and microscopic polyangiitis. *Genes Immun.*, 2008 Apr; 9(3), 214-223.
- [77] Suzuki, E; Amengual, O; Atsumi, T; Oku, K; Hashimoto, T; Kataoka, H; et al. Increased expression of phospholipid scramblase 1 in monocytes from patients with systemic lupus erythematosus. *J Rheumatol.*, Aug 1; 37(8), 1639-1645.
- [78] Kim, JR; Mathew, SO; Patel, RK; Pertusi, RM; Mathew, PA. Altered expression of signalling lymphocyte activation molecule (SLAM) family receptors CS1 (CD319) and 2B4 (CD244) in patients with systemic lupus erythematosus. *Clin Exp Immunol.*, Jun; 160(3), 348-358.
- [79] Canete, JD; Albaladejo, C; Hernandez, MV; Lainez, B; Pinto, JA; Ramirez, J; et al. Clinical significance of high levels of soluble tumour necrosis factor-alpha receptor-2

- produced by alternative splicing in rheumatoid arthritis, a longitudinal prospective cohort study. *Rheumatology* (Oxford). Apr; 50(4), 721-728.
- [80] Badot, V; Durez, P; Van den Eynde, BJ; Nzeusseu-Toukap, A; Houssiau, FA; Lauwerys, BR. Rheumatoid arthritis synovial fibroblasts produce a soluble form of the interleukin-7 receptor in response to pro-inflammatory cytokines. *J Cell Mol Med*. Nov; 15(11), 2335-2342.
- [81] Hasegawa, M; Nakoshi, Y; Muraki, M; Sudo, A; Kinoshita, N; Yoshida, T; et al. Expression of large tenascin-C splice variants in synovial fluid of patients with rheumatoid arthritis. *J Orthop Res.*, 2007 May; 25(5), 563-568.
- [82] Michel, J; Langstein, J; Hofstadter, F; Schwarz, H. A soluble form of CD137 (ILA/4-1BB), a member of the TNF receptor family, is released by activated lymphocytes and is detectable in sera of patients with rheumatoid arthritis. *Eur J Immunol.*, 1998 Jan; 28(1), 290-295.
- [83] Grisar, J; Munk, M; Steiner, CW; Amoyo-Minar, L; Tohidast-Akrad, M; Zenz, P; et al. Expression patterns of CD44 and CD44 splice variants in patients with rheumatoid arthritis. *Clin Exp Rheumatol.*, Jan-Feb; 30(1), 64-72.
- [84] Mourich, DV; Iversen, PL. Splicing in the immune system, potential targets for therapeutic intervention by antisense-mediated alternative splicing. *Curr Opin Mol Ther.*, 2009 Apr; 11(2), 124-132.
- [85] Garber, M; Grabherr, MG; Guttman, M; Trapnell, C. Computational methods for transcriptome annotation and quantification using RNA-seq. *Nat Methods*. Jun; 8(6), 469-477.

Chapter 27

POLY(ADP-RIBOSYL)ATION REGULATES ALTERNATIVE SPLICING

Yingbiao Ji and Alexei V. Tulin*

Cancer Biology Program, Fox Chase Cancer Center, Philadelphia, PA, US

ABSTRACT

Poly(ADP-ribosyl)ation is a major post-translational modification performed by Poly(ADP-ribose) Polymerases (PARP1). PARP1 utilizes NAD as the substrate to synthesize a long linear and branching poly(ADP-ribose) (pADPr), ranging in length from 2 to 200 units of ADP-ribose. PARP1 can modify a variety of proteins by attaching pADPr to the target proteins in either a covalent or noncovalent manner. In this way, Poly(ADP-ribosyl)ation is involved in a number of biological processes, including transcription regulation, stress responses, and apoptosis. Recent studies have demonstrated that several splicing factors, including hnRNPs and S/R proteins, are poly(ADP-ribosyl)ated via noncovalent binding by PARP1. Here, we describe how poly(ADP-ribosyl)ation regulates alternative splicing by modulating the activities of these splicing factors. In addition, we discuss a possible role of PARP1 in coupling transcription with splicing.

INTRODUCTION

Alternative splicing (AS) greatly enhances the proteomic complexity of an organism with otherwise limited numbers of encoding genes in the genome [1]. For instance, the *Drosophila* Dscam (Down syndrome cell adhesion molecule) gene can produce 38,016 different isoforms through alternative splicing [2]. AS is achieved by spliceosomes that choose alternative 5' splicing sites (5'ss) or 3'ss to remove the intron of a pre-mRNA. A spliceosome is composed of five small nuclear RNA (snRNA) molecules and associated proteins, such as snRNPs [3]. The interaction between splicing regulatory elements (SREs) and transacting factors defines

* Corresponding Author's Email: Alexei.Tulin@fccc.edu (Alexei V. Tulin, Ph.D. Fox Chase Cancer Center, 333 Cottman Avenue Philadelphia, PA 19111; Tel: (215) 728-7408; Fax: (215) 728-2412).

the exact splicing sites to generate spliced variants. SREs include exonic and intronic splicing enhancers (ESSs and ISSs) and silencers (ESEs and ISEs). The two major groups of transacting factors, named hnRNPs and SR (serine-arginine-rich) proteins, bind to SREs and enhance or inhibit the recruitment of a spliceosome to the splice sites. Therefore, controlling the expression levels and activities of hnRNPs and SR proteins is critical for achieving tissue- or developmental-specific splicing regulation.

Poly(ADP-ribosyl)ation (PARylation) is a specific post-translational modification performed by Poly(ADP-ribose) polymerases (PARPs), which utilize NAD as the substrate to synthesize poly(ADP-ribose) polymer (pADPr) [4]. Although the PARP superfamily includes 17 members (PARP1-PARP17) in the mammalian genome [5], PARP1 enzymatic activity is responsible for producing 90% of pADPr products in cells [4]. PARP1 mainly modifies itself in a covalent manner through Glu/Asp and/or lysine residues in its automodification domain [6, 7]. In addition, automodified PARP1, or free pADPr, can bind to a variety of targets, including histones and other chromatin-associated proteins, in a noncovalent manner [8]. To reverse the protein modifications by PARylation, Poly(ADP-ribose) glycohydrolase (PARG) degrades pADPr to a single unit [9-11]. Because pADPr is a highly negatively charged, and often branching, macromolecule varying in size from two to up to 200 ADP-ribose, PARylation often dramatically changes the chromatin structure and thereby regulates transcription processes [4, 12, 13]. Recent studies have shown that the activities of hnRNPs and SR proteins can be regulated by PARylation to modulate alternative splicing [14-16]. In this review, we summarize how poly(ADP-ribosyl)ation regulates alternative splicing by modulating the activities of the splicing factors. In addition, we address a possible role of PARP1 in coupling transcription with splicing.

1. REGULATION OF ALTERNATIVE SPLICING BY PARYLATION

1.1. Interaction of hnRNPs with Poly(ADP-ribose)

It was first reported that pADPr could bind to human hnRNP proteins in a noncovalent manner via a conserved domain [14]. This domain represents a 20 amino acid-long segment located between two RNA-binding domains of hnRNPA1 and belongs to the first pADPr-binding motif identified in DNA-damage-related proteins [17]. A mass spectrometry analysis coupled with co-immunoprecipitation demonstrated that *Drosophila* hnRNPA1 (Hrp38) is associated with PARP1 *in vivo* [18]. It was also found that PARP1 could modify two hnRNP proteins (hrp38 and Hrp40) by noncovalent pADPr binding [16], suggesting that PARylation of hnRNPs is a conserved mechanism between mammals and *Drosophila* [14, 16]. Since PARP1-pADPr-hnRNP complex has been detected in *Drosophila*, it appears that automodified PARP1 interacts with hnRNP proteins *in vivo* [16]. Both overexpression of PARP1 and PARG loss-of-function mutation increase pADPr levels and also cause an increased level of PARylated hnRNPs in *Drosophila* [16]. Heat-shock treatment also enhances hnRNP PARylation by the induction of PARP1 activity [16, 19]. These results demonstrate that hnRNP PARylation is a post-translational modification which can be regulated at the genetic and physiological levels.

1.2. Inhibition of RNA-Binding Ability of hnRNP Proteins by PARylation

Both *in vivo* and *in vitro* experiments showed that PARylation inhibits the RNA-binding ability of hnRNP proteins [16, 20]. *Drosophila* Hrp38 is associated with the actively transcribed loci known as puffs in the polytene chromosomes found in salivary gland cells of the wild-type fruit fly [16]. In *Parg* mutants, Hrp38 dissociates from puffs and moves into the nucleoplasm, where Hrp38 appears to be highly PARylated [16]. Therefore, the observed dissociation of Hrp38 from most of the transcripts in the *Parg* mutant background is most likely triggered by accumulation of pADPr and overwhelming PARylation in the absence of PARG. In another line of evidence supporting this argument, Hrp38 was shown to reduce binding to one long noncoding RNA (Hsr ω -n) upon heat-shock treatment in the *Parg* mutant [16]. Hrp38 and Hrp40 dissociate from most transcripts and exclusively bind to this heat shock-induced noncoding RNA after heat shock [16, 21]. Since heat shock induces PARP1 activity and, hence, increases hnRNP PARylation [16], it has been proposed that PARylation acts as a molecular tool which moves hnRNPs from the chromatin to the nucleoplasm. In addition, it was observed that much smaller amounts of Hrp38 and Hrp40 were associated with this noncoding RNA in the *Parg* mutant compared to the wild-type animal after heat shock treatment [16], suggesting that PARylation inhibits the target-binding ability of hnRNPs. *In vitro* experiments based on the RNA electrophoretic mobility shift assay (EMAS) also demonstrated that pADPr binding to hnRNPs inhibits the RNA-binding ability of hnRNPs [16, 20, 22]. It was well documented that human hnRNPA1 strongly binds to a G-rich RNA element [23]. Similarly, it was found that Hrp38 could bind to a G-rich motif in the 5'UTR of DE-cadherin mRNA [20]. Incubation of human hnRNPA1 or Hrp38 with pADPr greatly reduced the amount of hnRNPA1 bound to its target RNA element [16, 20]. Collectively, these data demonstrate that PARylation of hnRNPs inhibits their RNA-binding properties, similar to the effect of PARylation on other RNA-binding proteins, such as NONO [24], Ago2 [25] and Poly(A) polymerase (PAP) [26].

1.3. Splicing Regulation by PARylation of hnRNPs

Experimental evidence has shown that PARylation can regulate hnRNP-dependent alternative splicing because hnRNP PARylation inhibits the RNA-binding ability of hnRNPs [16] (Figure 1). It is generally agreed that hnRNPs act as splicing repressors by binding to exonic/intronic splicing silencers (ESSs and ISSs) [27]. The inhibitory effect of hnRNPs on splicing is mainly attributable to the competition with S/R proteins for binding the regulatory elements. In some cases, however, hnRNPs were also shown to promote splicing by binding to exonic and intronic splicing enhancers (ESEs and ISEs) [28-30]. Since Hrp38 and Hrp40 knockout significantly increased the ratio of the spliced isoform to the skipped isoform of the *Ddc* (dopa decarboxylase) gene in *Drosophila*, Hrp38 and Hrp40 appear to function as splicing repressors for splicing this gene [16]. Similarly, PARylation of these two proteins by either *Parg* knockout or PARP1 activation promoted splicing of *Ddc* pre-mRNA by inhibiting hnRNP association with the splicing silencers of this pre-mRNA [16]. Hrp38s were also shown to function as splicing activators by promoting pre-*hsr ω* mRNA splicing [16]. However, Hrp38 PARylation by *Parg* knockout caused Hrp38 dissociation from the *hsr ω* pre-mRNA, thus inhibiting splicing [16]. Therefore, hnRNP PARylation can either promote or inhibit alternative

splicing during differently regulated splicing events, depending on whether hnRNPs function as splicing repressors or activators.

1.4. Splicing Regulation by PARylation of S/R Proteins

PARP1 can also modify SR (serine-arginine-rich) proteins in a noncovalent manner, including ASF/SF2 [15], SF3B1 [31], SF3A1, and SF3B2 [32]. It is generally agreed that S/R proteins function as splicing activators by binding to exonic splicing enhancers to antagonize the inhibitory effects of hnRNPs on splicing [33]. The SR domain of an S/R protein is often phosphorylated by SR kinases to regulate splicing. DNA Topoisomerase I (Topo I) was identified as one of SR kinases that phosphorylates the serine residues of ASF/SF2 and thus promotes alternative splicing [34]. It has been shown that pADPr can inhibit the kinase activity of Topo I, thereby decreasing phosphorylation of ASF/SF2 [15]. Mechanistically, pADPr polymers bind to ASF/SF2 either via RRM1 or the RS domain of ASF/SF2 to compete with Topo I for substrate binding [15]. Consequently, PARylation of SR proteins likely regulates alternative splicing by inhibiting phosphorylation of SR protein.

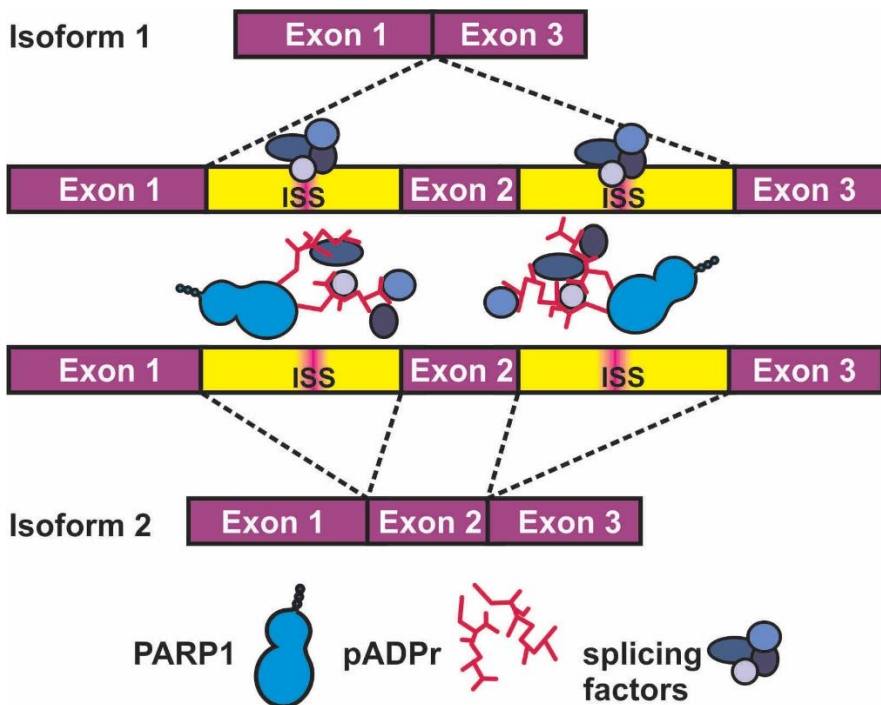


Figure 1. The model for the modulation of splicing by hnRNP poly(ADP-ribosyl)ation. Splicing enhancement by poly(ADP-ribosyl)ation of hnRNPs. In hypoderm tissues, the binding of Hrp38 and Hrp40 to the G triplets or quartets as intronic splicing silencers (ISSs) in introns 1 and 2 of the Ddc pre-mRNA inhibits splicing and results in exon skipping. The dissociation of hnRNPs from the transcript, such as the Ddc pre-mRNA, by poly(ADP-ribosyl)ation of hnRNPs induced by heat shock or elevated PARP1 activity in CNS enhances intron splicing.

2. REGULATION OF SPLICING BY PARP1 DURING THE TRANSCRIPTION PROCESS

2.1. Coupling Transcription with Splicing by Epigenetic Mechanisms

Genome-wide studies on splicing in yeast [35], *Drosophila* [36], and mammalian cells [37] have revealed that most splicing events are coupled with transcription. Co-transcriptional splicing implies that introns are spliced and that exons are ligated during the transcription process when the pre-mRNAs are still tethered to the DNA templates in the chromatin [1]. Because it takes longer to assemble a spliceosome than it does to complete elongation *in vivo*, splicing rate is much slower than the transcription elongation rate [1]. The key question for understanding co-transcriptional splicing is how the transcriptional machinery coordinates intron removal with the spliceosome before the termination of transcription. In addition, it is not clear what mechanisms recruit splicing factors to the nascent transcripts in order to achieve alternative splicing. Recent studies suggest that specific chromatin modifications may facilitate the coupling of splicing with transcription [38]. Eukaryotic chromatin is composed of highly organized nucleosomes formed by the core histones (H2A, H2B, H3 and H4) wrapping around the 146-bp DNA template. The N-terminal tails of histones undergo extensive post-translational modifications, including acetylation, methylation, phosphorylation, and PARylation, which modulate chromatin structure [39]. Changing the chromatin landscape has a profound effect on the status of transcription. Certain modifications, such as histone acetylation, lead to chromatin loosening, thereby facilitating transcription. In contrast, the heterochromatin, where transcription is generally repressed, is often marked by histone methylation, such as the enrichment of H3K27me3 on inactive X-chromosomes in mammals [40]. Similar to labeling active/ inactive chromatin regions, histone modifications are also found to facilitate splicing site selections by marking introns and exons [41, 42]. Nucleosome density is higher in exons than in introns [42]. For instance, trimethylation of Lys36 in histone H3 (H3K36me3) is preferentially associated with exons, not introns, in *C. elegans* [41] and human [42]. However, when compared to constitutively spliced exons, it appears that the alternatively spliced exons are marked with H3K36me3 to a lesser degree [41]. It has been proposed that histone codes can be read by an adaptor system, which then transmits the splicing signals from the chromatin to the spliceosome [43-46]. For example, the chromatin-binding protein MRG15 recognizes H3K36me3 in the alternatively spliced exon IIIB in the human fibroblast growth factor receptor 2 (FGFR2) gene and proceeds to recruit the splicing repressor PTB to the regulatory elements of this exon, repressing splicing [43]. The chromatin-associated protein Psp1/Ledgf can bind to H3K36me3 and recruit splicing factors for splicing regulation [46]. Another chromatin protein, CHD1, reads H3K4me3 (trimethylation of Lys4 in histone H3) codes and brings U2 snRNP to the chromatin, thereby activating splicing [44]. Because it was established that HP1a can recognize H3K9me3 in the chromatin [47], the interaction between HP1a with hnRNP proteins, such as DDP1, HRB87F and PEP, may represent another mechanism that brings hnRNPs to chromatin. Alternatively, transmission of the splicing signal from chromatin histone codes can be achieved by direct interaction between chromatin proteins and nascent transcripts [45]. For instance, HP1 γ can recognize H3K9me3 in alternatively spliced exons, such as CD44, to slow down the elongation rate of RNA polymerase II, facilitating exon inclusion [45]. It appeared that HP1 γ could also bind to the regulated pre-mRNA to effectively couple

transcription with alternative splicing [45]. In addition, the RNA-binding protein Hu can directly inhibit the activity of histone deacetylase 2 (HDAC2), thereby inducing local histone hyperacetylation [48]. Elevated histone hyperacetylation around the alternatively spliced exons made chromatin more relaxed, increasing the local transcription elongation rate and decreasing exon inclusion [48]. The aforementioned examples illustrate the multiplicity of mechanisms allowing the coupling transcription with splicing.

2.2. Chromatin Remodeling by PARP1

PARP1 plays a pivotal role in regulating chromatin structure [49,50]. Enzymatically silent PARP1 binds to the core histone proteins, thereby contributing to chromatin compaction for transcription inhibition [50]. PARP1 preferentially binds to H3 and H4 in mononucleosomes with linker DNA *in vitro* [18, 51]. H4 can activate PARP1, which, in turn, activates PARylation [18]. Activated PARP1 can PARylate itself, histone proteins, and other chromatin-associated proteins, leading to chromatin loosening, which facilitates the passage of RNA Polymerase II through the DNA molecule [13]. Phosphorylation of Ser137 in *Drosophila* H2Av histone variant (homolog of mammalian histone H2Az/H2Ax) induces PARP1 activation for transcription of such target genes as the heat shock-induced *Hsp70* gene [52]. In turn, phosphorylation of H2Av-Ser137 may stimulate the activity of dTip60, a histone acetyltransferase, leading to acetylation of H2AvK5, which facilitates the exposure of H4 tail for PARP1 activation [53]. PARP1 can modify histones and a number of chromatin-associated proteins, thereby modulating chromatin structure during transcription [54]. LPS (lipopolysaccharide) treatment of macrophages can stimulate PARP1-dependent transcription of inflammatory genes, including IL-1 β gene [55]. PARP1 can PARylate the histone demethylase KDM5B, thereby inhibiting its activity to maintain H3K4me3 levels in the promoter regions of the positively regulated genes [56]. A recent study showed that PARP1 interacts with a histone methyltransferase (Suz12) to maintain and localize it on the chromatin [57]. Therefore, inhibition of PARP1 activity by PARP1 inhibitor PJ34 can decrease the level of H3K9 and H3K27 modifications in pronuclei of male mice [57]. In addition, poly(ADP-ribose) can bind to the nucleosome remodeling ATPases, including ALC1 [58] and dMi2 [59], recruiting them to the transcriptionally active regions for modeling chromatin structure. Together, these studies suggest that PARP1 is the key regulator, modulating chromatin status during transcription.

2.3. Linking Transcription with Splicing by PARP1

As discussed above, PARP1 activation is required for the expression of inducible genes by making chromatin more accessible to the transcription machinery. Several lines of evidence support the hypothesis that PARP1 couples transcription with splicing directly. PARP1 activation leads to chromatin loosening in the ecdysone-inducible E74 puff in *Drosophila*, where both pADPr and Hrp40 are accumulated [16, 19]. Indeed, the first indication that splicing occurs co-transcriptionally came from the observation that the 60 kb primary transcript of the E74 gene was spliced to form 6.0 kb mRNA before transcription termination [60]. Histone modifications, such as H3 phosphoacetylation and chromatin remodeling by the chromatin

remodeling factor NURF301, are also involved in changing chromatin structure during the expression of ecdysone-inducible genes [61, 62]. Consequently, it seems plausible that PARP1 could couple chromatin remodeling and splicing by recruiting splicing factors, such as hnRNP proteins, to the transcription site during the transcription process (Figure 2).

3. PERSPECTIVES

Based on the experimental evidence summarized above, Poly(ADP-ribosyl)ation can regulate both transcription and splicing processes. Poly(ADP-ribosyl)ation of histones and chromatin-associated proteins mediated by PARP1 dramatically changes chromatin structure, facilitating the transcription of inducible genes during development. Poly(ADP-ribosyl)ation also regulates the interaction between RNA-binding proteins and their RNA targets to modulating alternative splicing. Therefore, PARP1 might function as an adaptor protein, transmitting histone modification information contained in chromatin to determine the splicing pattern by interacting with splicing factors, such as hnRNP proteins. However, several questions need to be addressed to further substantiate this assumption. Specifically, it is important to determine whether PARP1 preferentially recognizes specific histone codes, such as H3K9me3 and/or H3K36me3, in relation to exonic definitions. We might also ask 1) what role histone poly(ADP-ribosyl)ation plays in determining the transcription rates for splicing regulation and 2) whether PARP1 recruits splicing factors to the transcription sites in a pADPr-dependent manner. A detailed investigation of these questions will help us understand how splicing is coupled with transcription.

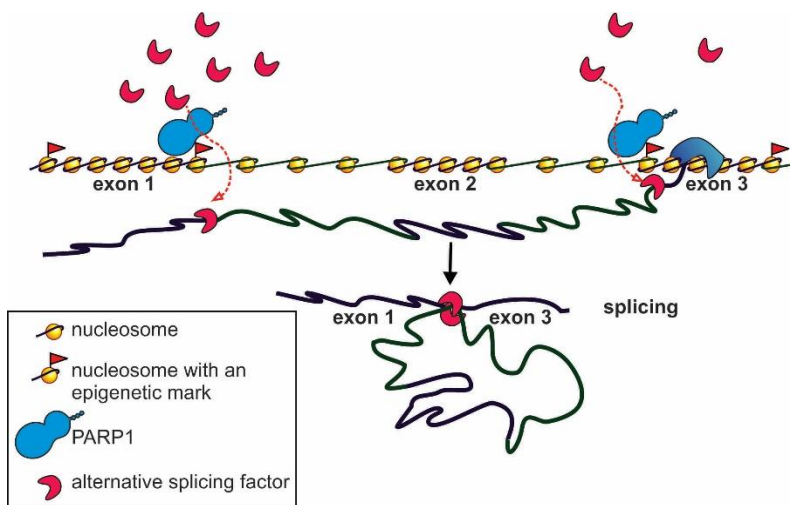


Figure 2. PARP1 binding to specific epigenetic marks in chromatin facilitates loading alternative splicing factors to specific locations of a pre-mRNA. Nucleosomes with specific epigenetic markers (red flags) localized in constitutively spliced exons preferentially bind PARP1 (blue snowman), which proceeds to attract splicing factors (magenta) and position them in the marked areas to cause exon skipping.

ACKNOWLEDGMENTS

We thank Dr. Kate Pechenkina for critical reading of the manuscript and valuable comments. The expenses were defrayed by a grant from the National Institutes of Health (R01 GM077452) (to A.V.T.).

REFERENCES

- [1] Chen, M. & Manley J. L. (2009). Mechanisms of alternative splicing regulation: insights from molecular and genomics approaches. *Nat Rev Mol Cell Biol*, 10(11), 741-754.
- [2] Graveley, B. R., et al. (2004). The organization and evolution of the Dipteran and Hymenopteran Down syndrome cell adhesion molecule (Dscam) genes. *RNA*, 10(10), 1499-1506.
- [3] Will, C. L. & Lührmann, R.. Spliceosome Structure and Function. *Cold Spring Harbor Perspectives in Biology*, 3(7).
- [4] D'Amours, D., et al. (1999). Poly(ADP-ribosyl)ation reactions in the regulation of nuclear functions. *Biochem. J.*, 342(2), 249-268.
- [5] Otto, H., et al. (2005). In silico characterization of the family of PARP-like poly(ADP-ribosyl)transferases (pARTs). *BMC Genomics*, 6(1), 139.
- [6] Tao, Z., Gao, P. & Liu, H.-w. (2009). Identification of the ADP-Ribosylation Sites in the PARP-1 Automodification Domain: Analysis and Implications. *Journal of the American Chemical Society*, 131(40): 14258-14260.
- [7] Altmeyer, M., et al. (2009). Molecular mechanism of poly (ADP-ribosyl) ation by PARP1 and identification of lysine residues as ADP-ribose acceptor sites. *Nucleic Acids Research*, 37(11), 3723-3738.
- [8] Krietsch, J., et al. (2012). Reprogramming cellular events by poly (ADP-ribose)-binding proteins. *Molecular Aspects of Medicine* (In press).
- [9] Hanai, S., et al. (2004). Loss of poly(ADP-ribose) glycohydrolase causes progressive neurodegeneration in *Drosophila melanogaster*. *Proceedings of the National Academy of Sciences of the United States of America*, 101(1), 82-86.
- [10] Koh, D.W., et al. (2004). Failure to degrade poly(ADP-ribose) causes increased sensitivity to cytotoxicity and early embryonic lethality. *Proceedings of the National Academy of Sciences of the United States of America*, 101(51), 17699-17704.
- [11] Tulin, A., et al. (2006). *Drosophila* Poly(ADP-Ribose) Glycohydrolase Mediates Chromatin Structure and SIR2-Dependent Silencing. *Genetics*, 172(1), 363-371.
- [12] Thomas, C. & Tulin, A. V(2013). Poly-ADP-Ribose Polymerase: Machinery for Nuclear Processes. *Molecular Aspects of Medicine* (In press).
- [13] Ji, Y. & Tulin, A. V. (2010). The roles of PARP1 in gene control and cell differentiation. *Current Opinion in Genetics & Development*. 20(5), 512-518.
- [14] Gagne, J. P., et al. (2003). A proteomic approach to the identification of heterogeneous nuclear ribonucleoproteins as a new family of poly(ADP-ribose)-binding proteins. *Biochemical Journal*, 371, 331-340.
- [15] Malanga, M., et al. (2008). Poly(ADP-ribose) binds to the splicing factor ASF/SF2 and regulates its phosphorylation by DNA topoisomerase I. *J Biol Chem*, 283, 19991 - 19998.

- [16] Ji, Y. & Tulin, A. V. (2009). Poly(ADP-ribosyl)ation of heterogeneous nuclear ribonucleoproteins modulates splicing. *Nucl. Acids Res.*, 37(11), 3501-3513.
- [17] Pleschke, J. M., et al. (2000). Poly(ADP-ribose) Binds to Specific Domains in DNA Damage Checkpoint Proteins. *Journal of Biological Chemistry*, 275(52), 40974-40980.
- [18] Pinnola, A., et al. (2007). Nucleosomal Core Histones Mediate Dynamic Regulation of Poly(ADP-ribose) Polymerase 1 Protein Binding to Chromatin and Induction of Its Enzymatic Activity. *Journal of Biological Chemistry*, 282(44), 32511-32519.
- [19] Tulin, A. & Spradling, A. (2003). Chromatin loosening by poly(ADP)-ribose polymerase (PARP) at *Drosophila* puff loci. *Science*, 299(5606), 560 - 562.
- [20] Ji, Y. & Tulin, A. V. (2012). Poly(ADP-ribose) controls DE-cadherin-dependent stem cell maintenance and oocyte localization. *Nat Commun*, 3, 760.
- [21] Prasanth, K., et al. (2000). Omega speckles-a novel class of nuclear speckles containing hnRNPs associated with noncoding hsr-omega RNA in *Drosophila*. *Journal of Cell Science*, 113(19), 3485-3497.
- [22] Ji, Y. (2011). *Noncovalent pADPr Interaction with Proteins and Competition with RNA for Binding to Proteins*, in *Poly(ADP-ribose) Polymerase*, A.V. Tulin, Editor, Humana Press. 83-91.
- [23] Burd, C. G. & Dreyfuss, G. (1994). RNA binding specificity of hnRNP A1: significance of hnRNP A1 high-affinity binding sites in pre-mRNA splicing. *Embo Journal*, 13, 1197-1204.
- [24] Krietsch, J., et al. (2012) PARP activation regulates the RNA-binding protein NONO in the DNA damage response to DNA double-strand breaks. *Nucleic Acids Research*, 40(20), 10287-10301.
- [25] Leung, A. K., et al. (2011). Poly(ADP-Ribose) Regulates Stress Responses and MicroRNA Activity in the Cytoplasm. *Molecular Cell*, 42(4), 489-499.
- [26] Giammartino, D., et al. (2013). PARP1 Represses PAP and Inhibits Polyadenylation during Heat Shock. *Molecular Cell*, 49(1), 7-17.
- [27] Huelga, Stephanie C., et al. (2012). Integrative Genome-wide Analysis Reveals Cooperative Regulation of Alternative Splicing by hnRNP Proteins. *Cell Reports*, 1(2), 167-178.
- [28] Talukdar, I., et al. (2011). hnRNP A1 and hnRNP F Modulate the Alternative Splicing of Exon 11 of the Insulin Receptor Gene. *PLoS ONE*, 6(11), e27869.
- [29] Vu, N.T., et al. (2013). hnRNP U Enhances Caspase-9 Splicing and Is Modulated by AKT-dependent Phosphorylation of hnRNP L. *Journal of Biological Chemistry*, 288(12), 8575-8584.
- [30] Martinez-Contreras, R., et al. (2006). Intronic Binding Sites for hnRNP A/B and hnRNP F/H Proteins Stimulate Pre-mRNA Splicing. *PLoS Biol*, 4(2), e21.
- [31] Gagne, J., et al. (2008). Proteome-wide identification of poly(ADP-ribose) binding proteins and poly(ADP-ribose)-associated protein complexes. *Nucleic Acids Res*, 36, 6959 - 6976.
- [32] Isabelle, M., et al. (2010). Investigation of PARP-1, PARP-2, and PARG interactomes by affinity-purification mass spectrometry. *Proteome Science*, 8(1), 22.
- [33] Zhou, Z. & Fu, X.-D. (2013). Regulation of splicing by SR proteins and SR protein-specific kinases. *Chromosoma*, 122(3), 191-207.

- [34] Soret, J., et al. (2003). Altered Serine/Arginine-Rich Protein Phosphorylation and Exonic Enhancer-Dependent Splicing in Mammalian Cells Lacking Topoisomerase I. *Cancer Research*, 63(23), 8203-8211.
- [35] Carrillo Oesterreich, F., Preibisch, S. & Neugebauer, K. M. (2010). Global Analysis of Nascent RNA Reveals Transcriptional Pausing in Terminal Exons. *Molecular Cell*, 40(4), 571-581.
- [36] Khodor, Y. L., et al. (2012). Nascent-seq indicates widespread cotranscriptional pre-mRNA splicing in Drosophila. *Genes & Development*, 25(23), 2502-2512.
- [37] Girard, C., et al. (2012). Post-transcriptional spliceosomes are retained in nuclear speckles until splicing completion. *Nat Commun.*, 3, 994.
- [38] Luco, R. F., et al. (2011). Epigenetics in Alternative Pre-mRNA Splicing. *Cell*, 144(1), 16-26.
- [39] Bannister, A. J. & Kouzarides, T. (2011). Regulation of chromatin by histone modifications. *Cell Research*, 21(3), 381-395.
- [40] Marks, H., et al. (2009). High-resolution analysis of epigenetic changes associated with X inactivation. *Genome Research*, 19(8), 1361-1373.
- [41] Kolasinska-Zwierz, P. (2009). Differential chromatin marking of introns and expressed exons by H3K36me3. *Nature Genet.*, 41, 376-381.
- [42] Tilgner, H., et al. (2009). Nucleosome positioning as a determinant of exon recognition. *Nat Struct Mol Biol*, 16(9), 996-1001.
- [43] Luco, R. F., et al. (2010). Regulation of Alternative Splicing by Histone Modifications. *Science*, 327(5968), 996-1000.
- [44] Sims III, R. J., et al. (2007). Recognition of trimethylated histone H3 lysine 4 facilitates the recruitment of transcription postinitiation factors and pre-mRNA splicing. *Molecular Cell*, 28(4), 665-676.
- [45] Saint-André, V., et al. (2011). Histone H3 lysine 9 trimethylation and HP1 γ favor inclusion of alternative exons. *Nature structural & molecular biology*. 18(3), 337-344.
- [46] Pradeepa, M. M., et al. (2012). Psip1/Ledgf p52 binds methylated histone H3K36 and splicing factors and contributes to the regulation of alternative splicing. *PLoS Genetics*, 8(5), e1002717.
- [47] Piacentini, L., et al. (2009). Heterochromatin protein 1 (HP1a) positively regulates euchromatic gene expression through RNA transcript association and interaction with hnRNPs in Drosophila. *PLoS Genetics*, 5(10), e1000670.
- [48] Zhou, H.-L., et al. (2011). Hu proteins regulate alternative splicing by inducing localized histone hyperacetylation in an RNA-dependent manner. *Proceedings of the National Academy of Sciences*. 108(36), E627-E635.
- [49] Tulin, A., Stewart, D. & Spradling, A. C. (2002). The Drosophila heterochromatic gene encoding poly(ADP-ribose) polymerase (PARP) is required to modulate chromatin structure during development. *Genes & Development*, 16(16), 2108-2119.
- [50] Kotova, E., Jarnik, M. & Tulin, A. V. (2011). Uncoupling of the transactivation and transrepression functions of PARP1 protein. *Proceedings of the National Academy of Sciences*, 107(14), 6406-6411.
- [51] Clark, N. J., et al. (2012). Alternative Modes of Binding of Poly(ADP-ribose) Polymerase 1 to Free DNA and Nucleosomes. *Journal of Biological Chemistry*, 287(39), 32430-32439.

- [52] Kotova, E., et al. (2011). Drosophila histone H2A variant (H2Av) controls poly (ADP-ribose) polymerase 1 (PARP1) activation in chromatin. *Proceedings of the National Academy of Sciences*, 108(15), 6205-6210.
- [53] Petesch, S. J. & Lis, J. T. (2012). Activator-Induced Spread of Poly (ADP-Ribose) Polymerase Promotes Nucleosome Loss a *Hsp70*. *Molecular Cell*, 45(1), 64-74.
- [54] Messner, S. & Hottiger, M. O. (2011). Histone ADP-ribosylation in DNA repair, replication and transcription. *Trends in Cell Biology*, 21(9), 534-542.
- [55] Martinez-Zamudio, R. & Ha, H. C. (2012). Histone ADP-ribosylation facilitates gene transcription by directly remodeling nucleosomes. *Molecular and cellular biology*, 32(13), 2490-2502.
- [56] Krishnakumar, R. & Kraus, W. L. (2010). PARP-1 regulates chromatin structure and transcription through a KDM5B-dependent pathway. *Molecular Cell*, 39(5), 736-749.
- [57] Osada, T., Rydén, A.-M. & Masutani, M. (2013). Poly(ADP-ribosylation) regulates chromatin organization through histone H3 modification and DNA methylation of the first cell cycle of mouse embryos. *Biochemical and Biophysical Research Communications*, 434(1), 15-21.
- [58] Ahel, D., et al. (2009). Poly(ADP-ribose)-Dependent Regulation of DNA Repair by the Chromatin Remodeling Enzyme ALC1. *Science*, 325(5945), 1240-1243.
- [59] Murawska, M., et al. (2011). Stress-induced PARP activation mediates recruitment of Drosophila Mi-2 to promote heat shock gene expression. *PLoS Genetics*, 7(7), e1002206.
- [60] LeMaire, M. F. & Thummel, C. S. (1990). Splicing precedes polyadenylation during Drosophila E74A transcription. *Molecular and cellular biology*, 10(11), 6059-6063.
- [61] Badenhorst, P., et al. (2005). The Drosophila nucleosome remodeling factor NURF is required for Ecdysteroid signaling and metamorphosis. *Genes & Development*, 19(21), 2540-2545.
- [62] Kellner, W. A., et al. (2012). Genome-wide phosphoacetylation of histone H3 at Drosophila enhancers and promoters. *Genome Research*, 22(6), 1081-1088.

Chapter 28

PATENT ROADMAP FOR THE BIOSENSOR SPACE

Mohamed C. Azeez^{*,†}, Unisha Patel[†] and Dennis Fernandez

Fernandez & Associates, LLP, Atherton, CA, US

ABSTRACT

Genetic and behavioral data mining will revolutionize health/lifestyle delivery and outcomes, in much the same way the Internet delivered meaningful social outcomes by providing the infrastructure for information to be collected and connections to be made in ways that had not been possible hitherto. Genetic databases and lifestyle correlations via the cutting edge field of bioinformatics, wearable technology and personalized medicine will be the next echelon of information to be reconnoitered and used through the Internet to enrich our lives. Thanks to smart phones and wearable technology, it will all commence from the palm of our hands. In this review, we will spotlight the emerging fields of bioinformatics, personalized medicine and biosensor technology: historical insights, current issues, and future trends. In particular, we will embark on an in-depth look into the sub-fields of personalized medicine and wearable systems for health and lifestyle management. The health-sector, being information intensive, will exploit IT-led platforms that capture personalized health data in real-time and have connectivity with a cloud backbone to intelligently process the information to deliver real-time analytics and actions. Various roadmaps are presented offering aggressive offensive and defensive IP strategies and solutions to companies in the bioinformatics and biosensor space. How does current patent law and policy in the wake of the AIA reforms and Supreme Court decisions in *Alice* and *Mayo* impact on this field and on the strategic guidepost? What's more, we will reopen the patent troll reform debate: will it create a sweeping sea change, or be just another talking point. How do the sweeping provisions of Obamacare touch upon the field of wearable health systems and bioinformatics? Moreover, how can we reconcile proprietary value with the growing trends of the open-source movement and the big health data initiatives that lie at the intersection of private and non-profit sectors? Finally, all of this big health data begs an inquiry into issues of privacy and a host of other ethical concerns. With the right strategies in their IP toolkit, bioinformatics and wearable startup companies will not only bolster their current patent position, but also be able to leverage it into a

* Corresponding Author's Emails: azeez.iploft@gmail.com; unisha.iploft@gmail.com.

† Associates have contributed equally.

significant competitive advantage. More importantly, the broader public will be able to avail themselves of the utility of these flash-bulb popping innovations that promise to capture personalized health data in real-time to deliver targeted and improved health outcomes.

INTRODUCTION

“Imagination is more important than knowledge” - Albert Einstein

Albert Einstein’s proclivity for imagination over knowledge is an initial point, because science, medicine and Intellectual Property (IP) are based on the power of imagination. Einstein understood that it is the ability to “stand on an existing foundation of accepted knowledge, and see beyond to the next frontier of discovery that is the source of personal, cultural and economic advancement. [1] The World Wide Web in the past two decades has opened our eyes to a plentitude of information and thus, molded the present. Likewise, in the next two decades the avant-garde fields of wearable technology, bioinformatics and personalized medicine will be the next stratum of information to be explored and used through the internet which will change our lives and help us make healthier and smarter lifestyle decisions.

Bioinformatics has been defined as “the application of computing power to biological data to reveal new patterns and information below the surface of that data.” [2] The patterns emerging from this treasure trove of data have enabled a vast array of genomic and proteomic tools, such as high-throughput RNAseq-expression profiling, next generation sequencing for personalized medicine and mass-spectrometry-facilitated protein profiling, to single out sequence variants across an entire genome and global expression. [3] With this tool in the biomedical investigation toolkit, researchers are able to pinpoint the molecular variants underlying disease states, and use these leads for developing novel targets in the cure and prevention of the disease. Some tools include enabling data mining, visualization, sequence alignment, pattern recognition, molecular modeling, and predictive tools. What happens when these bioinformatics tools leave the confines of a lab setting and become integrated into ubiquitous consumer devices? A powerful array of dynamic provisioning will now be at our disposal- literally from the palm of our hands.

In the new economy, a brand's competitive advantage will be determined by how effectively its team can build a strong Intellectual Property portfolio to protect a robust suite of products. The rise in proprietary value around bioinformatics and wearable device technology has led to unprecedented investments, underscoring the strategic value of Intellectual Property. A recent report from a New York based database research firm has reported an investment of \$1.4 billion in the wearable technology sector. [4] Moreover, venture firms have sunk more dollars into gene therapies and genomics-based molecular diagnostics since 2010, \$715.8 million, than they did all of last decade, \$653.6 million, according to venture source. [5] The US Dept. of Health and Human Services is predicting the market for personalized medicine will reach \$300B by 2020; it's no surprise that many of the same analysts feel the sector is now entering into the inflection phase of the Gartner Curve for what some are calling "the internet of healthcare." Exploiting IT platforms in order to mine and curate across a broad spectrum of health data will be pivotal.

However, as in any emerging technology, the concept-to-commercialization pipeline is fraught with peril: from Supreme Court decisions, sweeping legislative reforms, and the all-enveloping patent troll pandemic. Buried deep under hundreds of pages of amorphous codes and guidelines, lays a strategic guidepost. This review will elucidate a roadmap that offers a variety of offensive and defensive patent strategy for this emerging technology space. With these strategies in their IP toolkit, bioinformatics and wearable companies will be able to circumvent the legal pitfalls, and grab the high-hanging fruits of innovation that are sprouting in this sector.

PERSONALIZED MEDICINE

“It's far more important to know what person the disease has than what disease the person has” -Hippocrates

As individuals, we all have a stake in personalized medicine. Today, we are at a genesis of a new epoch of personalized medicine. The medical community has long recognized the inherent uniqueness of patients as evidenced by the prevalence of specific disease entities within families and ethnicities, variable responses to medications, and diverse manifestations of a single pathology. [6] Nonetheless, the scientific community and the advances in medical therapy until the turn of the 21st century have generally employed a broad treatment methodology to a heterogeneous population rather than a unique treatment approach to an individual patient. This practice is changing very rapidly because of the completion of the \$3.1 billion Human Genome project, which has heightened the scientific community's understanding of the individual, thus enabling the practitioners to treat individuals and patients based on their personal and unique characteristics- a practice known as Personalized Medicine (PM).

History of Personalized Medicine

The concept of personalized medicine dates back to the late 18th century when Mendel discovered the hereditary traits. But, it was not until the middle of the 19th century, when the developments in chemistry and microscopy allowed scientists to begin to understand the underlying causes of disease. [2] Since then, major developments in science and technology have allowed healthcare decisions to become increasingly granular over time. [7] With the growth of the pharmaceutical and medical device industries in the 20th century, came the rise of genetics, imaging, and data mining. Midway through the century, observations of individual differences in response to drugs gave rise to a body of research focused on identifying key enzymes that play a role in variation in drug metabolism and response and that served as the foundation for pharmacogenomics. [7] More recently, the sequencing of the human genome at the turn of the 21st century set in motion the transformation of personalized medicine from an idea to a practice. [7] The mapping of the human genome has shown that while mankind's genetic make-up is 99.1% identical; a small 0.9% inter-individual genetic variability creates and accounts for the vast variability that exists within the human species. [6] The rapidly

growing arena of personalized medicine will now enable the scientific community to focus on the 0.9% inter-individual genetic variability and create precise and specific therapies and drugs to individuals rather than a population. Our understanding of genetics and the genome has been greatly enhanced by the overwhelming revolution in sequencing technology in the recent years. [8] Next generation sequencing (NGS) has been widely implemented for whole-genome sequencing, whole-exome sequencing, transcriptome sequencing, targeted region sequencing, epigenetic sequencing, and other sequencing. [9] NGS has significant clinical potential in studying and identifying novel cancer mutations as well as understanding hereditary cancer. Additionally, another important and noteworthy implication of NGS is to advance personalized cancer treatment. For example, NGS has been used in the detection of epidermal growth factor receptor (EGFR) deletions in non-small cell lung cancer, which showed important pathogenetic and clinical implications for patients with non-small cell lung cancer. [10] In an inspiring study, genetics researchers at Washington University implemented a whole-genome and transcriptome sequencing for a researcher in their team, who had adult acute lymphoblastic leukemia. [11] The cancer relapsed twice in 10 years from the time of first diagnosis. It is at this point his colleagues found a significant clue about the disease through next generation RNA sequencing. Their results showed that a normal gene, *FLT3* was wildly active in the leukemia cells. Fortunately, the drug Sunitinib, which is approved to treat advanced stage renal cancer, acts as a *FLT3* blockade. With Sunitinib treatment, his blood counts appeared to be normal and the cancer was in remission. These developments in genomics, together with advances in a number of other areas, such as stem cell therapy, personalized medicine, bioinformatics, medical imaging, and more recently wearable technology, are creating the possibility for scientists to develop tools to truly personalize diagnosis and treatment.

Landscape of Personalized Medicine

“An Ounce of Prevention Is Worth a Pound of Cure” -Benjamin Franklin

National Institute of Science (NSA) describes ‘personalized medicine’ as, “the use of genomic, epigenomic, exposure and other data to define individual patterns of disease, potentially leading to better individual treatment.” [12] Personalized medicine can be defined broadly as a model of healthcare that is predictive, personalized, preventive and participatory. [13] Moreover, PM can also be referred to as the tailoring of medical treatment to the individual characteristics, needs and preferences of a patient during all stages of care, including prevention, diagnosis, treatment and follow-up. [7]

Personalized medicine generally involves the use of two medical products – typically, a diagnostic device and a therapeutic product – to improve patient outcomes. A diagnostic device is a type of medical device. Diagnostic devices include both *in vitro* tests such as assays used in measurement of genetic factors and *in vivo* tests, such as electroencephalography (EEG), electrocardiography (EKG), or diagnostic imaging equipment. [7] Many medical device therapies are now capable of being tailored to specific patient characteristics. These individual characteristics include patient anatomy (e.g., size), physiology (e.g., nervous and cardiovascular systems, metabolism, reproduction) and environment of use (e.g., intensive care unit, home use). [7] Additionally, physiological sensors can be used to predict treatment responses for individual patients. For example, 3D printing has been used to create personalized

medical devices based on imaging of a patient's anatomy. [7] In addition, the advent of mobile and wireless capability, better sensors, interoperable devices, and the Internet have led to technologies that allow for more effective patient monitoring and treatment outside the traditional medical care setting of hospitals, chronic care facilities and physician offices, instead, more people are treated at home and at work and are better able to maintain their lifestyle and quality of life. [7] As a result of these technological advances, medical diagnostics and therapeutics can be more finely tuned to better meet the needs of individual patients. [7]

In the long run, personalized medicine seeks to reduce the burden of disease by targeting prevention and treatment more effectively. [7] With the help of personalized medicine, the health care management paradigm will focus on prevention, moving from illness to wellness, and from treating disease to maintaining health. [7] By improving our ability to predict and account for individual differences in disease diagnosis, experience, and therapy response, personalized medicine offers hope for diminishing the duration and severity of illness, shortening product development timelines, and improving success rates. [7] At the same time, it may reduce healthcare costs by improving our ability to quickly and reliably select effective therapy for a given patient while minimizing costs associated with ineffective treatment and avoidable adverse events. [7]

The Exigent Need for Personalized Medicine

“Declare the Past, Diagnose the Present, Foretell the Future” -Hippocrates

Despite the major advances in the medical and scientific field, we still are a long way away from understanding of exactly how and why different individuals respond to the similar treatments in a different way. One of the major shortcomings is the ‘One size fits all’ mentality that we are currently converging on. ‘One size fits all’ and ‘Evidence-based medicine’ approach rarely works- be it in clothing, shoes or medicine. For example, a patient being treated for high blood pressure or diabetes might be placed on one of a number of blood pressure or diabetes medications available in the market. The patient's doctor makes a decision about what medication to prescribe based on only general-population- evidence-centric-based information about what might actually work for that particular patient. If the prescribed medication does not work after a few weeks, the patient might be switched to another medication. This somewhat “trial-and-error” approach can lead to patient dissatisfaction, adverse and antagonist drug responses and drug interactions and poor adherence to treatment regimens. [7] Figure 1, illustrates an alternative to the “one-size-fits-all” approach to treatment, whereby treatment is personalized and tailored to complement the genetic and metabolic profiles of a patient.

Evidence-based medicine may provide transitory savings in the short term, but the same patient who takes the cheapest available statin today may very well be the patient costing us—the taxpayer, the policymaker, the thought-leader, the sister, the spouse -- big bucks when that patient ends up in the hospital because of improperly treated cardiovascular disease. [14] The repercussions of choosing short-term thinking over long-term results and cost-based medicine over patient-based are pernicious to both the public purse and the public health. [14] Thus, quoting Mark McClellan, the chief architect of the American health care policy, “Looking at a gigantic uniform solution for everything is never going to work”, we must advance ahead to the 21st century of patient-centric and cost-efficient based medicine, where we can situate our

emphasis on treating patients via the 'Dose selection based on genotyping and phenotyping of an individual' approach. [14]

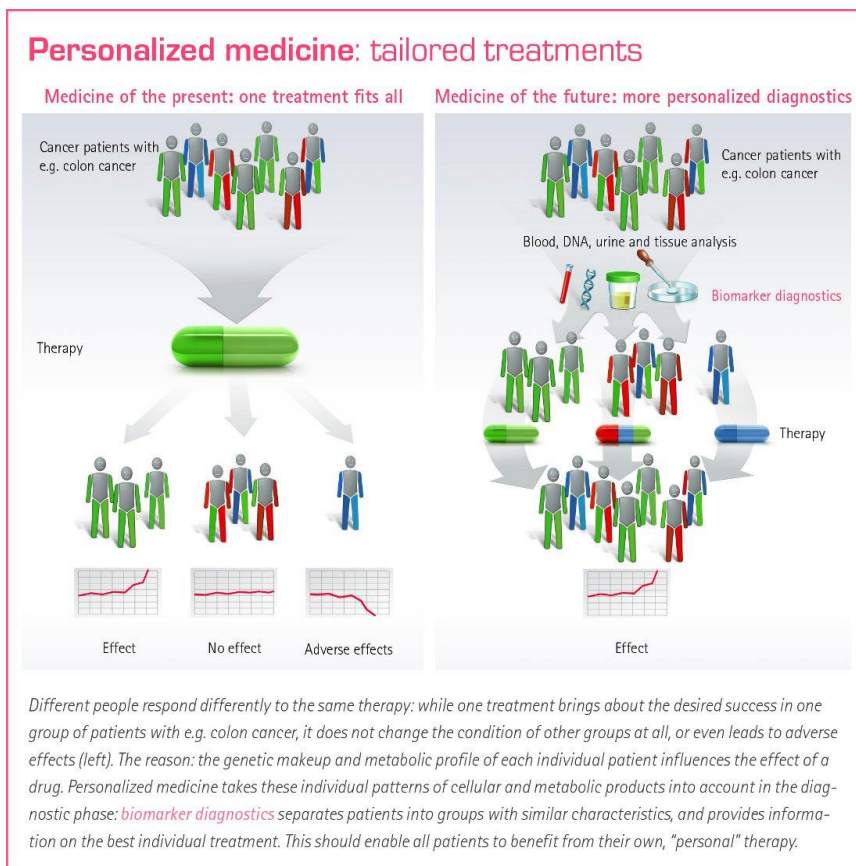


Figure 1. Illustrates the effects of genetic and metabolic profiles on pharmacodynamics and the value of developing a personalized and tailored approach to treatment. ("Personalized Medicine." Bayer HealthCare Pharmaceuticals RSS. Web. <<http://www.bayerpharma.com/en/research-and-development/research-focus/oncology/personalized-medicine/index.php>>).

Today, with the rapid advancements in the field of genomics and bioinformatics navigating into the age of personalized medicine, people can be screened with a variety of molecular diagnostics to reduce the side effects, increase compliance and improve outcomes. Increasingly in the near future, clinical outcomes will be monitored more accurately using computers that can control thousands of variables to help the scientific community, researchers and doctors to identify precise treatments that matter the most to improve individual-patient care. [14] It has increasingly become obvious that we, as future citizens of this country need to approach the newfangled and novel era of personalized medicine by developing the studies, tools and metrics that will provide policymakers, thought-leaders and Congress with the ability to evaluate, support and use new medicines and medical technologies to improve patient well-being in a patient-based, cost-efficient environment.

We must provide the 21st century health care evidence so crucial to the decision process of policy-makers and the policy debate of thought leaders -- and must do so in a coordinated, rigorous manner via evidence-based policy approaches that will (i) recognize the value of medical innovation, (ii) support the healthcare provider and patient as an empowered technology assessors and finally, (iii) reflect the emerging science of personalized medicine, bioinformatics and wearable technology to develop state of the art approaches in documenting and improving patient well-being in a cost-efficient way. [14] It is imperative to recognize the importance of partnership between the state and federal levels along with health systems, hospitals and academic research facilities whose end goal is to achieve a healthcare system where, the 'Equality of care' is matched to the 'Quality of care'.

BIOINFORMATICS

"Everything is going to be connected to cloud and data... All of this will be mediated by software"- Satya Nadella

Bioinformatics has been defined as "the application of computing power to biological data to reveal new patterns and information below the surface of that data." [15] The recent and rapid advances in the past decade in genomics and next generation sequencing (NGS) is beginning to reveal a large number of possible new and rapid, low-cost ways to sequence the genome, thus creating an mammoth amount of genetic data. The patterns emerging from this treasure trove of data have enabled a vast array of powerful and dynamic provisioning to be at our disposal on demand- literally from the palm of our hands. The key is to accurately and efficiently transform this 'big data' to benefit individuals in a clinical setting. Bioinformatics will provide the hub and tools for data analysis, interpretation, and ultimately the translation of relevant data in personalizing clinical medicine. [17]

While this has been the common application for bioinformatics for the past decade and a half, this field has recently branched into lifestyle and well-being applications. Bioinformatics is no longer a place where just lab-based science meets information science, but now has become a place where real-time, on-demand, physiological data seamlessly integrates with information science, thus, translating the data into a comprehensive lifestyle and wellness program. For instance, BodySync's core Gene/Lifestyle Integration technology integrates and interprets nutrition, fitness, and lifestyle data into actionable diet and exercise information. [18] Health and lifestyle choices could be informed on a personalized and analytical level, the likes of which we have never seen before. This brand of medicine and healthcare, coined as P4 medicine by Leroy Hood, the famed geneticist, for being personalized, preventive, predictive, and participatory, will finally usher in cost-containment, efficiencies, and outcomes to a healthcare model that has largely been reactive and plagued by administrative inefficiencies and poor health outcomes. [19] The health-sector, being information intensive, will exploit IT-led platforms that capture personalized health data in real-time and have connectivity with a cloud backbone to intelligently process the information to deliver real-time analytics and action.

WEARABLE TECHNOLOGY

“Doctors can be replaced by software – 80% of them can. I’d much rather have good machine learning system diagnose my disease than the median or average doctor”- Vinod Khosla

Are wearables headed for humankind’s sixth sense? During his keynote speech at the most recent Consumer Electronics Show in Las Vegas, Intel CEO, Brian Krzanich, unveiled his mobile or wearable computing technology devices as varied from baby sensors that transmit bio-data to ear-buds that would enable runners to get detailed health information in real-time. According to Mr. Krzanich, smart wearable technology represents the newest iteration of the mobile market. [17] This new slice of the mobile market hopes to make the human-computer synergy more seamless, intuitive, and immersive.

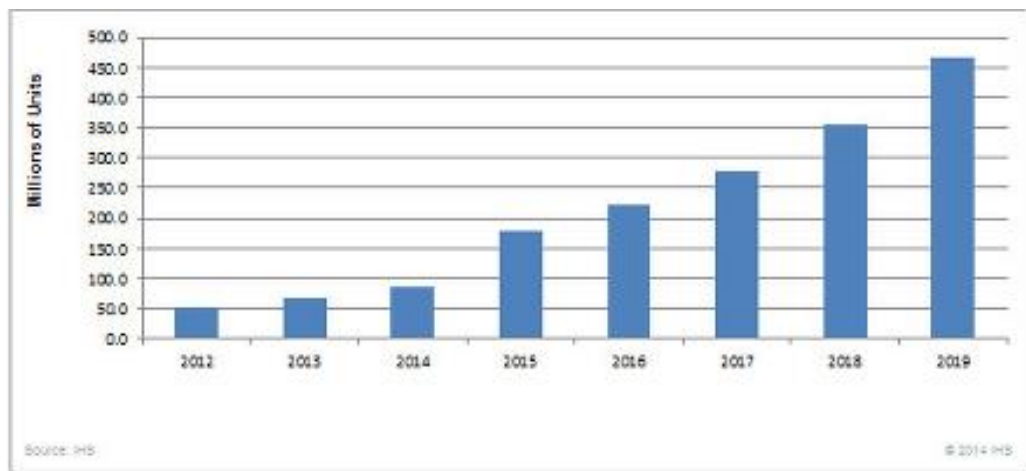


Figure 2. The number of sensor units to ship and market projections for the sensor market for wearables. (Clarke, Peter. "Sensors for Wearables Market to Double in 2015." EE Times (Europe), 23 Oct. 2014. Web. <http://www.analog-eetimes.com/en/sensors-for-wearables-market-to-double-in-2015.html?cmp_id=7&news_id=222906773>).

Wearable products span across five distinct applications, across approximately twenty-three product categories. Two categories are assessed for the purpose of this article and are defined as follows: (1) fitness/wellness wearables monitor and track activity and emotions; and (2) medical/healthcare wearables monitor vital signs and deliver therapeutics. [20] The wearable device market is expected to experience exponential growth over the next few years. Market analysis and forecast organization, HIS, thinks that the Apple Watch, will stimulate and set a standard for fitness and health monitoring features on wearable electronics devices. [21] HIS predicts a seven fold increase in the shipments of sensors used in wearable electronics. Additionally, HIS forecasts the market for wearable equipment will move from 50 million units in 2013 to 135 million units shipped in 2019. [22] Figure 2, represents the sensor market for wearables in particular- notice the doubling of units in 2015 from 2014, and then an upward trend into 2019. It is forecasted to grow to \$8.3 billion by 2018, with an 18% compound annual growth rate. Research analysts predict that total volume will increase from 35.7 to 134 million

units in the same time frame (30% CAGR). [23] Wearables are projected to have the same market trajectory as tablet devices: 21% of all U.S. adults use some form of wearable, which is exactly the same market penetration of tablets in 2012. Today, tablet penetration is 41%, leading many industry experts to surmise that wearable penetration may double in the next two years, as well. [24]

If venture capital interest is any indication of future growth, then we are all poised to see unprecedented growth in the near future. As of June of 2014, there has been an unprecedented \$2.3 billion worth of venture funding directed towards digital health start-ups, of which \$200 million went to health wearables. [25] Intel Capital, Intel's venture arm, has just pledged to invest \$355 million in 16 carefully selected tech-start-ups. [26]

Lifestyle Wearables

In the next decade, wearable technology is set to become the next big frontier in the personal electronics market. Leveraging integrated wearable devices for personalized health and fitness applications will rely on latent-free connectivity and processing as well as low-consumption and high-fidelity sensing. In the context of motion sensors, multi-axis accelerometers and gyroscopes are being integrated with wearables in order to deliver higher accuracy body movement detection versus traditional accelerometer-only wearable designs. [27] These 6 and 9-axis motion sensors are integrated with microcontroller subsystems with an extremely low memory and power footprint. Most of the more resource-taxing processes are shifted to the microprocessor unit (MPU), relieving the microcontrollers of this resource stress. This sensor system architecture allows for highly differentiated accuracy, while still having a small form factor and a small resource footprint. [28] This type of sensor deployment in health monitoring is fast becoming an emerging standard due to the advances in circuits, microcontrollers, front-end amplification, and wireless transmission. Due to the miniaturization, sensors are increasingly being integrated into garments, socks, wristbands, watches, stickers, glasses, etc. Batch fabrication techniques, wherein microprocessors and radio communication circuits are integrated into a single circuit, have brought about significant reductions in size and resources. Future developments in the field of flexible electronics, as well as material science, will be the touchstone of smaller, lighter wearable systems. [29] Figure 3, demonstrates the vast ubiquity in product design, form-factor, and provisioning, with respect to lifestyle and health wearables.

Wearable systems integrate a plurality of sensors into a sensor network by relying on the 802.15.4/ZigBee and Bluetooth. This standard, based on low-power and low-cost radio frequency technology, has an exceptionally high data rate. The most ubiquitous wearable sensor for health and wellness applications is the accelerometer. It can very discreetly track metabolic activity, gait parameters, and general activities of daily living. It measures the acceleration of objects in motion along three reference axes. It's a far more preferred method to measuring physical activity, compared to a pedometer, because the intensity of the movement is more precisely captured. There is also data suggesting that precision rates can be optimized for activity based on sensor positioning; suggesting just how precise the data gathering can be via the accelerometer sensor. [30] In addition to displacement and velocity, accelerometers can also sense for orientation, inclination, body posture, etc. Since the first batch fabricated MEMS accelerometer in 1979, many advances have occurred in MEMS technology, not the least of

Medical and Healthcare Wearables

Unsustainable trends including aging demographics, chronic disease, cost of care, and caregiver shortage are driving a global healthcare crisis and even threatening the global economy. [36] Biosensors representing the technological counterpart of living senses have found routine application like blood glucose measurement, drug development, and to some extent DNA chips for expression analysis and enzyme polymorphisms. [37] The use of wearable monitoring biosensors allows a constant monitoring of physiological signals which holds a prodigious promise in early detection, diagnosis and immediate treatment of various diseases. Wearable biosensors are an amalgamation of wearables and biosensors. Biosensors are analytical devices, which detect an analyte by combining a biological component with a physio-chemical detector. The biosensor consists of three major elements: (1) a bio-recognition component (e.g., microorganism, nucleic acids, enzymes, antibodies, cell organelles etc.), (2) a transducer/detector element which works in a physio-chemical way that transforms the signal resulting from the interaction of the analyte with the biological element into a signal which can be readily measured and quantified, and (3) an electronic system which includes a signal amplifier, processor and a display/reader device that displays the results in a comprehensible manner. [38] The graphics of figure 4, illustrates the workflow from sample to provisioning in a standard health wearable or embeddable, including the elements of a biosensor and transducer, converting the physiological sample into digital signals.

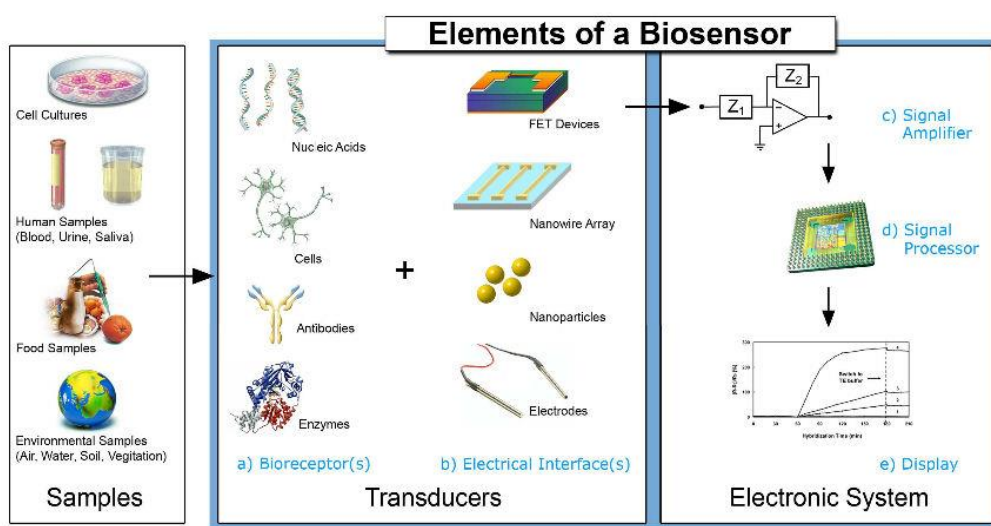


Figure 4. Elements of a biosensor in a standard health or medical-grade wearable or embeddable. ("Biosensor." *Wikipedia*. Wikimedia Foundation. Web. <<http://en.wikipedia.org/wiki/Biosensor>>.)

These wearable biosensors can be worn on the body in the form of smart shirts, watches, thin bandages or even tattoos. One such company based out of San Diego, Electrozyme, uses a disposable biosensor attached to a watch that measures an individuals' sweat and communicates with the wearer of the watch when to replenish lost electrolytes, rehydrate themselves or even take a break from their workout all in real-time where, the results are displayed on the individuals' smart phone. According, to its co-founder Jared Tangney, sweat has more than 800

biomarkers. Electrozyme combines the chemicals released in the sweat along with standard fitness wearables that analyzes just physical data to provide useful, actionable information to the individual in real-time. The biosensor strip which can be attached to the wearable watch is a low cost screen printed strip which can be produced for less than 10 cents a strip. Screen printing technology is a widely used technique for the fabrication of electrochemical sensors. [39] This methodology is likely to underpin the progressive drive towards miniaturized, sensitive and portable devices, and has already established its route from lab-to-market for a plethora of sensors. [39] The past decades have seen enormous progress in electro-analytical chemistry with the development of ultra-microelectrodes, tailored interfaces, molecular devices and smart sensors. These developments have resulted in substantial popularity of electro-analyses and to their expansion into new phases and environments. [40,41] As we enter the 21st century, we do not want to rely on cumbersome electrochemical cells and bulky electrodes, but rather would like to have fast, small, easy to use, portable, economical and disposable electrode systems. [39] The exploitation of new fabrication techniques allows the replacement of traditional beaker type electrochemical cells and bulky electrodes with easy to use sensors. [39] Fabrication of printed devices on bendable substrates has enabled the development of a wide range of new electrode systems. [39] Screen printing technology is a well-established technique for the fabrication of economical, portable and disposable electrode systems. [42,43] The whole electrode system, including reference, counter and working electrodes can be printed on the same substrate surface. [44] Figure 5, is a schematic of the configuration of a portable and disposable printable electrode system.

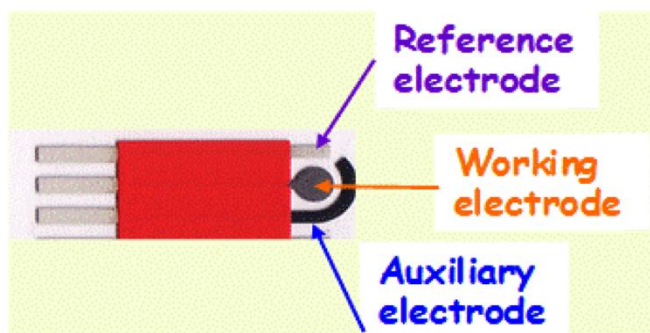


Figure 5. Electrode configuration using a printed screen fabrication on a bendable substrate. (Hayat, Akhtar, and Jean Louis Marty. "Disposable Screen Printed Electrochemical Sensors: Tools for Environmental Monitoring." *Sensors* 14.6 (2014): 10432-0453. Web.< <http://www.mdpi.com/1424-8220/14/6/10432/htm>>).

The adaptability of screen printed electrodes is of vital importance in the area of research. The ability to modify electrodes with ease through different inks commercially available for the reference, counter and working electrode, allows for highly specific and finally calibrated electrodes to be produced for specific target analytes. [44,45,46] Briefly, a screen printed electrode comprises a chemically inert substrate on which three electrodes, including working electrode, reference electrode and counter electrode, are printed through screen printing methodology. [39] The working electrode is the principal electrode on which electrochemical reactions are performed, while the reference electrode and counter electrode are used to

complete the electronic circuit. [39] The chemical or biological event on the screen printed electrode is converted into a detectable signal with the integration of a transducer element. [39] The fabrication of an electrochemical screen printed sensor usually involves three steps: fabrication of the screen printed electrode, surface design of the screen printed electrode and subsequently utilization for a sensing application. [39] The inks used in screen printed electrode fabrication consist of particles, polymeric binder and other additives for improved dispersion, printing and adhesion process. [39] The exact ink formulation and composition are patented by the respective companies, and are not disclosed to the users. [39]

The rate limiting force in achieving the small form factor of any commercially available wearable or embeddable is its power supply. Endless thought has gone to figuring out how to put low-bulk, low-latency batteries that are long lasting, energy renewing, into small form factor devices. Innovations in energy harnessing methods have enabled wearable technology to overcome this rate limitation. Novel power circuitry has done to wearables, what the smart phone did to the Internet of Things, or to the sharing economy. One such source, with low vitality, but good enough for wearable sensors, is the energy-harvesting method called Piezoelectric. Here, a small power supply converts ambient vibrations into electric current, which is then stored in a rechargeable battery. This is particularly fitting with fitness-tracking wearables that incorporate accelerometer sensors. [47] Thermoelectric is another novel approach, which exploits body heat by converting it into low-yield power source. Again, while suitable for wearables, it would not be ideal for ambient sensors. [48] However, the most promising technique appears to be Wi-Fi backscattering pioneered by researchers at the University of Washington. Researchers have figured out how to employ currently existing Wi-Fi infrastructure to not only provide internet connectivity- but just as importantly- to power devices battery-free. Ultra-low power tags communicate intelligently with Wi-Fi enabled devices leaving a very small power footprint. [49] The tags communicate by altering and reflecting ambient carrier waves that are freely bouncing around us. These reflected signals are subsequently encoded and then decoded by other devices in the network. [50] Presently, data rates and signal range is highly limited in the beta stage, but backscattering is positioned to one day potentially overcome the form-factor limitation previously described, and fill a long-awaited void. Thus, the combination and permutation of modern thermoelectric, electrochemical and biochemical systems along with screen-printing technology offers an introduction of powerful and potential unprecedented tools for an effective monitoring of various diseases and health in real-time.

Diagnostic Wearables: Real-time Medicine

As mobile technology penetrates into the daily facet of our lives, its extension into the new types of technologies and devices is inevitable. A patient self-diagnosing themselves is a very tangible reality thanks to the rapid advances in wearable technology, where the applications are infinite. These wearables encourage its wearers to be more affianced in their own fitness, help modify lifestyle habits by reminding them to exercise or take their medication, while providing a platform for patients and their doctors to share data in real-time and collaborate in making healthy strategic decisions.

Here, we present a few examples of how diagnostic wearables in the next decade will help instantaneously relay data and health information to our doctors, thus aiding them to diagnose better.

- **Smart Contacts:** Imagine managing glaucoma by simply wearing a contact lens. Sensimed, a Swiss company has developed a unique non-invasive soft contact lens-based solution, which revolutionizes glaucoma management by providing an automated recording of continuous ocular dimensional change over 24 hours. The data is received wirelessly by an antenna that is further, transmitted from the antenna to a portable recorder, which is connected to the practitioner's computer via a Bluetooth.
- **Personal Heart Rate Monitor:** The products available in the market today to measure heart rate are bulky, uncomfortable and largely inaccurate. Cardiio, is a simple application that converts the front camera of an iPhone/iPad to measure heart rate from a distance without even touching the camera. It measures the slight increase in blood volume that causes more light to be absorbed, and hence less light is reflected from your face. [51] Cardiio uses the camera to track these tiny changes in reflected light that are not visible to the human eye and calculate heartbeat. [51]
- **ECG-recording Device:** Atrial Fibrillation can occur without the patient having any symptoms. It could be frustrating for cardiologists to monitor patients once every few months during their routine checkups. AliveCor's heart rate monitor is an easy-to-use, FDA approved and accurate ECG-recording device that attaches to the back of a smartphone and wirelessly transmits data to the AliveECG app. The AliveECG app helps inform clinical decision-making through immediate detection of AF in ECGs and patient engagement by tracking medications, symptoms and lifestyle activities. [52]
- **Otoscope:** Ear infections are one of the leading reasons for a pediatric visit in the US. 90% of all children in any given country will develop an ear infection before the age of 7. Imagine if you could just page Dr. Mom, instead of a pediatrician in the middle of the night. Cellscope [53] developed an Oto, a smartphone-based otoscope which lets parents capture images of their child's inner ear and transmit them over a secure server to the doctor who, based on the images, can decide whether or not to call a patient in for treatment or prescribe over the phone.

Therapeutic Implantables/Embeddables: The Future of Personalized Medicine

Wearable products like the Fitbit, Pebble smartwatch and Samsung Galaxy Gear are already on the market, and Google's Android Wear promises to bring all the power of a smartphone to your wrist. [54] Even as wearable technology readies for its day in the sun, the next trend of implantables are already on the horizon. And it promises to move us closer to a sci-fi future than anything we've seen so far. [54] According to Transparency Market Research, the market for wearable medical devices is expected to reach \$5.8 billion globally growing at a CAGR of 16.4% from 2013 to 2019. [55] Implantable, also referred to as "embeddable", technology refers to a class of objects that can be inserted directly into the human body to modify, enhance or heal in ways that non-embedded devices cannot. [54]

Proteus Digital Health took wearables to unabridged next level. They have developed an ingestible biosensor in the form of a pill, which directly reads onto a smart phone. Each pill contains a one-square-millimetre sensor that is coated in two digestible metals: copper and magnesium. [56] These metals are not dangerous to consume because they currently exist in multi-vitamin supplements, as well as naturally in our diets. [56] Upon swallowing, the sensor is activated by electrolytes within the body. [56] The pill then transmits a signal to a small,

battery-powered patch worn on the user's torso and sends the data via Bluetooth to a caregiver's or family member's smartphone. [56]

Here, we present few illustrations of how mobile technology has penetrated into our lives to help us better understand our future by helping us make smarter choices through the use of implantables and embeddables.

- **Healing Chips:** Currently, patients are using cyber-implants directly connected to a smartphone application in order to monitor and treat diseases in real time. A new bionic pancreas being tested at Boston University, for instance, has a tiny sensor on an implantable needle that talks directly to a smartphone app to monitor blood-sugar levels for diabetics. [57]
- **Ingestible smart pills:** The most vital feature of implantables is that it not only communicates with the smart phone but, will also have a tête-à-tête with their doctors. The future of patient-doctor relationship is going to be in real-time. Proteus Digital's research team is developing a cyber-pill embedded with microprocessors and biosensors which is made entirely of ingredients found in food and is activated upon ingestion which can text doctors directly from inside of a body and relay information to your mobile device via a body-worn patch.
- **Implantable birth control pill:** The hunt for a perfect contraceptive has gone on for millennia. [58] MicroCHIPS and a laboratory at MIT collaborated, licensed and developed a device which can be implanted under the skin of a woman, where a hormone normally used in contraceptives is dispensed into the body which can be turned on and off via a remote control. It is designed to last for up to 16 years, which is nearly half of a woman's reproductive life and unlike many existing contraceptive implants; it can be deactivated without a trip to the clinic and an outpatient procedure.
- **Smart Dust:** Perhaps the most startling of current implantable innovations is Smart Dust. [57] Arrays of full computers with antennas, each much smaller than a grain of sand, that can organize themselves inside the body into as-needed networks to power a whole range of complex internal processes. [57] Imagine swarms of these nano-devices, called motes, attacking early cancer or bringing pain relief to a wound or even storing critical personal information in a manner that is deeply encrypted and hard to hack. [57] With smart dust, doctors will be able to act inside your body without opening you up, and information could be stored inside you, deeply encrypted, until you unlocked it from your very personal nano network. [57] Thus, the future of implantables/embeddables is the next big frontier.

The Cloud and Home Automation as the Next Frontier: Analytics and Provisioning

As the sensor readings become more multidimensional, dynamic, and non-linear, there becomes an even more practical need for an intelligent, proactive, and context-aware analytics and provisioning. All of the various devices currently on the market, such as Fitbit, Nike Fuelband etc., come equipped with companion apps that deliver targeted health and lifestyle data and recommendations. However, the real utility is in the deeper analytics that provide

personalized risk assessments for both acute and chronic conditions and actionable targets based on comparative population data sets. Even deeper value will come from integration of these data sets into health systems and processes; aggregated wearable data can be integrated with other pertinent health information across certain geography, demography, and diagnosis to render a health diagnosis with sharper resolution. [59] What's more, cross-device or device-agnostic analytical platforms will be strategic in integrating a whole host of different wearable apps with varying standards and actionable data. [60]

One such platform is from Vivametric. They are a device agnostic data analytics platform that ingests, standardizes and analyzes large information data sets from both wearables and biosensor devices using a data-as-a-service model. Standardizing and calibrating the disparate device and wearable population is at the core of Vivametric's novel analytical platform. The company launched a private beta in December 2014, which will feature a platform that incorporates the Bitfit, Google fit, Google Wear, Apple Healthkit, and a number of other disparate wearables. According to President Scott Valentine, this will give Vivametrics 90-95% of the market for accelerometry. [61] This data clearinghouse will enable Vivametrics to add a comparative layer to the analysis, enriching the analytics and consumer-end experience. The algorithms and analytical processes result in unquestionable accuracy: correlation coefficients of 0.89-0.99 with 'gold standard' ActiGraph sensors. [62] The Samsung SAMI (Samsung Architecture for Multimodal Interactions) is another agnostic and open platform that processes data from multiple sources via open API (Application Programming Interface). This platform is ideally configured for the Samsung Simband, which is being provided as open-source reference architecture for health/ lifestyle wearables. The open API model of analytics will enable other developers to layer SAMI-based fitness data with ancillary data, such as food diary app data or calendar app data. [63]

The standard system architecture comprises three major logical layers: storage, query, and analysis. The data storage layer consists of a non-relational, relational, ontology, cloud storage, and 3rd party databases. Data sets will communicate with different storage layers depending on the type of sensor and captured data. [64] Data points are curated by sensor reading data, data attachment index, tag mapping, and block mappings in order for it to be retrieved quickly. Triggers, with condition information, perform specific actions when conditions are satisfied. [65] The query and analysis layer consists of the workflow engine and ontology engine. The former is involved in the scheduling and execution of workflows and processes. The latter is involved in the annotation of data sets and providing the semantic substrate to manipulate the various formats of health data sources. Moreover, it is the ontology engine that enables integration and aggregation of different data formats. [66] Finally, the application layer consists of components that are involved in the management, access, and distribution of the data. The API service components offer an easy-to-use interface for access, management, visualization, etc. [67] Figure 6, illustrates a standard system architecture featuring the API service component in the application layer.

The API open service model is even more critical in the biosensor and medical implantables/embeddable space. The real-time predictive processing of a confluence of biological data from various sources and formats will spur the digital health revolution. Combining data streams in order to deduce meaningful patterns may only be possible in an open API analytical framework. The framework can be adapted from HealthData.gov API, used to provide software developers programmatic access to the contents of the public-wide health data catalog. The API allows for search and advanced query capabilities using API-enabled

datasets, as well enable developers to build a new data catalog tool. The HealthData.gov API uses the CKAN version 2.0beta. [68] The following is an example of a dataset that could be converted into an app by a developer using this API analytical framework: The TXT4Tots Message Library datasets provide recommended text for age-appropriate nutrition and physical activity reminders. Use the Data API to build an app that sends push notifications to wearable device users with targeted reminders appropriate to the age and activity of the wearable device user. This example illustrates how the API methodology can enhance health outcomes by leveraging web and cloud architecture to increase the efficiency of health data transfer and distribution. [69]

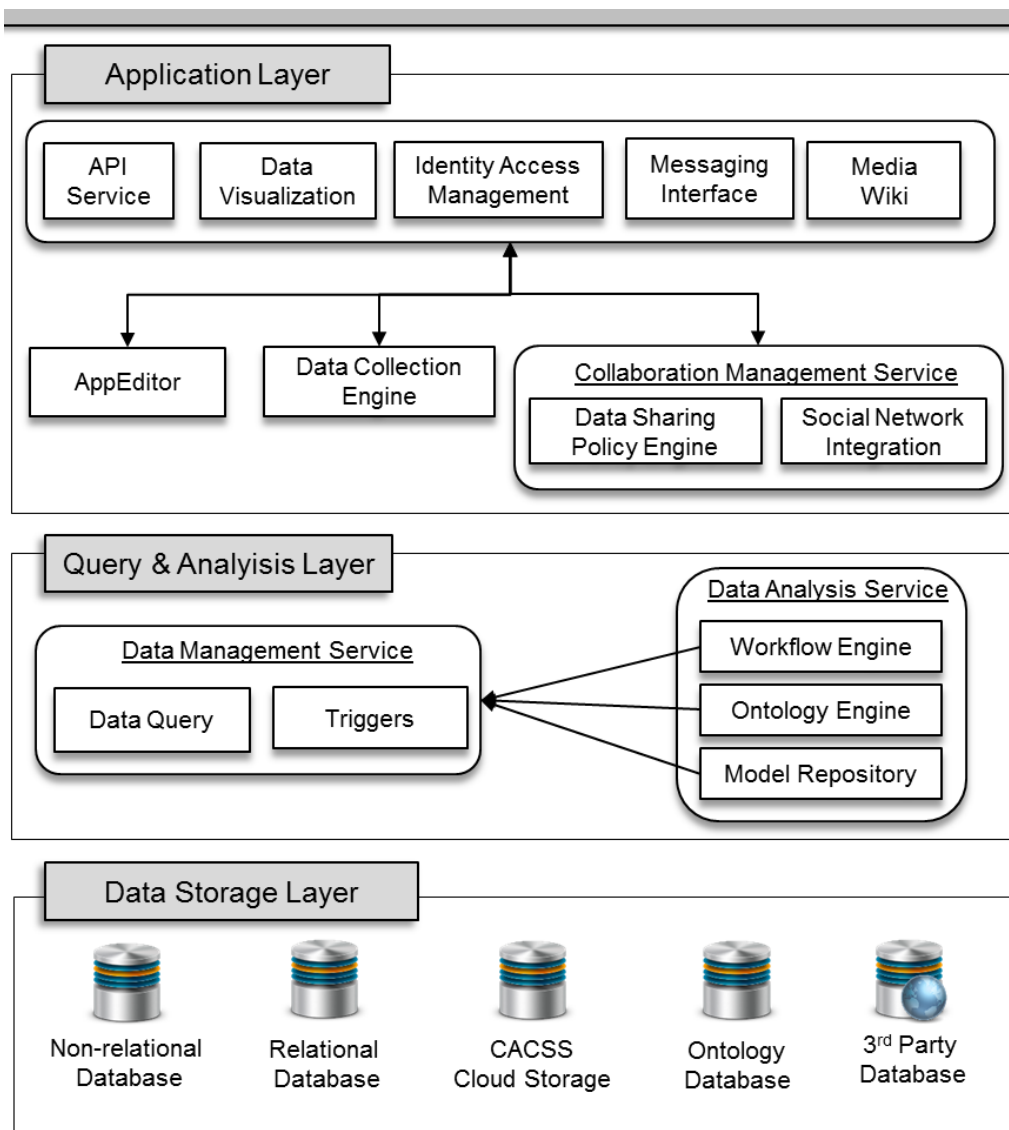


Figure 6. Standard system architecture featuring an API open service model for the biosensor space.

Gallileo Analytics co-founder, Anna McCollister-Slipp, has called on wearable/embeddable (device) manufacturers, hospital systems, and electronic health record companies to make the data accessible to patients and developers in order to provide a more thorough digital health ecosystem. “From a technological perspective, this problem is easy to fix. We could do it this week or this month, if we chose to do it. But so far, we have lacked the political will to force or the corporate will to choose to make personal data access a priority,” said Ms. McCollister-Slipp. [70] This digital health revolution is poised to not only deliver patient-centric positive outcomes, but is also touted to deliver substantial health care efficiencies, such as improving access and lowering costs.

INTERNET OF THINGS ECOSYSTEM

“As the Internet of things advances, the very notion of a clear dividing line between reality and virtual reality becomes blurred, sometimes in creative ways”- Geoff Mulgan

The above-mentioned companies range from wearables, Internet of Things, big data analytics, and smart devices. What do these companies all have in common? They are all cogs in the same machine: components and devices in an interoperable ecosystem with a cloud backbone involved in real-time sensing, processing, and provisioning.

Regardless of the type of sensor technology, none of this fine-tooth data capturing has any consumer-end value unless it is coupled to a cloud-based analytics and provisioning. Due to the large data sets and heterogeneous objects, in order for these sensor-laden devices to have any true appreciable value to the consumer, it must be seamlessly and intelligently linked with cloud-based provisioning across interoperable access layers/gateways. Unless the devices are coupled to a larger interoperable ecosystem, such as an Internet of Things (IoT) ecosystem, the provisioning capacity of the data will be limited in scope. While wearables are currently being deployed in industries as diverse as healthcare, entertainment & media and retail, the focus of this section will be on wearables that deliver health and lifestyle provisioning. Wearables in this space have the potential to deliver personalized diet and exercise mapping on a very granular level; deliver personalized health information, thereby removing the information asymmetries between health provider and patient and in addition, deliver on-demand, real-time diagnosis and tailored-therapeutics. This analytics-enabled biosensor data can not only provide efficiencies and positive outcomes for the individual with the wearable or embeddable, but it will also provide healthcare organizations and providers with the customized analytics to deliver a tailored-approach to treatment. Insurers can use the captured data to better manage healthcare and to lower overall healthcare costs. Finally, the analytics will provide clinicians and researchers with robust, real-time safety and efficacy information during late-stage clinical trials. [71]

The IoT ecosystem refers to the interconnection of uniquely identifiable embedded computing-like devices within the existing Internet infrastructure across a wide variety of protocols, browsers, and applications. This emerging technology converged with communication technology, such as RFID, NFC, etc., along with the huge increase in Internet

Protocol address space with the advent of IPv6 and the vast storage space of the cloud. Even by conservative estimates, it is estimated that 30 billion devices will be wirelessly connected to the Internet of Things by 2020. What's more, by 2030, the average person will come into contact with 3,000-5,000 IoT objects on a day-to-day basis, [72] the lion share of which will be wearables and embeddables.

Challenges

These wearables/emdeddables will require a network architecture that is adaptable to the massive scale and dynamic nature of the Internet of Things—a planetary-scale network infrastructure. On a high-level, it must be able to mine and analyze the vast swaths of data in real-time. Conversely, on a granular level, it must be able to sense and actuate the smallest of devices with the utmost discrimination. Additional limitations in the art include a lack of standards and interoperable technologies, data mining and analytics in real-time, securing such data from unauthorized use and attacks, and finally, network inefficiencies. Novel programming, content-delivery, and network management approaches are very much needed to catch up to the fast emerging ecosystem of IoT. In terms of network management, current architecture found in the art is designed for small-scale, closed-loop networks, which are unable to communicate across these networks in real-time. These access and latency issues effectively impair the execution of the distributed tasks of sensing, computing, and actuating. There exists an essential need for novel software-defined network architectures to effectively and efficiently deliver real-time, contextually based IoT services.

As the ubiquity of network-enabled wearables grow, consumer demand for IoT products and services will surge. All of the current heterogeneous networking infrastructures will have to be integrated into a clean, seamless networking platform. Another limitation facing wearables and the IoT is its vulnerability to cyber-attacks. This is especially a concern in the area of health wearables, where big health data in relation to individuals may be compromised and trigger an assortment of health privacy issues. New security controls of legacy security processes and technology will be needed. Besides the stop-gap patches deployed from the various manufacturers, a comprehensive security life-cycle- that is contextually-aware and behaviorally-adaptive- will need to be implemented in wearable-coupled IoT networks. These new approaches center around the way security in an Internet of Things network is architected, delivered, and self-monitored.

Finally, bottlenecks and latency represent the final hurdle. Currently, in order to accommodate newly introduced objects, the network elements need software upgrades in order to deliver dynamic provisioning. Even firmware upgrades of existing objects may require a matching software change on the network elements. Such a requirement presents tremendous bottlenecks when introducing new objects or object firmware. A dynamic and elastic IoT infrastructure is necessary to circumvent these data bottlenecks. It is expected that there will be three Internet Protocol-connected devices per capita and network traffic will reach 17 GB per capita by 2018. Global IP traffic will grow at a compound annual growth rate of 21% from 2013-2018, as illustrated by figure 7. Over half of all of this IP traffic will originate with non-PC devices, such as machine-to-machine (M2M) or wearables. [73] Cloud-centric, predictive and automated provisioning and deprovisioning of IoT-linked wearables will be the key to the seamless fusion between the physical and digital world.

With the lack of uniformity in standards across various frequency bands and communication protocols, adequate wireless connectivity is vital for Internet of Things network architecture, and provisioning ability of the wearable. The wireless connectivity technologies that are predominant in the market are adequate for the purposes of the IoT network. The standard communication protocol utilizes a set of rules and standards to format and control data exchange. The two standard communication models are the OSI model and the TCP/IP stacks. Both, the model and the stack, are further differentiated into a number of informational layers, allowing for integration with scalable and interoperable networks. In the IP stack, Wi-Fi represents the link layer and bi-directionally converts bits of data, and provides data framing for wireless communication. The network layer routes data through the network and provides a unique IP address (IPv4, IPv6, etc.) to every heterogeneous object (wearables, thermostats, light switches, etc.) in the IoT network. The Transmission Control Protocol allows multiple applications to run on one device. Finally, the HTTP is the predominant application layer protocol and governs data flow and is still used to transfer web content over the Internet.

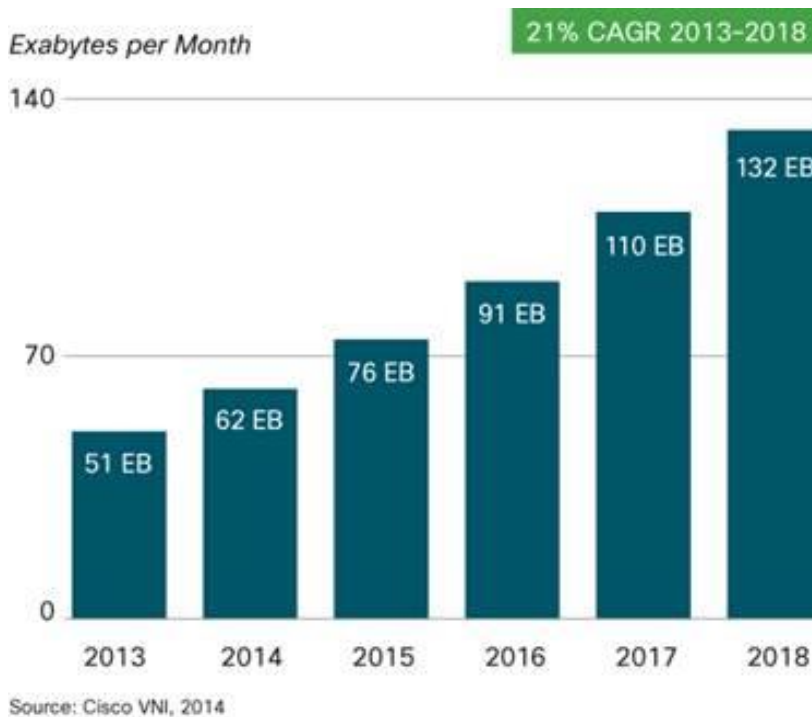


Figure 7. Cisco VNI Forecasts 132 Exabytes per Month of IP Traffic by 2018.

While a layered network implementation enables scale and interoperability, it also introduces more complexity. Internet gateways may be used to connect IP devices to the Internet by using a TCP/IP stack to restructure the simpler data flow from the local network. This enables faster and more flexible connectivity. Wearables may encompass the full range of network classes: PAN, LAN, NAN, and WAN. The standard PAN short-wave radio technology used to communicate an object with a hand-held device are Bluetooth, NFC, ZigBee, 6LoWPAN, Zwave, ANT, DECT ULE, etc. The standard LAN is the Wi-Fi based on the IEEE 802.11 standard. Until a few years ago, adding Wi-Fi connectivity to smaller, household object

would have been unthinkable; however, it is possible now due to the advances in silicone devices. Finally, the standard WANS used for the external network is 2G, 3G, 4G, and WiMax. Thus, internet gateway functionality between interoperable and distributive devices is, as imaginable, crucial to the real-time health provisioning of the wearable.

The Broad-Sweeping Implications

According to RockHealth, a stalwart in the fields of digital health accelerators and venture funding, has recently announced near \$1.9B raised for start-ups incorporating predictive analytics into their business model, from 2011-Q3 of 2014. [74] RockHealth has been instrumental in providing seed funding and access to a network of industry leaders for digital health start-ups, many of which are biosensor wearable companies that deploy predictive analytics. RockHealth claims that venture funding for biosensor wearables has increased five-fold from 2011-2013, [75] reflecting the consensus that these convergent technologies, bridging once very disparate sectors, has finally arrived to deliver life-altering outcomes. This personalized medicine revolution has begun to catalyze disruptive innovations across many different sectors of PM and an entire cadre of legal issues has emerged, from patents to privacy.

What are the implications of the new post-grant proceedings of the America Invents Act, which are being used to invalidate previously issued weak patents? The weakening of software patents by the recent Supreme Court decision in *Alice* will also have an undeniable impact on the biosensor/wearable patent landscape, especially with respect to the back-end. What about the fee-shifting provisions of the proposed troll reform legislation being mulled over by Congress? This could undoubtedly have an effect on the commercialization of digital health start-ups. Moreover, the aggregation and distribution of patient data triggers all sorts of privacy issues under HIPAA. On the other hand, Obamacare, as polarizing as it has been, could serve as a legislative accelerator for start-up digital health companies, for these digital health start-ups could be at the forefront of integrating personalized health data with hospitals and proxy systems. The transitioning of medical records into an integrated electronic system that could serve as a platform for a multi-disciplinary approach to healthcare is a legislative priority of the law, after all. What's more, will the patent troll reforms being mulled over in Congress, namely the fee-shifting provisions, have any impact on digital health landscape? In particular, will digital health start-ups be hesitant to go through seed-funding, research, and commercialization in lieu of looming, incendiary trolls? Not to mention, will digital health start-ups be hesitant to enforce their patents through litigation knowing that losers may be responsible for the legal fees of the winner under the fee-shifting provisions currently being debated? If patents are not enforced, will there be enough incentives for innovation? Finally, the groundswell of support for an open-source technology environment will be explored and whether or not this growing sentiment can be reconciled with a protective patent strategy. All of these potential hazards need to be accounted for in developing a strategy for navigating through the tricky biosensor/wearable landscape.

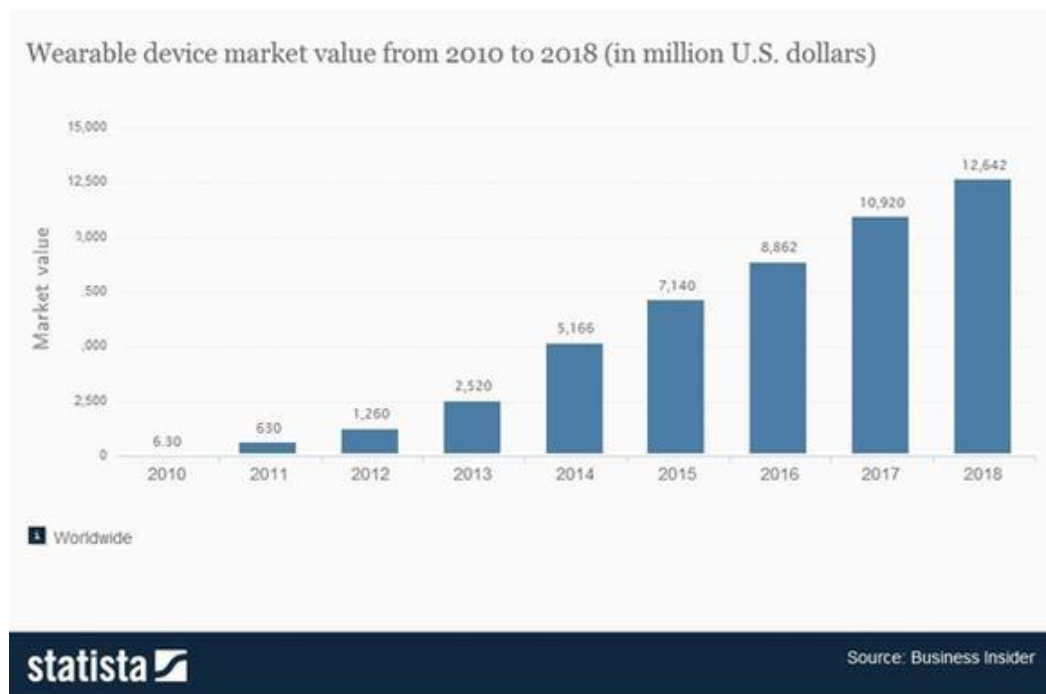


Figure 8. U.S. wearable device market.

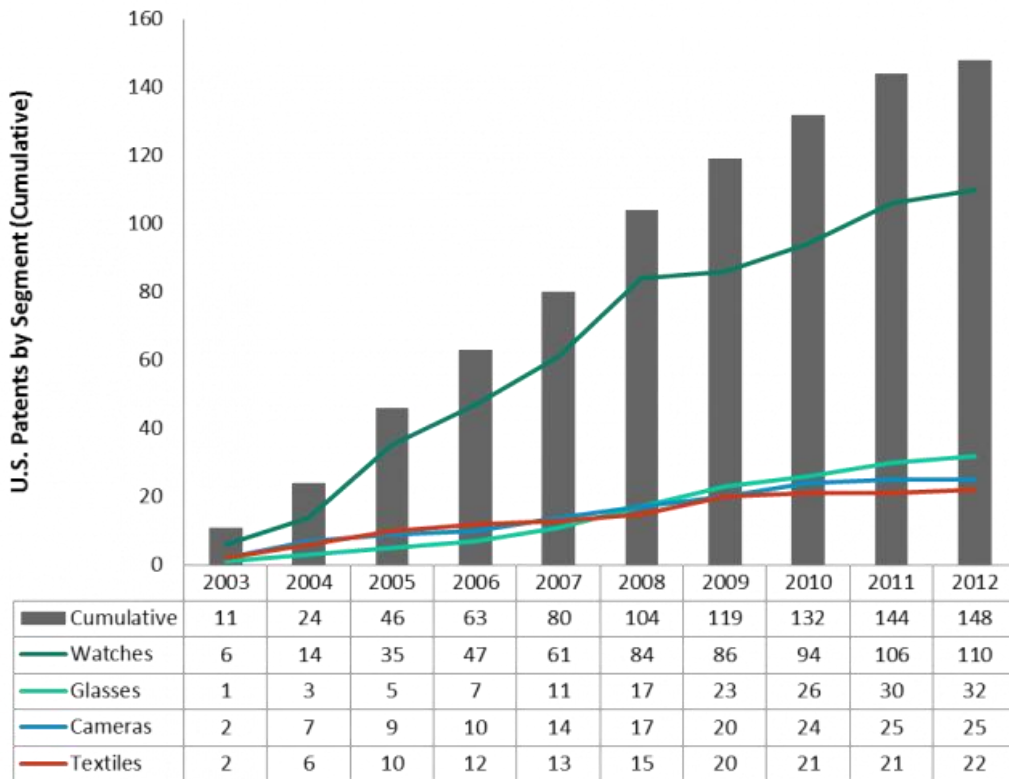
VALUE OF PATENTS

The market forecasts are mirrored very closely by the patent landscape, based on wearable device market data and U.S. patent issue data from the U.S. Patent & Trademark Office (Figures 8 and 9), suggesting the value of patents as drivers of innovation and market dominance. [76] Having a thorough understanding of the patent landscape can help identify emerging areas of innovation within the biosensor/wearable space, even prior to being deployed into consumer form factor products. Building a patent portfolio with broad, robust patents that cover emerging areas of innovation within the space that address fundamental needs is, crucial for all digital health start-ups as well as established bell-weather. In addition to launching products with innovational features that are legally protected with market exclusivity, these strategies may also enable a digital health company to engage in a lucrative licensing program. Finally, it can serve as a defensive initiative, whereby competitors are blocked from entering into these newly found niches in the overall space, or from preventing competitors from designing-around.

The current wearable tech market is dominated by a few devices, and by an even fewer number of stalwart companies. It is often described as a “mass niche” by industry experts, despite the \$3 billion in revenue forecast for this year. [77] Large tech-conglomerates are in the midst of building their wearable tech patent portfolios. In 2013, Google was awarded 2,000 patents, nearly double the number of all patents previously awarded combined. According to IFI CLAIMS 2013 top 50 US Patent Assignees, Google ranked 11th on the list this year with 1851 patents. [78] Of these patents, the trending shares are increasingly related to wearable tech, shifting Google’s patent portfolio from the litigation-exposed smart phone market to the

emerging wearable tech market. Figure 10, indicates the top U.S. patent filing companies for the past decade, and their share of the overall wearable patent landscape. Strategically, Google has opted to adopt Apple's model for technology acquisition: buying patents crowding a tech landscape in order to get complete vertical control of an entire technology platform. This is in contrast to Google's approach to their smart phone patent portfolio, where licensing and cross-licensing deals were in place for various proprietary positions. [79]

U.S. Patents for Smart Devices by Segment (2003-2013)



Note: Cumulative filings do not equal total filings due to several patents containing more than one segment

Source: U.S. PTO, Hanover Research analysis

Figure 9. U.S. patents by wearable segment.

During its infancy stage, interoperability and industry standards will influence standard-essential patents and FRAND-licensing patents (Fair, Reasonable and Non-Discriminatory). Patents that cover technology verticals that are essential for the implementation of a given standard will be licensed on FRAND terms, much like the mobile and semiconductor industry during its infancy stage. This is to ensure interoperability among various components and devices, across different networks, while still providing the necessary incentives for continued innovation in the wearable tech space. [80]

Top U.S. Patent Filing Companies for Wearable Tech, 2003-2013

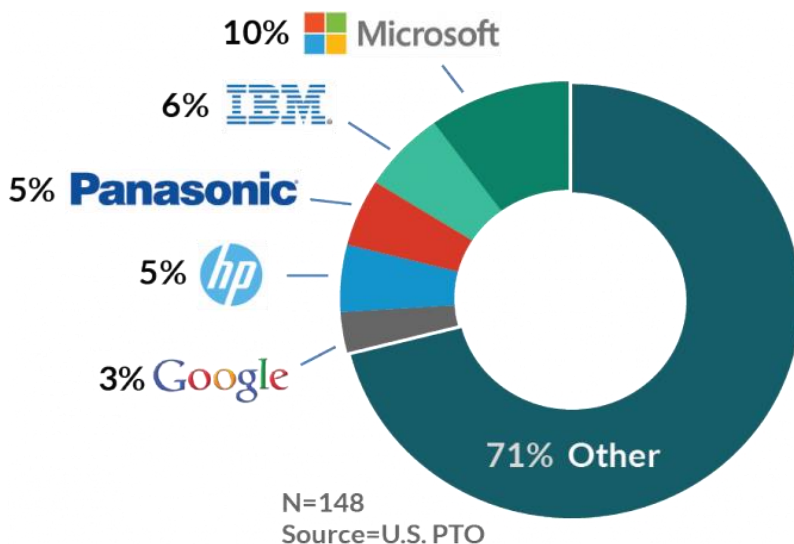


Figure 10. Top tech companies and their share of the overall wearable patent landscape. (<http://www.hanoverresearch.com/insights/the-rise-of-the-wearable-tech-market/?i=food-beverage>).

Motorola Mobility sought from Apple, 2.25% of net sales on all iOS products (amounts to roughly \$12/iPhone) that use essential industry standard patents (video-streaming and Wi-Fi technology, in particular). [81] Apple, however, believed that these terms were not consistent with FRAND terms and accused Motorola of seeking egregiously excessive payments on essential technology for mobile devices. District Court judges have ruled that Motorola was obligated to FRAND license the patented technology under the contractual agreements between Motorola and the standardization issuing organizations, namely the ETSI and IEEE. This has called Google's patent strategy of acquiring Motorola Mobility for \$12.5 billion into serious question; vertical acquisition or developing in a nascent tech sector must be weighed against industry standardization processes and possible FRAND licensing obligations. [82] Despite much of the Motorola Mobility portfolio being labeled 'FRAND', Google has still decided to retain much of the patent portfolio, as it prepares to sell off Motorola Mobility to Lenovo. [83] Any basic benchmarks and methodologies used in developing a FRAND, particularly in the field of smart wearable technology, must strike the most equitable balance between maintaining innovation incentives and royalty stacking. In a recent development that may have far-reaching implications on how IT-related patents are licensed, Qualcomm's model of aggressive licensing is facing a governmental investigation for alleged monopolistic practices in China, in addition to ongoing regulatory investigations in the U.S. and Europe. The IEEE, the electronics standards group that establishes protocols so technology can be used across devices, may adopt a proposal to prevent excessive fees charged for licensing technology in Wi-Fi standards now essential to connect all devices- hand-held or not- to the Internet. The measure would limit patent owners from getting courts to block product sales of companies using the technology without paying what has amounted to exorbitant fees. This standard, if adopted, could have far-flung implications on licensing models across a wide swath of technology, bioinformatics driven devices included. [84]

Patent Strategy

Protection for these strategic business assets will at its core, include a robust patent portfolio with a broad family of patents and an accompanying patent leveraging program. The beguiling jewel of any leveraging program will be its licensing and assignment program. Any leveraging program in this space will incorporate a four-step process: (1) compiling a target list of infringers or potential infringers within the niched space; (2) preparing claim charts, itemizing how infringing product features read or potentially read on every element of at least one claim; (3) conducting a patent valuation; and (4) negotiating a license/assignment. [85] The terms of the licensing/assignment agreement will vary depending on the targeted companies, and whether the targeted companies are actively infringing, or operating tangentially to the outer boundaries of the scope of the patent claims. Terms will also depend on whether it is an exclusive license or a non-exclusive license; and whether assignees are whole or partial.

Step 1

In compiling a list of targets, priority should be given to targets that are actively infringing every single element of at least one claim in the patent. Market-efficiencies may call for allowing the target to continue commercialization, provided they engage in a “stick” agreement, versus a “carrot” agreement. By being in a state of active infringement, this provides tremendous leveraging power to the patent owner and allows for certain terms to be dictated in the favor of the patent owner, hence the term “stick”. Moreover, provided that the specification of the parent application is expansively constructed, additional continuation applications may be filed that claim priority to the earlier parent application, and have claims that are drafted that read on currently commercialized product. This type of post facto infringement targeting is dependant on the newly drafted continuation claims finding support in the specification of the earlier parent application. This strategic insight underscores the importance of drafting the parent specification and claims narrow enough to navigate through the crowded patent landscape, while still being expansive enough to maximize value, deter design-arounds, and foresee emerging trends in order to leverage strategic positioning.

Companies that are not actively infringing, but operate in a space that is tangential to the scope of the patent, should also be targeted. Market efficiencies may dictate that a company adopts or piggy-backs on new features that are covered by the patent, and should enter into a licensing agreement with the patent owner in order to capitalize on this protected feature and potential market. The company already operating in the space may have the requisite know-how on reducing the concepts covered by the claims into a commercially-viable embodiment, with more efficiency than the owner or assignee of the patent. In such a case, “carrot” agreements may be entered into, which may have more market incentives for the licensee, since active infringement is not present.

Step 2

Once targets are compiled, claim charts and infringement reports are created, aligning distinct product features of the potentially infringing product against distinct elements of the patent claim. An additional column provides the claim construction, whereby key elements within the claim are defined, preferably defined as broadly as possible within the constricts of the specification in order to trap the product features of the potentially infringing product. In

addition to serving as the basis of an infringement analysis and licensing negotiations, it can also be instrumental in providing free and clear pathways to patentability and validity in order to help: (1) determine portfolio coverage and gaps; (2) discover product options that weren't earlier apparent; and (3) serve as a precursor to a freedom-to-operate framework, paving the way for commercialization, leveraging, and mitigation of litigation risk. Given the growing competitive landscape of the wearable market, litigation rates will likely mirror other tech sectors [86], underscoring the strategic value of having a panoramic view of the patent landscape.

Step 3

Prior to committing into commercialization or licensing efforts, a rigorous valuation of the patent or family of patents will need to be performed. There are three classic valuation methodologies that are used: (1) cost; (2) income; and (3) market. The cost approach aims to determine the value of the patent by determining the cost of acquiring or developing an equivalent approach. [87] The income approach aims to determine patent value by forecasting future revenue over the course of the life of the patent. The estimated future revenue is then discounted to arrive at a single net present income. [88] The market approach aggregates transaction data of similar or comparable technology in order to yield a valuation.

The market approach yields the most accurate estimate, provided that information of similar or comparable transactions are available. [89] Unfortunately, patent transactions operate in secondary markets and are often confidential. Even if one was privy to certain patent transactions, it is exceedingly difficult to find comparable transactions with analogous IP and market parameters. [90] Likewise, the other approaches are also replete with flaws. Subsequently, in order to perform a thorough assessment and arrive at a fair value, its probably best to use a hybrid approach, whereby distinct metrics spanning across all three methodologies are accounted for: (1) remaining years in the patent life; (2) number of claims, especially independent claims, and the breadth and strength of these independent claims; (3) size of patent family/priority analysis; (4) patent validity analysis; (5) foreign status; (6) and the number of forward and reverse references; (7) technology/product review- parameter differentials between patented product and substitute product; and (8) market coverage analysis/sales forecasts. [91]

If comparable transactions aren't available and the patented product has been commercialized and substantive financial data is available, then the hybrid approach to valuation referred to as the comparative advantage valuation model may be used to derive a fairly accurate patent valuation. Given the patent segmentation spanning across individual wearable and biosensor products, the CAV model may be the optimal means of valuation. The CAV model can accurately derive a base patent value for each patent covering distinct components and parameters of individual products. Adapting Ted Hagelin's CAV analysis of a hypothetical case study, we will similarly conduct a valuation analysis on a series of hypothetical patents covering various components of a hypothetical wearable/biosensor product.

Hypothetical Valuation Case Study

OpticProbe, Inc. is a start-up firm specializing in the research, manufacturing, and distribution of biosensor and wearable devices. Its flagship product is the biosensor device, YouSense. YouSense binds selective biomarkers using embedded nano-tube sensors with receptor substrates that produce an optical change upon binding. This optical change is then translated into a quantifiable digital signal by a transducer component, and then processed by an analyzing component and cloud-based back-end analytics in order to identify the biomarker. OpticProbe operates in the burgeoning subfield of nano-biosensor devices, with a total of ten firms with a total market value of \$850 million.

Step 1

The first step involves identifying the patents and its associated device parameters. For instance, the competition parameter for the sensor component is sensitivity; how much biomarker is needed in a given unit of sample in order to trigger a reliable detection and a robust signal. The competition parameter for the transducer is the translation fidelity or conversion rate; how much deviation or spill-over is there between sensor input and transducer digital signal output. Finally, the competition parameter for the device and back-end processing components is the provisioning rate; how much latency exists in the real-time data provisioning of the analytics. Once competition parameters are identified, the associated patents within the patent portfolio need to be identified. A quick survey of the OpticProbe portfolio reveal that patent '234 is associated with the sensor component and sensitivity parameter; patent '567 is associated with the transducer and conversion rate parameter; and patent '890 is associated with the processing components and the provisioning rate.

Step 2

The next step is to research parameter values of the product and substitute product in order to determine the relative competitive advantage in comparison to other patents incorporated in the associated product. The table below includes the hypothetical parameter values and hypothetical competitive advantage for an average nano-biosensor and YouSense.

The technical Intellectual Property value is first determined by subtracting the tangible asset value from the total market capitalization value. The remaining value, total intangible asset value, is parsed from the technical Intellectual Property value, by dividing the sales, general, and administration expenses by the SG&A minus advertising plus business processing spending. This value is then further multiplied by the total intangible asset value percentage to yield the intangible asset percentage $[(\$20 \text{ m}/\$20 \text{ m} + \$10 \text{ m}) \times (100\%-15\%)] = 56.67\%$. Finally, the technical Intellectual Property Value is derived by subtracting the intangible percentage from the total intangible percentage, and then multiplying the intellectual property value percentage by the net present value $[(85-56.67) \times 150\text{m}] = 84\text{m}$. This year's sales records, projected growth, forecasted sales, remaining patent term, and present discount rate are all required to arrive at the net present value of \$150m, and ultimately with a final technical IP value of 84m. [92] The base value for each patent corresponding to a unique product parameter is 22.7m, 27m, and 63m, respectively.

	Sensor Amount/unit for threshold detection	Transducer Deviation between sensor input and signal output	Processor Latency (seconds) of data provisioning
Patent	‘234	‘567	‘890
Substitute Product Parameter Values	500	0.25	0.8
YouSense Parameter Values	250	0.10	0.2
Base Competitive Advantage (YouSense PV-Product PV)/Product PV	0.5	0.6	0.75
Relative Competitive Advantage (Base Competitive Advantage/Total Base Competitive Advantage)	$(0.5/1.85)=$ 0.27	$(0.6/1.85)=$ 0.32	$(0.75/1.85)=$ 0.41
Base Value RCA x 100 % (Technical Intellectual Property Value)	$(27\% \times 84 \text{ million})=$ 22.68 m	$(32\% \times 84 \text{ million})=$ 26.88 m	$(75\% \times 84 \text{ million})=$ 63 m

Step 3

The base values then have to be discounted by their corresponding risk value. IP risk factors include obsolescence of rights, loss of rights, and strength of rights. Obsolescence of rights is the measure of risk against the patent becoming functionally obsolete prior to the expiration of the product life cycle. [93] Obviously, obsolescence, loss of rights due to post-grant events, and the neutering of patents with crafty design arounds, all have tremendously devaluing effects on the value of the patent. This potential devaluing risk can be priced into the overall valuation by multiplying the base value of each patent by $(1 - \text{total IP risk factors})$. [94] For instance, let us suppose that OpticProbe patent counsel have discovered a publication on nano-biosensors that predates the priority date of the ‘234 patent and believes that there is a 5% chance of the ‘234 patent being invalidated in a post-grant proceeding. Also, let us suppose there is an additional 5% chance that competitor engineers would be able to design around the sensor optical output by designing a sensor that creates an electrical output upon binding of the receptor substrate. In this scenario, the total IP risk factor for ‘234 is 10%. For the purposes of this case study, let us assume that the remaining patents retain the same risk factor. The following table illustrates the final adjusted base value for all three patents, adjusting for their respective patent risk factors.

As it has been illustrated through this hypothetical case study, the ‘890 patent is the most valuable patent in the portfolio, valued at an adjusted \$56.7m. As one can imagine, the provisioning aspects of a biosensor/wearable product would be expected to be the most valuable component/parameter of such a product. Without real-time analytics and provisioning, these devices serve very little use to patients or consumers. The biological recognition element that the signal output and analysis is derived from may, no longer be representative of the current physiological state due to the back-end latency. Patents ‘234 and ‘567 are valued at \$20m and \$24m, respectively. Now that we have adjusted base values for the three patents in the OpticProbe YouSense portfolio, what are some strategic steps to leverage this valuation?

	Sensor Amount/unit for threshold detection	Transducer Deviation between sensor input and signal output	Processor Latency (seconds) of data provisioning
Patent	‘234	‘567	‘890
Adjusted Base Value Base Value x (1- total risk factor)	22.68 (1- 0.10)= \$20.412m	26.88 (1-0.10)= \$24.192m	63 (1-0.10)= \$56.7m

IMPACT OF ALICE AND MAYO ON IP

The Supreme Court decisions in *Bilski*, *Mayo*, and more recently in *Alice*, have resulted in a legal sea change in terms of patent subject matter eligibility. Moreover, the recent reforms of Congress in the America Invents Act, has dramatically changed the landscape of patent law. Figure 11, depicts these developments in a timeline fashion. One need not look any further than the changes in the first to invent system- the U.S. being the lone holdout- to a first to file standard. Aside from the file-to-file changes, there are a number of other prosecution changes that the AIA has introduced, such as making it easier to invalidate applications and patents via a newly expanded suite of pre-issuance and post-grant proceedings. The USPTO has codified these judicial and legislative changes into examination guidelines that may be found in the Federal Register or the Manual of Patent Examination Procedure. Neither of which are helpful, without some prosecution guidance helping navigate the complex legal labyrinth that is patent prosecution and enforcement.

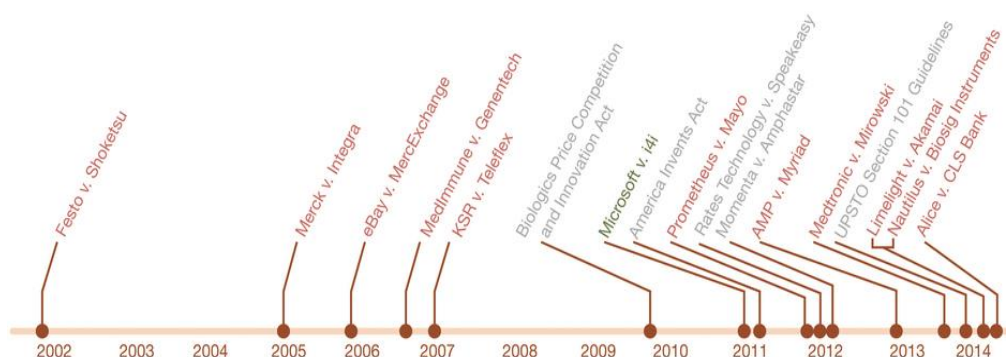


Figure 11. Timeline of recent developments in patent law. (Brougher, Joanna, and Konstantin Linnik. "Patents or Patients: Who Loses?" *Nature Biotechnology* 32 (2014): 877-80. Web. <http://www.nature.com/nbt/journal/v32/n9/fig_tab/nbt.3005_F1.html>).

Patent Prosecution

All of these changes have one obvious effect: making it more difficult to get a patent, most particularly in the fields of software and biotechnology. Bioinformatics, laying at the intersection of these fields, is most exposed to the litany of prosecution reforms. In the wake of these reforms, there are a number of bioinformatic parameters associated with wearable back-

end analytics that have been found to be worthy of protection by the PTO, namely the software tools for biological data mining, visualization, pattern recognition, sequence alignment, modeling, predictive tools, etc. Additionally, the mechanical and biochemical sensors, along with the small factor forms of wearable devices have been found to be patent eligible, even under the constraints of the new regime. The key is to identify these constraints, and circumvent them with sound diligence, claim drafting, office action responses, and enforcement.

IP Due Diligence

The first step is to assess the current patent portfolio of the company. Portfolios of patented technology and licensed technology should be categorized into different competition parameters. For instance, the OpticProbe YouSense portfolio could be categorized into several parameters: form factor, display interface, sensor sensitivity, signal transduction, real-time analytics/provisioning, data mining, recognition, structure modeling, predictive modeling, dashboard alerts/visualization, app interfacing, etc. Competition parameters should be further categorized into “definitely utilize”, “likely utilize”, and “not likely utilize”. Product features of YouSense should be aligned with patent claims of the associated patents in the portfolio, categorized by parameter and utility. Categorized claims without any product features or an underrepresented amount of features represent areas in which OpticProbe may decide to roll-out a new and improved iteration of YouSense embodying these new features. [95] If it is consistent with OpticProbe’s product development strategy, they may want to roll-out other wearable products that embody these claim features. Crafting patent strategy to align with OpticProbes business plan is crucial. Conversely, the empty parameter bins- competition parameters with a lack of associated claims- should be seen as an opportunity to leverage the portfolio by filing additional continuation claims, claiming priority to the original parent filing, provide the scope of the continuation claims can find complete support in the parent filing. Filing continuation claims can have the benefit of gaining a market foothold on an emerging feature, or merely, play a defensive role- blocking potential competitors from developing and commercializing the feature. Having a robust portfolio with broad claims provides OpticProbe with significant competitive advantages, not the least of which is the luxury of engaging in a potentially lucrative licensing program.

Claim Drafting in the Wake of *Bilski*, *Mayo*, *Myriad* & *Alice*

With the ever-changing landscape of patent prosecution, in the wake of *Bilski*, *Myriad*, *Mayo* & *AIA*, one cannot underscore enough the competitive value of having a sound claim drafting strategy. Any sound strategy would start with frequent inventor and examiner interviews. Interviews will bring the high-level picture of the invention into focus, and additionally, place the granular-level details into proper context. Having this high-low level or top-down perspective of the invention will underlie more value-laden, tailored claims that cover the breadth, as well as the more nuanced elements of the invention. It has been demonstrated, at least anecdotally, that frequent interviews with the examiner often facilitate a more streamlined prosecution pipeline. [96] This is presumably due to the fact that subject matter issues become elucidated earlier in the pipeline, enabling the prosecutor to adjust claim

language to fit within subject matter eligibility guidelines in lieu of the recent case law and legislative reforms.

Patentable subject matter has been the most contentious issue in patent law, polarizing people from kitchen tables in the Bible belt, all the way to senate committee hearings on the Hill and the hallowed halls of the Supreme Court. At the heart of this debate, is whether certain claims, particularly method claims, were attempting to monopolize mere abstract ideas, such as algorithms and DNA, without any tangible implementation or intervening step. *Bilski v. Kappos* (2010, Supreme Court) and *Mayo Collaborative Services v. Prometheus Laboratories* (2012, Supreme Court) are two major recent developments regarding patentable subject matter. *CLS Bank v. Alice* (2014, Supreme Court) represents the third leg of the subject matter eligibility trilogy. The Supreme Court is unified in all three cases: the claims must go beyond the natural occurrence/principle/phenomenon/law; naturally occurring things are not patent eligible. In *Mayo*, in light of *Bilski*, the Supreme Court reaffirmed the shortcomings of the machine-or-transformation test, and determined the patent under consideration to be ineligible since it attempted to monopolize a mere correlation that is found in the law of nature. [97] In *Myriad*, the Supreme Court ruled that, while isolated genomic DNA is not patent-eligible (naturally-occurring) under section 101 of the Patent Act, cDNA (resulting from the intervening step of reverse transcription) is. [98] Most recently, *Alice*, has reinforced this patent eligibility standard, in the context of computer software and business method claims. The Supreme Court, in *Alice*, applied the two-prong patent eligibility framework from *Mayo* against the computer-implemented method claims at issue, and held them ineligible. The Court ruled that the claims were (1) a natural phenomenon, abstract idea, etc.; and (2) the claims do not recite sufficient additional elements to transform the natural phenomenon, abstract idea, etc. [99] The Court held that, “merely requiring generic computer implementation” did *not* “transform that abstract idea into a patent-eligible invention.” *Alice* ushers in, yet another watershed moment- claims to a method, a computer system configured to carry out the method, and a computer-readable medium containing program code for performing the method all fell invalid under Section 101 (subject matter eligibility). This line of jurisprudence could have a significant impact on the validity of bioinformatics applications.

With respect to YouSense, the novel and non-obvious aspects that deal with providing technical solutions to consumer demands, will fall outside of the 101 hazard, and will be patent eligible. Emphasizing these aspects in the claim and specification is crucial, especially regarding the back-end analytics and provisioning portions. It’s the back-end bioinformatic tools that are the most exposed to the risk of 101 ineligibility. Claims covering the back-end will need to be drafted in the context of a substantive computer implementation that results in a tangible, transformative outcome. The following are some of the claim drafting best practices in capturing Intellectual Property for software-implemented, bioinformatics-related innovations in a post-*Alice* landscape. According to the preliminary examination instructions issued by the USPTO in the wake of the *Alice* decision, claims that are drawn to software or computer-implemented methods should have meaningful limitations beyond generally linking the use of an abstract idea to a particular technological environment. [100] Claims that require no more than a generic computer to perform generic computer functions- that are well understood, routine and conventional activities previously known to the industry- will not be patent eligible in this new post-*Alice* era. [101] In terms of specific approaches, drafting multiple independent claims, each embodying different ways to perform the computer-implemented method, is sound strategy. However, as a word of caution, there will always be a

risk of a restriction requirement- a request to elect claims that pertain to the same invention, and cancel the set of claims that pertain to a wholly, different invention. Notwithstanding, potentially receiving a restriction requirement will have significantly less adverse patent consequences than a section 101 eligibility rejection. Additionally, system, computer-readable medium and computer-implemented method claims with greater detail and clarity-without losing any strategic value- will be crucial in avoiding 101 rejections. Moreover, including a particular machine or particular transformation, which implements or integrates the abstract idea, is critical. Finally, add features that are more than conventional, routine, insignificant extra-solution, or simply a mere field of use. The following hypothetical claims drawn to YouSense shall serve as an illustrative guide.

A computer-system-implemented method, the computer system having at least one programmed processor to implement the method, the method comprising:

- *continuously collecting and processing biomarker data with respect to an individual from a wearable sensor device;*
- *the computer system—*
 - (i) obtaining physiological information for the individual based at least in part on output from a contextual sensor; and*
 - (ii) based on physiological data collected from the wearable sensor device, patterns from user logged history, and other contextual information, determining a recommendation for the individual.*

A non-transitory computer-readable medium storing instructions that, when executed, cause a wearable monitoring device to:

- *receive biomarker data transmitted by a device worn by the user, measuring biomarker levels of the user and detected by the device;*
- *receive contextual data transmitted by a device worn by the user, environmental cues of the user, along with activity levels of the user and detected by the device;*
- *cause real-time display presence of a first indication of the biomarker data on a first display of a dashboard app and a second indication of the contextual data on a second display of the dashboard app; and cause a recommendation to the user via a text-crawl in at least one dashboard display, text message, e-mail message, or any other form of secure communication.*

Notice the detailed steps in the computer readable medium claims versus the computer-implemented method claims. This may have been necessary as a way to design around or not encroach on a crowded prior art space. Dependent claims will refer to each of these claim sets, further adding limitations and alternative embodiments. Apparatus and system claim sets will also be included to round out the metes and bounds of the invention. Finally, a descriptive specification- with high-level system diagrams and flow diagrams- will refer to the diagrams in granular detail and provide exemplary support for the claims. Once applications have been filed, the first office action response is typically delivered within a year of filing. If patent eligibility issues are raised by the Examiner, it is imperative to first argue that the supposed “abstract” idea is not “abstract” within the meaning of the *Alice* decision. *Alice* rejections predicated on an abstract concept of recent origin- such as small-factor sensors, real-time data mining and provisioning, etc.- should be overcome by arguing that these concepts are not long-prevalent and fundamental within the meaning of *Alice*. In other words, they are not abstract concepts, such as risk hedging or intermediated settlements, both of which are concepts that

have been entrenched in financial systems around the world since time immemorial. [102] Alternatively, express limitations in the claims should be pointed out, which would run counter to any Examiners contention that the claims merely recite an abstract idea and instructions on how to apply it. Another often misconstrued holding of *Alice* has been that all method claims that recite computer implementation will be scrutinized for an *Alice* rejection. However, it is important to remember that only claims reciting abstract ideas will be subject to the *Alice* inquiry; claims not drawn to abstract ideas- not fundamental or long-standing- will not trigger the inquiry, irrespective of containing computer implementation language. [103] In the six months following the *Alice* decision, Examiners have issued a number of *Alice* rejections, anchored on the notion that a computer implementation is per se patent ineligible. Despite the examination guidelines in place, Examiners may still need to be reminded of their burden of proof and the two-prong analysis of *Alice*. The following is an excerpt from a recent *Alice* rejection issued by an Examiner in reference to a software-related invention; note the circular fallacy. [104]

Claims... are rejected under 35 U.S.C. 101 because the claimed invention is directed to non statutory subject matter. In the instant invention, the claims are directed towards the concept of... [This] is considered a method of organizing human activities, therefore the claims are drawn to an abstract idea. The claims do not recite limitations that are "significantly more" than the abstract idea because the claims do not recite an improvement to another technology or technical field, an improvement to the functioning of the computer itself, or meaningful limitations beyond generally linking the use of an abstract idea to a particular technological environment. It should be noted the limitations of the current claims are performed by the generically recited processor. The limitations are merely instructions to implement the abstract idea on a computer and require no more than a generic computer to perform generic computer functions that are well- understood, routine and conventional activities previously known to the industry. Therefore, claims... are directed to non-statutory subject matter.

Patent enforcement strategies in bioinformatics will require considerable rethinking as well, in light of all of the seismic changes in the landscape of patent law. Most notably, how should a start-up or a young, fledgling tech company, such as OpticProbe, enforce their patents in the face of alleged infringement. Conversely, how should a company, like OpticProbe, enter a tech consumer space riddled with so-called patent trolls? After all, what good are robust, iron-clad claims, if you don't have a comprehensive enforcement strategy in lieu of all of the recent developments in law and public policy?

Just a decade ago, life science companies with core technology revolving around bioinformatics tools were weakly anchored to IP holdings. This has largely been due to the fact that much of the commercial offerings integrated public domain tools and data structures that weren't proprietary in nature. Moreover, IP budgets have traditionally been relegated to products that weren't tethered to back-end tools- with questionable proprietary value- and to products with more distinct boundaries and clearer revenue streams. There has been a shift in sentiment over the past decade. This has largely been due to three major events: (1) the USPTO's decision to allow expression sequence tags in its utility as probes to be patentable;

(2) the *State Street* decision, whereby the Supreme Court ruled that tangible applications of mathematical algorithms and business methods, as well as data processing, and computer programs, were patentable; and (3) the explosion of high-throughput methods for sequence and data gathering. [105] Leading this trend in patenting bioinformatics inventions is none other than IBM; with several thousand patent filings in a burgeoning portfolio subclass. [106] This should be a signal to the proprietary value of bioinformatic inventions. IBM, as well as other life science and IT companies are all jumping on the bandwagon; attempting to all plant their flag in this relatively uncharted territory. As of 2012, there were 90,064 bioinformatic patents- a dramatic rise from the end and turn of the century. [107] There has been a steady state of patenting activity from 2006 and 2011, and a sharp rise in 2012, which is forecasted to continue to rise into the following decade. This is reflective of industry trends and the current renewed interest in code, algorithms, network architecture, cloud provisioning, and small form-factor devices. Despite the amorphous standards for an abstract idea under *Alice*, recent patenting activity has nevertheless centered on algorithms, methods, and database methods that might be specific to handling life sciences data. In terms of the biochip and biosensor space, the space is still relatively immature from a patent perspective; only 1,800 of the 90,064 patents are related to biochip manufacturing and use. [108] Getting in on the game early provides significant strategic advantages, namely excluding competitors and positioning themselves for favorable leveraging opportunities. However, without sound enforcement strategy, the painstakingly built portfolio could be potentially worthless.

Licensing

The first enforcement tool has to be leveraging the current portfolio, by way of licensing and acquisitions/sales. Licensing in bioinformatics is segmented; with some companies exclusively licensing software tools; others license access to databases; while others license devices, etc. [109] On one hand, the diversity of market participants can make it difficult to draft claims in anticipation of competitor positioning. However, on the other hand, this dizzying licensing array should embolden executive committees and directors across the country to expand their portfolio in the direction of algorithms, software tools, interfaces, computer-implemented methods, etc.

In the competitive world of bioinformatics it comes extremely important for a company, whether start-up or Fortune 50, to have a well-developed IP portfolio that can add strategic value to a company. A company should align its potential IP with its business strategy to ensure that any IP that it is targeting flows with its current business strategy. One such strategy is for engaging in a licensing program. The strategic value stems from the fact that risks can be shifted to the licensee, from manufacturing to marketing. Risks are shifted in proportion to the exclusivity of the license, as well as overlap in markets between licensor and licensee. Some key clauses in any term sheet between a licensor and licensee will include license scope, enforcement rights, future prosecution, and ownership of improvements, liability, indemnification, warranties and representations. [110]

Currently held patents can be licensed out as a steady stream of revenue or used to cross-license technology proactively or in reaction to allegations of infringement.

Licensing can bring in a flow of capital to a company and help to build a patent portfolio. For example, Qualcomm has used licensing agreements as a core element of its business plan

and collected over \$30.5 billion during the last five years in licensing fees alone. [111] There are two basic types of licensing: exclusive and non-exclusive. An “exclusive license prevents the patent owner (or any other party to whom the patent owner might wish to sell a license) from competing with the exclusive” license holder with the licensing agreement setting forth the terms of the geographic region of use, length of time of license and even the field of use”. [112] A non-exclusive license allows the grantor of the license agreement to also use the technology. It would behoove a company seeking to license a patent to look for an exclusive patent license because it would prevent a competitor from also practicing the same licensed technology. It is no secret that a small or start-up biotechnology firm can rely on exclusive licensing rights to ensure access to high-risk capital and increase its competitiveness. [113] However, in the case of OptiProbe’s licensing of patents related to the YouSense technology, it would make sense to enter into non-exclusive deals that are drafted narrowly in terms of scope, both temporally and geographically. This strategy may allow OptiProbe the opportunity to license the patent to others, across the spectrum, to generate more revenue.

In reviewing licensing in bioinformatics, one must be cognizant of different licensing schemes across a range of technologies, including software licensing, medical device licensing, data licensing, and licensing systems. Licensing royalty rates in bioinformatics can vary based on which technology sector the IP is based. For example, electrical and chemical patent royalty rates tend to be lower at a reported approximate of 4.25% of the product in comparison to a pharmacy patent royalty rate of 7.0%. [114] Royalty rates are conventionally determined according to the 25% Rule, which stipulates that a licensee only pays a portion of the profits to the licensor, due to the additional costs and uncertainties incurred by the licensee in converting the technology to revenue. [115] The rate is calculated by multiplying the expected operating profit margin percentage for the product embodying the IP at issue, by 25% in order to arrive at the estimated royalty rate. [116] The 25% Rule can be advantageous when trying to determine the royalty rate for infringements of new technology as industry standards on licensing may not have been developed. With respect to OptiProbe’s licensing scheme, royalty rates between 4-7% may be reasonable based on the earlier stated projected financials. As a bioinformatic company seeking a “stick” license through a demand letter against an accused infringer, OptiProbe may seek the higher end of this royalty range, whereas, the lower range may be more appropriate in a “carrot” proposal targeting a company who may be projecting into the biosensor embeddable space. Depending on the specific area that a bioinformatics patent was aimed, the royalty rate and licensing scheme can vary drastically. For example, a license can be provided for free for non-commercial, academic institutions, and personal use, but for purchase by commercial entities for a one-time set up fee plus an annual fee per user. [117] This would be an example of a non-exclusive license with an open source component stemming from a university. Another example of a licensing comes from the National Institute of Health (NIH), which is specifically aimed at start up companies. The NIH has two licensing options aimed at “minimizing barriers to entry faced by start-up companies under exclusive licenses and provide a structure that encourages and supports the commercial development of early stage NIH and FDA technologies”. [118] The program offers either an exclusive evaluation license with a \$2,000 execution fee or an exclusive commercialization license with includes tired upfront execution royalties upon certain events, a set earned royalty rate of 1.5% and a set sublicensing royalty rate of 15%. [119] In a recent development that may have far-reaching implications on how IT-related patents are licensed, Qualcomm’s model of aggressive licensing of royalty stacking and coercive cross-licensing is facing a governmental

investigation for alleged monopolistic practices in China, in addition to ongoing regulatory investigations in the U.S. and Europe. The IEEE, the electronics standards group that establishes protocols so technology can be used across devices, may adopt a proposal to prevent excessive fees charged for licensing technology in Wi-Fi standards now essential to connect all devices- hand-held or not- to the Internet. The measure would limit patent owners from getting courts to block product sales of companies using the technology without paying what has amounted to exorbitant fees. This standard, if adopted, could have far-flung implications on licensing models across a wide swath of technology, bioinformatics driven devices included. [120]

Patent acquisition is another way that a company can build its patent portfolio without having to actually invent the subject matter of the patent or engage in licensing. In patent acquisition a company will buy the rights to the patent from the owner of the patent, whether it is an entity or the actual inventor. When looking to acquire patents, the strategic value of the patents being purchased and the company's business plan should be considered. In order to acquire certain patent rights, a company can either directly contact the owner of the patent or work through a third party broker in order to acquire the rights. Use of a brokerage to acquire rights will cost either a flat fee or hourly rate and a success fee if the sale is consummated. A potential benefit to patent acquisition or even patent application acquisition is that the patent acquired or the granted patent from the patent application acquired previously can then be licensed to other companies/individuals in order to bring in more revenue. Moreover, it can be used as a tool for avoiding costly litigations; such was the case in twitters acquisition of IBM's 900 or so patents covering the efficient retrieval of uniform resource locators, for presenting advertising in an interactive service, and for programmatic discovery of common contacts. In a bulk-acquisition of OptiProbes portfolio covering the sensor, transducer, and processor patents ('234, '567, and '890), a reasonable offer of sale would be approximately \$100 million based on the earlier valuations performed. Alternatively, the individual patents could be sold to different companies, but with the caveat that they each would have significantly less value than their valuations due to the fact that they each cover distinct components and may potentially serve to block each of the companies from an unfettered freedom to operate in the embeddable biosensor space.

Enforcement

Many in this field are start-ups and small companies, and lack the financial means necessary to enforce their IP rights or defend their products against competitors and patent trolls should they become involved in costly and time-consuming IP litigation. Thus, demand letters and licensing proposals should be a first option in dealing with an accused infringer. If this tact doesn't work, then litigation must be pursued as a last case resort. Patent litigation has many different components, from claim construction, validity, infringement, damages, to counterclaims and affirmative defenses. [121] In the case of bioinformatics litigation in this newly ushered *Alice*-era, issues of validity will be the area of most risk exposure. Patents may be subject to invalidity based on *Alice* subject matter eligibility standards, not to mention on obviousness, inventorship, enablement, and written description grounds. Demand letters and licensing ultimatums could potentially expose the patent owner to declaratory judgments- a litigation tool for providing a legally binding determination of legal uncertainties in order to

avoid adjudication. The following is an illustrative example of the use of a declaratory judgment, with validity issues as the underpinning.

In one of many examples, Counsyl is counter-suing Myriad, asking the court for a *declaratory judgment* that (1) eight specified Myriad patents are invalid and (2) even if they are valid, Counsyl isn't in infringement. Declaratory judgments are an often employed litigation tool because they allow prospective infringement defendants to ask a court to resolve the uncertainty of infringement, and more importantly, to potentially invalidate the claims altogether. Moreover, the declaratory judgment allows Counsyl to mandate Myriad to counterclaim infringement on Counsyl in this suit, or otherwise be estopped from asserting infringement in any future suit. [122]

Alice-era validity issues may also be remedied through the new post-grant proceedings in the USPTO, as codified by the recently enacted America Invents Act. Notwithstanding the shift from the first-to-invent to the first-to-file system and the substantial expansion of prior art- the salient features of the AIA are pertaining to the remedial procedures for invalidating patents through post-grant proceedings. Under the AIA reforms, the mechanism for invalidating patents has been broadly expanded- including prior art, lack of written description, lack of enablement, and ineligible subject matter. It is adjudicated before the Patent Trial and Appeal Board and is based on motion practice. There are, however, significant limitations. For one thing, it can only be initiated within 9 months of the issuance of the patent or a reissue that broadens the scope of the original claims. Additionally, a challenger is precluded from a Post Grant Review (PGR) if a district court challenge was earlier initiated, and vice versa, precluded from district courts if a PGR was earlier initiated. Finally, the standard for initiating the PGR is that it is 'more likely than not' that one or more of the challenged claims are unpatentable. [123] In the aftermath of the AIA reforms and the sudden rise of bioinformatics patents being issued, the PGR pathway to invalidity grounded in *Alice* should be anticipated.

Another potential concern, especially for smaller bioinformatics firms, is the threat of demand letters and litigation from so-called patent trolls, more specifically, non-practicing entities. The NPE's that are the focus of this analysis are the patent aggregators, such as Acacia and Intellectual ventures (IV), and not the upstream actors, such as universities, individual inventors, etc. The issue is not the secondary markets for patents or even some upstream innovators and intermediaries. This secondary market creates liquidity and alienability of these valuable assets. It fosters innovation, the aggregation of which creates market efficiencies. Using patent monetization markets and VC financing as a metric for the current state of innovation and patent economy, we looked at data from the Robin Feldman's forthcoming UCLA J.L & Tech article, as well as data from the NVCA report, "Patent Demands & Start-up Companies: The View from the Venture Capital Community". This report looks beyond just patent litigation data, but also the patent demands, which include pre-litigation licensing demands, not to mention litigation threats and service of complaints. The report and all of the ensuing scholarship and public commentary by critics and practitioners is overwhelmingly in favor of curbing NPE's and especially the aggregating monetizers (i.e., Intellectual Ventures, Acacia, etc.). There is consensus among these sources that these demand letters have had a negative impact on venture capital financing. Most of the findings are empirical and based on poll data of actual venture capitalists that hold companies that have been subjected to these demand letters. *The vast majority of patent demands against startup companies come from entities that license or litigates patents as their core activity (Specifically, 59% of the venture capitalists and 66% of the startup companies reported that all or most came demands come*

from such entities). [124] One hundred percent of venture capitalists indicate that if a company had an existing patent demand against it, it could potentially be a major deterrent in deciding whether to invest. Roughly half indicate that it would simply be a major deterrent on its face, and the other half indicates that it could be a major deterrent, depending on the circumstances. [125] By some estimates, there has been a \$21 billion loss in VC investments over the past five years due to troll demanding. [126]

Bioinformatic companies, start-up or otherwise, are just as exposed to the troll pandemic. In fact, they may be even more exposed, than companies that do not exploit IT-based platforms, according to Bessen and Meurer, in a 2012 study (Figure 12).

Summary Characteristics of Lawsuits	
Publicly listed defendants Number	
Mean	15.3
Median	5
Percent	
Sole defendant	17
In litigation with 10 or more defendants	32
Software patent	62
Patent Technology Classes (NBER) Percent	
Chemical	1
Computers and communications	75
Drugs and medical	1
Electrical and electronics	12
Mechanical	4
Other	8

Figure 12. Number of defendants by sector subjected to NPE litigation. (Bessen and Meurer, 2012)

With this in mind, while Congress- in typical gridlock fashion- continue to mull over options on how to combat the troll problem, some bioinformatics start-ups have taken measures into their own hands. Case in point, Reflx, a Seattle based wearable and biosensor company, who recently launched their flagship product, Boogio, a shoe sensor. Reflx built a portfolio around Boogio and future products, and instead of waiting to be served a demand letter or complaint from a troll, they decided to form a strategic partnership, not with just any troll- but the market leader of NPE aggregators, Intellectual Ventures (IV). Why would a bioinformatics start-up get into bed with the king of trolls? Because Reflx gave IV equity in the company in exchange for help developing and selling software development licenses, according to Reflx CEO, Jose Torres. [127] In other words, strategic alliances may avoid threats of ex facto licensing, and leverage a start-ups' current portfolio in the secondary markets. The complexity of cloud server applications and software packages will allow us to interface with our physiological data streams in real-time. These systems will usher in a new era of personalized health care delivery, promising efficiencies in outcomes, access, and cost-containment.

However, it all begs serious ethical concerns, namely issues of privacy and encroaching on the public commons.

FDA'S ROLE IN THE 'APP' WORLD

"FDA clearance is an important step on the path towards getting genetic information integrated with routine medical care" - Anne Wojcicki

As mentioned earlier, the future of medicine in the succeeding decades will be an amalgamation of IT-lead personalized health care devices. Hence, the next big question is what type of device will the FDA classify as a personalized healthcare device and how long will it take to pass through FDA regulatory approval process? Also, what is the best strategy for a startup to navigate through this regulatory labyrinth?

One of the foremost aspects to consider is what kind of device the FDA considers to be a personalized medical device. According to the FDA's perspective as mentioned earlier: "Personalized medicine generally involves the use of two medical products – typically, a diagnostic device and a therapeutic product – to improve patient outcomes. [7] Therefore, according to current FDA standards and regulatory schemes, personalized medical devices diverge from class I to class III. Additionally, under existing guidelines, all devices regardless of its class are subject to general controls which includes the following: 1) establishment registration by manufacturers, distributors, repackages, and re-labelers, 2) medical device listing with FDA of devices to be marketed, 2) manufacture devices with GMP (Good Manufacturing Practices), 3) labeling regulations, and 4) MDR (Medical Device Reporting) of adverse events as identified by the user, manufacturer and/or distributor of the medical device. Class I devices are general controls and are commonly exempt from premarket notification 510(k). One example of a class I devices are Band-Aids. Class II devices are special controls which require Premarket Notification 510(k). One example of a class II device is diagnostic equipment. In fact for class II devices includes special labeling requirements, mandatory performance standards (domestic and international), post market surveillance, FDA medical device specific guidance, and pre-market notification by submission, and FDA review of a 510(k) clearance to market submission. Class III devices require Premarket Approval (PMA). One example of a class III device is replacement heart valves. The class III devices have the utmost stringent regulatory controls and are higher risk devices, where all class I general controls and class II special controls apply. Moreover, they are subject to FDA approval based on safety and effectiveness as well as PMA. [128] The class of the personalized health care device is unknown and indefinite under current FDA regulations. One of the first steps a start-up currently should take is to send a Request for Designation Document (RFD) to the Office of Combination Products which will determine the type of product and which the Lead Center of that product will be regulated in. [129] This will inform the start-up on what regulatory path to follow and dictate crucial engineering and design decisions.

Since the FDA regulations are very difficult and may take many years, an alternative opportunity for a start-up is gaining approval in the European Union, which usually takes considerably less effort and time than gaining approval through the FDA in the United States. However, this will only allow the start-up company to start selling the personalized healthcare

devices in Europe while waiting for FDA approval in the United States which might prove to be detrimental for the company's patenting strategy in the United States. This regulatory strategy should only be used if there is a relevant patent presence in the European Union. [130] Even though the Supreme Court upheld the ACA (Affordable Care Act), it left the Medicaid expansion option up to the individual states which could significantly alter the potential customers for the personalized healthcare market.

Reimbursement

One important aspect regarding the personalized medical devices is the ability to sell them to individuals and patients in the United States. The CMS (Centers for Medicare and Medicaid) develops and creates Medicare/Medicaid coverage and payment policies for many healthcare services but is not limited to ambulatory surgical centers, physician offices, hospitals, and nursing facilities. Typically many private insurance companies base their coverage on what the CMS decides to reimburse. Once CMS decides to cover the personalized medical device, the start-up company has access to all the Medicare, Medicaid, and private insurance patients that accept the coverage and reimbursement of CMS. Although, there are numerous steps involved in the CMS process, where initially the company must submit a request after which, the NCD (National Coverage Determination) process is initiated. Within 6 months of the NCD including a 30-day public comment period, a proposed decision memo is released in the public followed by another 30-day public comment period. Lastly, after a sixty day period, a final decision on the coverage is released.

The CMS process is very vital for any start-up specifically in the field of personalized wearable device. A start-up in need of revenue will momentarily benefit from CMS as it has potential access to the patients of Medicare and Medicaid. Also, along with the recent healthcare reform, the coverage provided by CMS becomes even more imperative is because it controls access to Medicaid which could insure another 21.3 million Americans in the next decade. That is another 21.3 million potential customers for the personalized healthcare and wearable market. [131] In fact, there is also legislation at the federal level which includes the ACA that will make it easier to access public data on patients, clinical trials, health insurance, and medical advances in the future. Some of these policy levels include the following: the HHS (Health and Human Services) which are starting to liberate data from agency such as CMS due to the 2009 Open Government Directive, the ACA includes a provision that authorized HHS to release data to promote transparency in the markets and health insurance, the HITECH (Health Information Technology for Economic and Clinical Health) Act authorized about \$40 billion in incentive payments for providers to use EMRs (Electronic Medical Records), with an overall goal of driving adoption to 70 to 90 percent of all providers by 2019. [132] With all these incentives, it provides many opportunities for personalized and wearable health care devices include for EMRs along with data continuing to be released to promote transparency in markets and health insurance.

CAN INTELLECTUAL PROPERTY AND OPEN SOURCE CO-HABITATE?

“Intellectual property is an important legal and cultural issue. Society as a whole has complex issues to face here: private ownership vs. open source, and so on” - Tim Berners-Lee

A patent gives its owner the right to exclude others from making, using and selling the claimed invention, whereas an open source license gives anyone who obtains a license to a particular product or software to use and view the code and modify it for his own purpose without the permission of the author. Samsung's Simband is an open source sensor platform for healthcare and wearable technology. Simband is a first prototype, which can utilize its sensors to measure bioimpedance. Simband's arsenal of sensors generates a mountain of data which requires a robust, purpose-built software engine to make sense of it. [133] SAMI is a cloud-based open software platform which makes more information available, breaks open information silos and gives applications and services access to large amounts of data to provide better insights. [133] Simband along with SAMI are intended to be used by the medical industry both in academia as well as by startups to develop new applications and software for sensor technology. Samsung ambitioned into the development of Simband and SAMI to empower a broader shift of wearables from fitness tracking to health monitoring with the goal of enabling preventive healthcare and wellness. The Simband is equipped with six sensors, but it gives flexibility to developers and innovators to add their own proprietary sensor. The six sensors measure daily steps, heart rate, blood pressure, skin temperature and the production of sweat. Simband's sensors are able to track bodily functions at an impressively detailed level. [134] For example, the optical sensor tracks the blood pressure. It uses LED light to track changes in how the light is absorbed by our blood, which allows it to “detect blood-volume change at a microvascular level.” [134] Another example is the Galvanic skin sensor (GSR), which measures the sweat thus, gauging the stress level of our body. [134] The Simband's sensors sense and display a real-time feed of all the stats being collected. [134] The data collected from Simband can also be traced on the web when connected to wifi. However, the bigger question is how these sensors will evolve, how Samsung's cloud will use the data, and how other researchers and companies will be able to develop tools for Simband. Will the Simband's universal platform be a boon or a curse to the future of wearable technology? Simband technology also gives us perspective in acknowledging the fact of how giant of a knoll wearable health tech still has to ascent.

Tesla Motors CEO, Elon Musk recently announced on the company's blog that it would make its patents open to the public, which would allow the automotive community to use its patents and technology in “good faith”. So what does “good faith” use really mean? Musk failed to publicly clarify its exact definition. Moreover, it does not provide a blanket and unmitigated protection against litigation and lawsuit for potential patent users. So is this Tesla's and Samsung's marketing and PR move? Undeniably. An excellent and first-rate business move? Indubitably. This is evident from Samsung's filing of 4652 patents in 2013, where it ranked number 2 in filing at the United States Patent and Trademark Office (USPTO). [135] Additionally, it could also mean that by opening up the Tesla patents to the general public, Tesla wants more automotive manufactures to invest in electric cars, which would in turn lead

to additional and easily accessible front-end infrastructure for electric cars for example, more electric gas stations thus, solving a giant and forthcoming challenge of global supply chain needed for electric cars to run on roads. So does this make Tesla steer into capitalism with a conscience or is this a simply an exceptional and a unique business deal along with a remarkable and a noteworthy public relations move?

Furthermore, OptiProbe's patent portfolio as suggested above is unquestionably going to bring in significant revenue to the company. So should OptiProbe follow the steps of Tesla and Samsung and lean towards making the patents, devices and platforms open for the public to use? That is the consummate question, which conflicts public's opinion about innovation, novelty, inspiration and creativity. What are the consequences of an open source movement? Is it going to halt innovation and novelty? Only time will tell if the future of wearable technology is the future of smart and shrewd business. Till then, we can embrace cutting edge wearable technology that will help us make smarter and healthier personal decisions.

BIG DATA AND PRIVACY: BIG CHALLENGE OF THE FUTURE

"In digital era, privacy must be a priority. Is it just me, or is secret blanket surveillance obscenely outrageous?" - Al Gore

Big data drives big benefits, from innovative businesses to new ways to treat diseases. [136] The Obama administration in its second term have pushed for a digital government and making health data easily accessible to the public. One of the major tools in the information technology tool box is via the use of APIs. APIs make information more accessible. [137] First, the use of APIs makes it easier to replicate government information across more places than ever before. APIs enable automatic updates of information when content is syndicated on other websites, while reducing actual person hours currently spent manually updating content. [137] Second, APIs make information and data easily available to developers, who can create web and mobile applications that make information increasingly more useful to the public. [137] Creating these automated connection points between the data and external computer systems will ease the data's use and its transfer to outside systems creating true data liquidity to fuel algorithms, support data visualizations, or be used in other tools and services. [138] The government is striding towards a digital government and making information and data more liquid, accessible, manageable and useful, so that the public can use the data in a more meaningful way.

The next big challenge for the government is the privacy issues that ascend because of easy access to big data. The challenges to privacy arise because technologies collect so much data (e.g., from sensors in everything from phones and medical devices to parking lots) and analyze them so efficiently (e.g., through data mining and other kinds of analytics) that it is possible to learn far more than most people had anticipated or can anticipate given continuing progress. [136] Technology by itself cannot protect privacy, and policy intended to protect privacy needs to reflect what is (and is not) technologically feasible. [136] The term "privacy" encompasses not only avoiding observation, or keeping one's personal matters and relationships secret, but

also the ability to share information selectively but not publicly. [136] The promise of big-data collection and analysis is that the derived data can be used for purposes that benefit both individuals and society. [136]

Big data can provide enormous benefits in many areas of technology from surveillance to advertising to medicine, but on the flip side can also be accompanied with several privacy challenges. With an unprecedented growth in the field of personalized medicine, healthcare and wearable technology, there is an exponential increase in the volume of data generated. Additionally, privacy concerns regarding patient's health, histories and records could also be jeopardized. The following are a few examples in the field of personalized medicine and healthcare via mobile devices, where privacy and concealment of data is an emerging concern.

- **Personalized Medicine**

Not all patients who have a particular disease are alike, nor do they respond identically to treatment. [136] Researchers will soon be able to draw on millions of health records (including analog data such as scans in addition to digital data), vast amounts of genomic information, extensive data on successful and unsuccessful clinical trials, hospital records, and so forth. [136] In some cases they will be able to discern that among the diverse manifestations of the disease, a subset of the patients have a collection of traits that together form a variant that responds to a particular treatment regime. [136]

Since the result of the analysis could lead to better outcomes for particular patients, it is desirable to identify those individuals in the cohort, contact them, treat their disease in a novel way, and use their experiences in advancing the research. [136] Their data may have been gathered only anonymously, however, or it may have been de-identified. [136]

Solutions may be provided by specific new technologies for the protection of database privacy. [136] These may create a protected query mechanism so individuals can find out whether they are in the cohort, or provide an alert mechanism based on the cohort characteristics so that, when a medical professional sees a patient in the cohort, a notice is generated. [136]

- **Wearable Technology:**

Many baby boomers wonder how they might detect Alzheimer's disease in themselves. [136] What would be better to observe their behavior than the mobile device that connects them to a personal assistant in the cloud (e.g., Siri or OK Google), helps them navigate, reminds them what words mean, remembers to do things, recalls conversations, measures gait, and otherwise is in a position to detect gradual declines on traditional and novel medical indicators that might be imperceptible even to their spouses? [136]

At the same time, any leak of such information would be a damaging betrayal of trust. What are individuals' protections against such risks? [136] Can the inferred information about individuals' health be sold, without additional consent, to third parties (e.g., pharmaceutical companies)? [136] What if this is a stated condition of use of the app? [136] Should information go to individuals' personal physicians with their initial consent but not a subsequent confirmation? [136]

HIPAA and the Age of Big Data: Old Laws New Problems

As mentioned above, Big Data has tremendous potential in the medical and healthcare industry, but is potentially perilous with respect to the privacy of a patient. By amassing and analyzing massive quantities of digital information from multiple sources, including an emerging class of wearable devices and smartphone apps, medical professionals will be well equipped to solve major health problems and warn people of emerging threats like the Ebola virus. [139] Furthermore, since the beginning of the Obama administration and the turn of the decade, the federal government has used its enormous power to push the medical and healthcare industry to synchronize care, embrace new technologies and cut costs, all in an effort to transition towards a more cost-effective model of healthcare. But representatives from a broad swath of the health economy – providers, insurers, and health information technology professionals – say the same resources haven't been leveraged to update existing regulations or inform the healthcare community about how old rules apply in the new world. They worry that they're increasingly navigating a regulatory minefield with an out-of-date map. [139] One of the most imperative apprehension of these professionals is the Health Insurance Portability and Accountability Act (HIPAA) law which was passed in 1996 by then President Bill Clinton that oversees the transfer of health information data under the threat of fines and legal penalties for data breaches and unsanctioned practices. Amazon's Paul Milsener, pointed out to Congress that outdated aspects of the law are impeding his company's move into the health information technology sector and that Congress needs to work with the Department of Health and Human Services to modernize the implementation of HIPAA so that healthcare providers can readily employ the benefits of cloud computing without a compromise of the strong protections HIPAA now affords health information in an effort to accelerate the delivery of new biomedical treatments and cures. [139] With the advances in technology, parts of this law has become obsolete. There is, additionally, the concern that many actors in the wearable space would not be considered covered-entities under the extant HIPAA laws, therefore being shielded of any breaches.

HIPAA is also triggering dyspepsia among the medical community. The law was passed when the doctors were not likely to be carrying around thousands of patient files in the palm of their hands. The law still levies penalties based on the volume of data that's been compromised in a security breach. [139]

The increase in wearable technology in the near future also poses a significant challenge between Tele-health and HIPAA laws. Consultations via FaceTime, videoconferences and Skype are not via a secure network and the breach of information would jeopardize the physicians practice. This causes reluctance among doctors to adapt to novel technologies out of apprehension over potential penalties stemming from the unintentional misuse of digital platforms.

The next big question in the era of cyber and digital world is who owns the big data? As more and more data is generated, storage of this data is amplified. When we talk of about data ownership, we refer to the storage of the generated data. Thus, 'my data' have multiple owners, and the number of owners increases with each share. [140] This arises another question of who possesses the right to use the information and data which is collected that truly does not belong to one's self. This is an issue that transcends borders of commerce, ethics, and morals, leading to privacy issues and protection of privacy. [140]

The fourth Amendment states, “The right of the people to be secure in their persons, houses, papers, and effects...” puts only the physical body in the rhetorically more prominent position, where individuals’ privacy is physically protected in the confines and boundaries of their house. The prevailing interpretation of the Fourth Amendment is inadequate for the present-day digital world. We, along with the “papers and effects” contemplated by the Fourth Amendment, live increasingly in cyberspace, where the physical boundary of the home has little relevance. [136] Hence, it becomes extremely imperative that law makers as well as policy and decision makers at the federal and state level address these issues without delay rather than address them when harm ensues.

Privacy has a noteworthy anthropological value. The advances of technology both threatens personal privacy and provide opportunities to enhance its protection. [136] The challenge for the U.S. Government and the larger community, both within this country and globally, is to understand what the nature of privacy is in the modern world and to find those technological, educational, and policy avenues that will preserve and protect it. [136] Furthermore, governments globally will need to play a more active role to protect citizens’ privacy rights, in light of the evolving and digital world we live in today. [140] Notwithstanding, the break-neck pace of development in the wearable health sector promises to usher in a healthcare era in which the inalienable right to personally tailored healthcare can be achieved by anyone, anytime, and anyplace. Innovations will be inextricably intertwined on the branches and nodes of the ever-growing tree of healthcare technology. In particular, the wearable healthcare sector will sprout strategic assets that will, undoubtedly, require a crafted patent strategy and roadmap. Strategic guideposts and a patent roadmap will be necessary to circumvent the myriad of legal pitfalls in order to properly leverage the low-hanging fruit of wearable biosensor innovations.

ACKNOWLEDGMENTS

We are exceptionally grateful to Dr. Yanjiao Zhou, an instructor at Washington University School of Medicine, St. Louis, for her valuable insight and critical reading of the manuscript. She was one of the leaders in analyzing the Human Microbiome Project metagenomics data studies and is trained in the fields of bioinformatics, microbiology and immunology. We are exceptionally grateful and appreciative to Dr. Hemraj Juwarker, a research fellow at Duke University Medical Center, where he is involved in developing novel therapeutics for chronic inflammation and thrombosis. We are also grateful to Michael Andraos and Carina Kraatz, both associates at Fernandez & Associates, LLP, for their valuable insight and discussions during the course of our research and for their critical reading and analysis of the manuscript. We are also thankful to Dennis Fernandez, managing partner at Fernandez & Associates, LLP, for giving us the opportunity to write this article and for his insightful and judicious suggestions.

REFERENCES

- [1] Idris, Kamil. *"Intellectual Property: A Power Tool for Economic Growth."* World Intellectual Property Organization (WIPO), 1 June 2003. Web.

- [2] Tom Myers, "Patenting and Financing Bioinformatics Inventions" (*Symposium on Bioinformatics and Intellectual Property Law, April 27, 2001*) 1 B.U. J. Sci. & Tech. L. 157.
- [3] Walker, et al., Primer: Genomic and Proteomic Tools for the Molecular Dissection of Disease, *Nat Clin Pract Rheumatol*. 2007 Oct;3(10):580-9.
- [4] Schubarth, Cromwell. "Wearables Are Hot: Here's Who Invested and Who Got the Most VC Funding - Silicon Valley Business Journal." *Silicon Valley Business Journal*. 14 Sept. 2014. Web.
- [5] Gormley, Brian. "Gene Therapies Push Investors to Examine New Areas of Human Body." *Venture Capital Dispatch* RSS. 14 Sept. 2014. Web.
- [6] Novelli, G. Personalized genomic medicine. *Int Emerg Med*. 2010;5(Suppl 1):S81-90. doi:10.1007/s11739-010-0455-9.
- [7] Paving the way of personalized medicine. FDA's Role in a New Era of Medical Product Development. *US Dept of Food and Drug Administration Report*, October 2013.
- [8] Guan, Yan-Fang et al. "Application of Next-Generation Sequencing in Clinical Oncology to Advance Personalized Treatment of Cancer." *Chinese Journal of Cancer* 31.10 (2012): 463–470.
- [9] Welch JS, Link DC. Genomics of AML: clinical applications of next generation sequencing. *Hematology Am Soc Hematol Educ Program*, 2011, 2011: 30-35.
- [10] Marchetti A, Del GM, Filice G, et al. Complex mutations & subpopulations of deletions at exon 19 of EGFR in NSCLC revealed by next generation sequencing: potential clinical implications. *PLoS One*, 2012, 7: e42164.
- [11] Lev-Ari, Aviva. "Sunitinib Brings Adult Acute Lymphoblastic Leukemia (ALL) to Remission - RNA Sequencing - FLT3 Receptor Blockade." 9 July 2012. Web. <<http://pharmaceuticalintelligence.com/2012/07/09/sunitinib-brings-adult-all-to-remission-rna-sequencing/>>.
- [12] National Research Council: Committee on a Framework for Developing a New Taxonomy of Disease. (2011). *Toward Precision Medicine: Building a Knowledge Network for Biomedical Research and a New Taxonomy of Disease*. Washington, DC: The National Academies Press.
- [13] Casey Lynnette Overby & Peter Tarczy-Hornoch, Personalized medicine: challenges and opportunities for translational bioinformatics, *Per Med*. Jul 1, 2013; 10(5): 453–462.
- [14] Pitts, Peter J., and Jeffery W. Young Jr. "One-size Fits All Medicine?" Web. 13 Dec. 2014. <<http://www.cmpi.org/PDFs/Reports/onesize.pdf>>.
- [15] Tom Myers, "Patenting and Financing Bioinformatics Inventions" (*Symposium on Bioinformatics and Intellectual Property Law, April 27, 2001*) 1 B.U. J. Sci. & Tech. L. 157.
- [16] Walker, et al., Primer: Genomic and Proteomic Tools for the Molecular Dissection of Disease, *Nat Clin Pract Rheumatol*. 2007 Oct; 3(10): 580-9.
- [17] Hong H X, Zhang W Q, Shen J, et al. Critical role of bioinformatics in translating huge amounts of next-generation sequencing data into personalized medicine. *Sci China Life Sci*, 2013, 56: 110–118, doi: 10.1007/s11427-013-4439-7.
- [18] "Bioinformatics: Where Lab-science Meets Information Science." *Bioinformatics*. Body Sync Personal Genetics. Web. <<https://bodysync.com/bioinformatics-technology.aspx>>.
- [19] Vullers, Rudolf. "Wearable Smart Systems for Personal Healthcare and Lifestyle." imec, The Netherlands. imec/Holst Centre, Lecture.

- [20] Wearable Technology- Market Assessment- An IHS White Paper (Sept. 2013).
- [21] National Research Council: Committee on a Framework for Developing a New Taxonomy of Disease. (2011). *Toward Precision Medicine: Building a Knowledge Network for Biomedical Research and a New Taxonomy of Disease*. Washington, DC: The National Academies Press.
- [22] Clarke, Peter. "Sensors for Wearables Market to Double in 2015." *EETimes*. 22 Oct. 2014. Web.
- [23] "Wearable Electronics Market and Technology Analysis (2013 - 2018): By Components (Sensors, Battery, Display, Networking); Applications (Consumer, Healthcare, Enterprise); Products (Smart -Textiles, Glasses, Watches);e-Materials & Geography." 1 Apr. 2013. Web. <<https://www.asdreports.com/shopexd.asp?id=73716>>.
- [24] Bothun, Dehorah et al. "The Wearable Future." 1 Jan. 2014. Web. <http://www.pwc.com/es_MX/mx/industrias/archivo/2014-11-pwc-the-wearable-future.pdf>.
- [25] PWC Health Research Institute, *Health wearables: Early days*, page 1. (2014, January 1). Retrieved January 1, 2014, from http://www.pwc.com/en_US/us/health-industries/top-health-industry-issues/assets/pwc-hri-wearable-devices.pdf.
- [26] Takahashi, Dean. "Intel Capital Invests \$62M in 16 Tech Startups." *VentureBeat*. 4 Nov. 2014. Web.
- [27] "Health and Fitness Wearable Sensors." *InvenSense*. Web. <<http://www.invensense.com/mems/wearable-sensors.html>>.
- [28] *Id.*
- [29] Patel et al., A review of Wearable Sensors and Systems in Rehabilitation, *Journal of NeuroEngineering and Rehabilitation* 2012, 9:21.
- [30] Atallah, et al., *Sensor Placement for Activity Detection Using Wearable Accelerometers*, Dept. of Computing, Imperial College London.
- [31] Che-Chang Yang and Yeh-Liang Hsu, A Review of Accelerometry-Based Wearable Motion Detectors for Physical Activity Monitoring, *Sensors* 2010, August 2010.
- [32] Chakravarthy, Jayanth. "UP3: The First Truly Wearable Heart Rate Monitor - The Jawbone Blog." *The Jawbone Blog*. 4 Nov. 2014. Web. <<https://jawbone.com/blog/up3-wearable-heart-rate-monitor/>>.
- [33] *Id.*
- [34] Lee et al., A Low-power and Compact-sized Wearable Bio-impedance Monitor with Wireless Connectivity, *J. Phys.: Conf. Ser.*: 434.
- [35] Vanhemert, Kyle. "Jawbone's New Wearable, the Up3, Is Ambitious in All the Right Places." *Wired.com*. Conde Nast Digital, 5 Nov. 2014. Web. <<http://www.wired.com/2014/11/jawbones-new-wearable-up3-ambitious-right-places/>>.
- [36] Ford, Omar. "Vital Connect's HealthPatch next Step in Wearable Biosensor Tech." *Vital Connect*. 4 Oct. 2013. Web. <<http://www.vitalconnect.com/news/vital-connects-healthpatch-next-step-in-wearable-biosensor-tech>>.
- [37] Scheller F. W, et al. Future of Biosensors: A personal view. *Advances in Biochemical Engineering/Biotechnology*, Volume 140, 2014, pp 1-28.
- [38] "Biosensor." Wikipedia. Wikimedia Foundation. Web. <<http://en.wikipedia.org/wiki/Biosensor>>.
- [39] Hayat Akhtar et al., Disposable screen printed electrochemical sensors: Tools for environmental monitoring, *Sensors*, 2014, pg 10432-53.

- [40] Thiagarajan, N.; Chang, J.-L.; Senthilkumar, K.; Zen, J.-M. Disposable electrochemical sensors: A mini review. *Electrochem. Commun.* 2014, 38, 86–90.
- [41] Honeychurch, K.C.; Hart, J.P. Screen-printed electrochemical sensors for monitoring metal pollutants. *TrAC, Trends Anal. Chem.* 2003, 22, 456–469.
- [42] Hayat, A.; Andreescu, S.; Marty, J.-L. Design of PEG-aptamer two piece macromolecules as convenient and integrated sensing platform: Application to the label free detection of small size molecules. *Biosens. Bioelectron.* 2013, 45, 168–173.
- [43] Hayat, A.; Haider, W.; Rolland, M.; Marty, J.-L. Electrochemical grafting of long spacer arms of hexamethyldiamine on a screen printed carbon electrode surface: Application in target induced ochratoxin A electrochemical aptasensor. *Analyst* 2013, 138, 2951–2957.
- [44] Metters, J.P.; Kadara, R.O.; Banks, C.E. New directions in screen printed electroanalytical sensors: An overview of recent developments. *Analyst* 2011, 136, 1067–1076.
- [45] Hart, J.P.; Wring, S.A. Recent developments in the design and application of screen-printed electrochemical sensors for biomedical, environmental and industrial analyses. *TrAC, Trends Anal. Chem.* 1997, 16, 89–103.
- [46] Hayat, A.; Marty, J.-L.; Radi, A.-E. Novel Amperometric Hydrogen Peroxide Biosensor Based on Horseradish Peroxidase Azide Covalently Immobilized on Ethynyl-Modified Screen-Printed Carbon Electrode via Click Chemistry. *Electroanalysis* 2012, 24, 1446–1452.
- [47] Iozzio, Corinne. "Power to the Internet of Things." *Scientific American Global RSS*. 18 Nov. 2014. Web. <<http://www.scientificamerican.com/article/power-to-the-internet-of-things/>>.
- [48] *Id.*
- [49] Ma, Michelle. "No-power Wi-Fi Connectivity Could Fuel Internet of Things Reality." *UW Today*. 4 Aug. 2014. Web. <<http://www.washington.edu/news/2014/08/04/no-power-wi-fi-connectivity-could-fuel-internet-of-things-reality/>>.
- [50] Anderson, Mark. "WiFi Backscatter: The Internet of Things Could Talk by Turning Reflective." *IEEE Spectrum*, 13 Aug. 2014. Web. <<http://spectrum.ieee.org/tech-talk/computing/networks/internet-of-things-talks-by-turning-reflective>>.
- [51] "Cardiio." - *Touchless Heart Rate Monitor for Cardio Health and Fitness*. Web. <<http://www.cardiio.com/>>.
- [52] "AliveCor-For Physicians." *AliveCor*. Web. <<http://www.alivecor.com/for-physicians>>.
- [53] "CellScope to Build a Smartphone-enabled Scope Toolkit for Diagnostics." *CellScope Oto*. Web. <<https://www.cellscope.com/>>.
- [54] David Kender, Reviewed.com / USA TODAY. "Implantable Technology Will Get under Our Skin." *USA Today*. Gannett, 27 Mar. 2014. Web. <<http://www.usatoday.com/story/tech/reviewed-com/2014/03/27/implantable-tech-is-the-next-wave/6914363/>>.
- [55] AK, Sheela. "Report: Wearable Medical Devices Market Is Expected to Reach \$5.8B Globally in 2019 - Drugs.com MedNews." *Report: Wearable Medical Devices Market Is Expected to Reach \$5.8B Globally in 2019 - Drugs.com MedNews*. 10 Dec. 2013. Web. <<http://www.drugs.com/news/report-wearable-medical-devices-market-expected-reach-5-8b-globally-2019-49476.html>>.
- [56] Aspler, Sarah. "How Ingestible Sensors and Smart Pills Will Revolutionize Healthcare - MaRS." *MaRS*. 26 Sept. 2014. Web.

- [57] De'Medici, Laurenti. "Top 10 Implantable Wearables Soon To Be In Your Body." *WT VOX*. 20 Oct. 2014. Web. <<https://wtvox.com/2014/10/top-10-implantable-wearables-soon-body/>>.
- [58] Kinkead, Gwen. "A Contraceptive Implant with Remote Control." *MIT Technology Review*, 4 July 2014. Web. <<http://www.technologyreview.com/news/528121/a-contraceptive-implant-with-remote-control/>>.
- [59] Afshar, Vala. "Wearable Technology: The Coming Revolution in Healthcare." *The Huffington Post*. TheHuffingtonPost.com, 4 July 2014. Web. <http://www.huffingtonpost.com/vala-afshar/wearable-technology-the-c_b_5263547.html>.
- [60] "Vivametrica and SensorUp Join Forces to Create Industry Standards for Wearable Device Data Collection." 21 Oct. 2014. Web.
- [61] McLellan, Charles. "Vivametrica Unveils Open Source Healthcare Analytics Platform for Wearables." *ZDNet*. 16 Sept. 2014. Web.
- [62] *Id.*
- [63] Weinberger, Matt. "The Samsung Simband Is Cool, but the Health Platform Is Cooler." *CITEworld*. 28 May 2014. Web.
- [64] Yang Li, et al., *Building a Cloud-Based Platform for Personal Health Sensor Data Management*, pg. 1, <http://www.doc.ic.ac.uk/~yl4709/papers/Building-a-Cloud-based-Platform-for-Personal-Health-Sensor-Data-Management.pdf>.
- [65] *Id.* at 2.
- [66] *Id.* at 3.
- [67] *Id.*
- [68] "Catalog API." [www.healthdata.gov](http://healthdata.gov/catalog-api). Web. <<http://healthdata.gov/catalog-api>>.
- [69] "Developer Challenges Overview: Evolving the HealthData.gov Platform." [www.healthdata.gov](http://healthdata.gov/developer-challenges-overview). Web. <<http://healthdata.gov/developer-challenges-overview>>.
- [70] "Galileo Analytics." 24 Oct. 2014. Web. <<http://www.galileoanalytics.com/author/annaslipp/>>.
- [71] *Id.* at 28.
- [72] "More Than 30 Billion Devices Will Wirelessly Connect to the Internet of Everything in 2020." 9 May 2013. Web. <<https://www.abiresearch.com/press/more-than-30-billion-devices-will-wirelessly-conne>>.
- [73] "Cisco Visual Networking Index: Forecast and Methodology, 2013–2018." *Cisco*. 10 June 2014. Web. <http://www.cisco.com/c/en/us/solutions/collateral/service-provider/ip-ngn-ip-next-generation-network/white_paper_c11-481360.html>.
- [74] <http://rockhealth.com/resources/blog/>.
- [75] <http://rockhealth.com/tag/wearables/>.
- [76] *Hanover Research Report*, The Rise of the Wearable Tech Market, March 13, 2014 <http://www.hanoverresearch.com/insights/the-rise-of-the-wearable-tech-market/?i=food-beverage>.
- [77] *WIPO Magazine*, March 2014.
- [78] "IFI CLAIMS® 2013 Top 50 US Patent Assignees." *IFI CLAIMS® Patent Services*. 16 Jan. 2014. Web. <http://www.ificlaims.com/index.php?page=misc_top_50_2013>.
- [79] Roberts, Jeff John. "Google's Patent Rank Shoots up as Company Stakes out Wearable Tech." *Gigaom*. 13 Jan. 2014. Web.

- [80] World Intellectual Property Organization. "The Brave New World of Wearable Technology: What Implications for IP?" Web. 13 Nov. 2014. http://www.wipo.int/wipo_magazine/en/2014/03/article_0002.html.
- [81] Musil, Steven. "Apple Scores Courtroom Wins in Motorola FRAND Case - CNET." *CNET*. 13 Aug. 2012. Web.
- [82] Von Florian Mueller, Eingestellt. "U.S. Court Grants Apple Partial Summary Judgment on FRAND Contract Issues in Motorola Case." *Foss Patents*. 13 Aug. 2012. Web. <<http://www.fosspatents.com/2012/08/us-court-grants-apple-partial-summary.html>>.
- [83] Trenholm, Rich. "Apple Talks to the FRAND, Reopens Motorola Patent Fight - CNET." *CNET*. 9 Apr. 2014. Web.
- [84] Decker, Susan, and Ian King. "Qualcomm Wireless Grip Risked as Intel, Microsoft Seek New Rules." *Bloomberg.com*. Bloomberg, 3 Dec. 2014. Web.
- [85] "How to License your Patent." *National Inventor Fraud Center, Inc.* Web. <<http://www.inventorfraud.com/License.aspx>>.
- [86] While litigation has been limited to just one thus far (Addidas v. Under Armour over its MapMyFitness App), litigation rates will likely parallel the rates of other high-tech sectors.
- [87] "The Brave New World of Wearable Tech: What Implications for IP?" World Intellectual Property Organization (WIPO), 3 June 2014. Web. <http://www.wipo.int/export/sites/www/wipo_magazine/en/pdf/2014/wipo_pub_121_2014_03.pdf>.
- [88] There were 5,189 patent actions in 2012, dramatically up from approximately 1,000 cases in 1991.
- [89] Barry, Chris, Ronen Arad, Kristofer Swanson, and Evan Clark. "2013 Patent Litigation Study: Big Cases Make Headlines, While Patent Cases Proliferate." PWC, 1 June 2013. Web. <http://www.pwc.com/en_US/us/forensic-services/publications/assets/2013-patent-litigation-study.pdf>.
- [90] O'Brien, Vince. "Patent Valuation Methodologies." Web. <<https://www.law.berkeley.edu/files/obrien.pdf>>.
- [91] "Valuing Patents • Intangible Business US." Web. <<http://www.intangiblebusiness.us/brand-services/legal-services/valuing-patents>>.
- [92] *Id.*
- [93] "INFOGRAPHIC: Patent Valuation." *ClearViewIP*. 13 Mar. 2013. Web. <<http://www.clearviewip.com/publications/patent-valuation-infographic/>>.
- [94] <http://www.ipofferings.com/drawings/patentvaluation/IPOfferings%20Patent%20Valuation%20-%20EnhancedrReport%20Sample.US.6584450.pdf>.
- [95] *Id.* at 7.
- [96] *Id.* at 8.
- [97] *Id.*
- [98] *Id.*
- [99] Kappos, David. "Ten Tips for Streamlining Patent Prosecution." *Director's Forum: A Blog from USPTO's Leadership*. 28 June 2010. Web. <http://www.uspto.gov/blog/director/entry/ten_tips_for_streamlining_patent>.
- [100] Fernandez, Dennis. "Intellectual property protection for bioinformatics and computational intelligence." *IEEE Symposium*. IEEE., Singapore. 16 Apr. 2013. Lecture.

- [101] Conley, John. "Myriad, Finally: Supreme Court Surprises by Not Surprising." *Genomics Law Report* 13 June 2013. Web. <<http://www.genomicslawreport.com/index.php/2013/06/18/myriad-finally-supreme-court-surprises-by-not-surprising/>>.
- [102] Rantanen, Jason. "The Supreme Court's Alice Decision on Patent Eligibility of Computer-Implemented Inventions: Finding an Oasis In the Desert." *PatentlyO*. 23 June 2013. Web.
- [103] Hirshfeld, Andrew H. "Preliminary Examination Instructions in View of the Supreme Court Decision in *Alice Corporation Pty. Ltd. v. CLS Bank International, Et Af.*" United States Patent and Trademark Office, 25 June 2014. Web. <http://www.uspto.gov/patents/announce/alice_pec_25jun2014.pdf>.
- [104] *Id.*, slip op. at 15.
- [105] Plotkin, Robert. "A Post-Alice Playbook: Practical Strategies for Responding to Alice-Based Rejections - IPWatchdog.com | Patents & Patent Law." *IPWatchdog*. www.ipwatchdog.com, 7 Oct. 2014. Web. <<http://www.ipwatchdog.com/2014/10/07/a-post-alice-playbook-practical-strategies-for-responding-to-alice-based-rejections/id=51577/>>.
- [106] *Id.*
- [107] Quinn, Gene. "Examiners Begin Issuing Alice Rejections for Software." *IPWatchdog.com*. 14 July 2014. Web. <<http://www.ipwatchdog.com/2014/07/14/examiners-begin-issuing-alice-rejections-for-software/id=50412/>>.
- [108] Locke, Scott D., and David A. Kalow. "Buffalo Intellectual Property Law Journal." *Preparing for Bioinformatics Litigation: How Will the Courts Confront the next Generation of Biotechnology Patents?* 1 (2001): 76-94. Web. <<http://wings.buffalo.edu/law/biplj/biplj113.pdf>>.
- [109] "IBM Extends Hyperactive IP Strategy to Life Science IT: Is Litigation Far Behind?" *GenomeWeb*. 20 Jan. 2003. Web. <<https://www.genomeweb.com/informatics/ibm-extends-hyperactive-ip-strategy-life-science-it-litigation-far-behind>>.
- [110] Martini, Antonella, Francesco Paolo Appio, Gualtierio Fantoni, and Riccardo Aprea. "At the Roots of an Interdisciplinary Radical Innovation: Preliminary Insights from the Patenting Activity in the Bioinformatics Industry." ISBN 978-90-77360 17 (2014): 85-107. Web.
- [111] *Id.* at 93.
- [112] Achiriloaie, Mirceae, and Dennis Fernandez. "Keeping Up Intellectual Property Lifelines for Life Sciences Ventures." *www.iploft.com. Suffolk University Law School Journal of High Technology Law*. Web. <<http://www.iploft.com/Mirc-Article.pdf>>.
- [113] Kesslen, Mark P. "Thinking Strategically to Achieve the Client's Goals." *Structuring Patent Licensing Transactions*. Vol. ISBN-10: 1-59622-626-9. Aspatore. Print.
- [114] Weinschenk, Carl. "Qualcomm Licensing Practices Under Fire." *IT Business Edge*. 4 Dec. 2014. Web.
- [115] "Ownership/Assignability of Patents and Applications [R-11.2013]." United States Patent and Trademark Office, 1 Dec. 2014. Web. <<http://www.uspto.gov/web/offices/pac/mpep/s301.html>>.
- [116] Henry MR, Cho MK, Weaver MA, Merz JF: DNA patenting and licensing. *Science* 2002;297:1279.

- [117] Jonathan E. Kemmerer, CPA, Jiaqing Lu, Ph.D., CFA, Profitability and Royalty Rates Across Industries: Some Preliminary Evidence, *Annual Conference of the International Academy of Business and Economics*, 2008.
- [118] Robert Goldsneider, et al., *The Use of the 25 Per Cent Rule in Valuing IP*, les Nouvelles, Dec. 2002, pg. 123.
- [119] *Id.*
- [120] "UCSC Genome Browser Store." Web. <<https://genome-store.ucsc.edu>>.
- [121] "NIH Start-Up Exclusive License Agreements." *NIH*. www.ott.nih.gov. Web. <<http://www.ott.nih.gov/nih-start-exclusive-license-agreements>>.
- [122] *Id.*
- [123] Decker, Susan, and Ian King. "Qualcomm Wireless Grip Risked as Intel, Microsoft Seek New Rules." Bloomberg.com. Bloomberg, 3 Dec. 2014. Web.
- [124] Gentry Gallery v. Berkline Corp., 134 F. 3d 1473, 1476 (Fed. Cir. 1998).
- [125] Conley, John. "Myriad Back in Court Again - This Time as a Defendant." *Genomics Law Report RSS*. 8 Oct. 2013. Web. <<http://www.genomicslawreport.com/index.php/2013/10/08/myriad-back-in-court-again-this-time-as-a-defendant/>>.
- [126] Ellison, Eldora, and Jeremiah Frueauf. "10 Things To Know About Post-Grant Review Under The America Invents Act of 2011." www.skgf.com. Sterne, Kessler, Goldstein & Fox P.L.L.C. Web. <http://www.skgf.com/uploads/24/doc/tent_things_to_know_about_post_grant_review.pdf>.
- [127] Feldman, Robin, Tom Ewing, and Sara Jeruss. "Patent Trolling: An Empirical Analysis of Patent Monetization Entities in US Litigation." 1 Jan. 2014. Web. <http://www.hoganlovells.com/custom/LITGM/patent-trolling-analysis_02_24_14.pdf>.
- [128] *Id.*
- [129] Tucker, Catherine, The Effect of Patent Litigation and Patent Assertion Entities on Entrepreneurial Activity (June 22, 2014).
- [130] Roberts, Jeff John. "Watch Out, Tech World: Patent Troll Intellectual Ventures Takes Stake in Wearable and Video Startups." *Gigaom*. 3 Nov. 2014. Web.
- [131] "US Regulations for Medical Devices" Jennifer Mateus. Lecture. 2/13/14.
- [132] "Combination Products: Jurisdictional Information." www.fda.gov, 15 July 2009. Web. <<http://www.fda.gov/combinationproducts/jurisdictionalinformation/default.htm>>.
- [133] "ObamaCare Medicaid Expansion." *Obamacare Facts*. 3 Apr. 2014. Web. <<http://obamacarefacts.com/obamacares-medicaid-expansion/>>.
- [134] "A Guide to the Supreme Court's Decision on the ACA's Medicaid Expansion." *A Guide to the Supreme Court's Decision on the ACA's Medicaid Expansion*. Kaiser Family Foundation, 1 Aug. 2012. Web. <<http://kff.org/health-reform/issue-brief/a-guide-to-the-supreme-courts-decision/>>.
- [135] "The 'big data' revolution in healthcare". McKinsey&Company, Peter Groves, Basel Kayyali, David Knott, Steve Kuiken. *The_big_data_revolution_in_healthcare.pdf*.
- [136] Poeter, Damon. "Meet Simband, Samsung's Next-Gen Health Tracker." *PCMAG*. 28 May 2014. Web.
- [137] Bell, Karissa. "First Look at Simband, Samsung's Health-tracking Wearable of the Future." *Mashable*. 14 Nov. 2014. Web.

- [138] "Top 300 Organizations Granted U.S. Patents in 2013. Are More Patents Better?" www.ipo.org. 6 June 2014. Web. <http://www.ipo.org/wp-content/uploads/2014/06/2013-Top-300-Patent-Owners_5.9.141.pdf>.
- [139] "Big Data and Privacy: A Technological Perspective." www.whitehouse.gov, 1 May 2014. Web. <http://www.whitehouse.gov/sites/default/files/microsites/ostp/PCAST/pcast_big_data_and_privacy_-_may_2014.pdf>.
- [140] Kivak, Bryan. "Making Information More Accessible, The Digital Government Strategy." www.healthdata.gov, 22 Aug. 2012. Web. <<http://healthdata.gov/blog/making-information-more-accessible-digital-government-strategy>>.
- [141] Davis, Damon. "Health Data Initiative Strategy & Execution Plan Released and Ready for Feedback." www.healthdata.gov, 23 Oct. 2013. Web. <<https://healthdata.gov/blog/health-data-initiative-strategy-execution-plan-released-and-ready-feedback>>.
- [142] Easley, Jonathan. "Uncharted Territory: HIPAA In the Big Data Era - Morning Consult." *Morning Consult*. 27 July 2014. Web.
- [143] Al-Khouri, Ali M., Data Owner: Who owns "My Data"? *International Journal of Management & Information Technology*, Volume 2, No. 1, November 2012.

Chapter 29

AVOIDANT/RESTRICTIVE FOOD INTAKE DISORDER IN A FEMALE PATIENT AFFECTED BY MARFAN SYNDROME

A. P. Verri², A. Serio¹, M. Grasso¹, I. Brega² and F. Clerici²

¹Centre for Inherited Cardiovascular Diseases, IRCCS,
Fondazione Policlinico San Matteo, Pavia, Italy

²Fondazione Istituto Neurologico Nazionale C. Mondino,
IRCCS, Pavia, Italy

ABSTRACT

Marfan syndrome is a heritable disorder of connective tissue that is transmitted as an autosomal dominant trait and is characterized by the mutation in the fibrillin 1 gene (FBN1). At present, the diagnosis of Marfan syndrome, based on Ghent criteria, considers both clinical evaluation (in which it is possible to observe alterations in eye, osteoarticular apparatus and cardiovascular system) and genetic evaluation (that reveals a FBN1 gene mutation). Moreover, we take into account the presence of affected relatives suffering of the same syndrome. The diagnosis of Marfan syndrome is often complex because of the evolution of the phenotype with age (Lipscomb, 1997) and because of the inter-individual variation in the clinical presentation even among the affected family members. According to a recent review (Baeza-Velasco, 2014), Marfan syndrome is often associated with a range of psychiatric problems like anxiety disorders, depressive disorders, schizophrenia, neurodevelopmental disorders (autism spectrum disorder and attention deficit/hyperactivity disorder) and eating disorders. We report a 16 years old female patient (MF) (kg 43, H 165 cm, BMI 16.9) who came to our outpatient service for a selective feeding successively diagnosed as Avoidant/Restrictive Food Intake Disorder (ARFID; DSM 5). “Selective feeding” refers to children who restrict the ingestion of food to a limited number of “favourite” foods, typically five or six different foods (Lask and Bryant-Waugh, 2000). ML., indeed, eats only pasta, sweets and a few other foods. Beyond the altered feeding behaviour, she was diagnosed with a mild intellectual disability (IQ=57; VIQ=61; PIQ=62) and reduced adaptive levels in the socialization and communication domains of the Vineland Adaptive Behavior Scales (VABS). This patient presented ligamentous laxity, long-limbed body habitus suggestive for Marfan syndrome. Therefore, she was addressed to the cardiologist: the echocardiographic evaluation showed mild

dilatation of aortic root, with rectilinear sinotubular junction and enlarged ascending aorta (z score=2). Z-scores of the aortic root at the level of the aortic annulus, sinuses of Valsalva, sinotubular junction, and ascending aorta measured from the parasternal long axis in diastole using leading-edge-to-leading-edge technique. The eye examination was normal. Sanger sequencing of the FBN1 gene identified the c.7501G>A (p.Val250Ile) variant/mutation. The missense mutation/variant is not reported in UMD, Ensemble, Exome variant server and dbSNP. The SIFT, PoliPhen2, Mutation Tester (software tool) for predicting damaging of missense mutations and variants gave the following results: PoliPhen2: benign; Mutation Tester: disease causing, SIFT: tolerant. Her mother carries the same mutation. ML. is affected by ARFID, (mild) intellectual disability and adaptive disorders often associated with Marfan syndrome, confirming and extending the results obtained in the review previously described (Baeza-Velasco, 2014). Moreover, we underline the importance in multidisciplinary diagnosis and care (also) in these cases.

INTRODUCTION

1. Marfan Syndrome

Heritable disorders of connective tissue consist of a group of genetic diseases that affect the proteins of the connective tissue matrix such as collagen, elastin, fibrillin and tenascin (Baeza-Velasco, 2014).

Marfan syndrome is a heritable disorder of connective tissue characterized by mutations in fibrillin 1 gene (FBN1) (Dietz, 1991). This protein is present in the ocular, skeletal and cardiovascular systems. At present, the diagnosis of Marfan syndrome, based on Ghent nosological criteria, (Loeys, 2010), considers both clinical evaluation (in which it is possible to observe alterations in eye, osteoarticular apparatus and cardiovascular system) and genetic evaluation (that reveals a FBN1 gene mutation). Moreover, we take into account the presence of affected relatives suffering of the same syndrome transmitted as an autosomal dominant trait. The diagnosis of Marfan syndrome is often complex because of the evolution of the phenotype with age (Lipscomb, 1997) and because of the inter-individual variation in the clinical presentation even among the affected family members. As shown in a recent review (Willis, 2009), the Authors showed that early diagnosis is very important: the patients that received the Marfan syndrome diagnosis before 18 years of age had less heart surgery (33% vs 59%) and have a better outcome.

2. Association between Marfan Syndrome and Psychiatric Disorders

Kaplan and Katz (1992) were the first to underline the possible coexistence of eating disorders and connective tissue diseases. A recent review (Baeza-Velasco, 2014) confirmed the association between heritable disorders of connective tissue and a wide range of psychiatric problems like anxiety disorders, depressive disorders, schizophrenia, neurodevelopmental disorders (autism spectrum disorder, attention deficit/hyperactivity disorder and developmental coordination disorder), eating disorders and personality disorders (anxious obsessive-compulsive personality disorder). Goh and colleagues (2013) affirm that joint hypermobility, peculiar feature of heritable disorders of connective tissue, is “a possible indicator of familial

disorders of connective tissue elasticity, which potentially plays a causal role in the development of the eating disorder” (p.1). Recent studies (Miles, 2007) dealt with the association between anorexia nervosa and specific connective tissue diseases, like Ehlers-Danlos syndrome.

Behaviour problems related to food have been reported in some genetic neurodevelopmental syndromes, particularly in Prader-Willi syndrome, characterized, in fact, by obesity and hyperphagia. Only recently eating disorders have not been specifically investigated in others genetic syndromes, such as Angelman syndrome, the “sister imprinted disorder” of Prader-Willi syndrome, Cornelia de Lange, Fragile X, and 1p36 deletion syndrome (Welham, 2014).

As regards the neurodevelopmental disorders, instead, Shetreat-Klein and colleagues (2014) observed that children with Autism Spectrum Disorder (ASD) had significantly greater joint mobility than their matched normally developing peers. Moreover, cases have been described in which a ASD (precisely, Asperger syndrome) coexisted with lifelong ligamentous laxity and muscular incoordination suggesting a “Marfan-like” disorder of connective tissue (Tantam, 1990).

3. Selective Feeding

Lask and Bryant-Waugh (2000) provided an important contribution to the classification of eating disorders in developmental age. They suggested a number of guidelines, known as Great Ormond Street Criteria (GOS), directed at identifying eating disorders for those aged under 14 years. The GOS include six different clinical conditions: anorexia nervosa, bulimia nervosa, emotional disturbance of food refusal, selective feeding, functional dysphagia and pervasive refusal.

The definition of “selective feeding” refers to children who restrict the ingestion of food to a limited number of “favourite” foods, typically five or six different foods. The diet is rich in carbohydrates, especially bread, chips and cookies. Even the drinks could be selected, they are usually milk or milk-based beverages. The attempt to increase the number of food or beverages meets extreme resistance.

Sometimes, selective feeding is associated with some psychiatric disorders, such as autism spectrum disorders (Bandini, 2010) or complex patterns of behaviour, such as a tendency not to change their habits and the incapability to tolerate new situations.

According to the fifth edition of the Diagnostic and Statistical Manual of Mental Disorders (DSM-5; APA, 2013), a disorder similar to that previously described, is diagnosed as a disorder avoidant / restrictive of food intake (ARFID), characterized by avoidance or restriction in food intake (Criterion A), associated with at least one of the following: relevant weight loss, significant functional deficit, dependence it from parenteral or from oral nutritional supplements, marked interference with psychosocial functioning. This disorder does not include a non-availability of food or a cultural practice (Criterion B).

Also, it is not better explained by an excessive concern about weight and shape of own body (Criterion C) or concomitant medical factors or mental disorder (Criterion D).

4. L.'s Clinical Case

At the age of about 16 years, L. came to our clinical attention because of selective feeding. The patient was used to eat only pasta, sweets and a few other foods for more than two years. Her height was 165 cm and weight 43 kg (BMI = 16.9) thus being underweight. Individuals with selective feeding, in fact, could have a low, normal or high body weight, (Lask and Bryant-Waugh, 2000). According to her medical history, there was a mild perinatal distress after pregnancy complicated by threatened abortion. Subsequently, the patient was bottle-fed with a difficult weaning because of the refusal of solid foods and the preferential intake of milk until about 6 years of age. Furthermore, there was a delay in sphincter control (daytime age of 4 and a half years, night at the age of about 6 years), in motor (independent walking after 16 months), and in language development (first words at about 3 years) and poor communicative intentionality. Moreover, there were emotional-behavioural problems characterized by poor socialization, tendency to play alone, narrow field of interest, motor restlessness, poor compliance and easy distractibility to external stimuli. For the developmental delay at the age of 4, L. was admitted to hospital and the presence of any congenital metabolic diseases was excluded. At the age of 6 years, the Wechsler Preschool and Primary Scale of Intelligence (WIPPSI) allowed the detection of a borderline cognitive level (IQT = 76; IQv = 76; IQP = 81). At the age of 9 years after loss of consciousness episodes in situations of emotional stress (venipunctures for hematochemical test), whereas the an electroencephalogram was normal echocardiogram detected mitral valve prolapse. As a consequence of the persistent emotional-behavioural problems previously described and due to difficulties at school, for which L. was followed by a special education teacher from the third year of nursery school, a clinical and neuropsychological assessment was performed. The clinical neurological examination, at the age of about 12 years, revealed dysmorphic features (mild prominence of the frontal bossing (CC: 55.2 cm, 90-97%), dysmorphic face, pectus excavatum, generalized muscular hypotonia, hypomobile and stereotyped facial expressions, bilaterally claw- feet). In the same year, the Wechsler Intelligence Scale for Children- Revised (WISC-R) detected the presence of mild mental retardation (IQ = 55), with a slightly disharmonious profile between verbal (verbal IQ = of 52) and non-verbal skills (performance IQ = of 65) in favour of the latter. Moreover, autistic traits with a greater endangerment in reciprocal social interaction and communication were present. In particular, the Autism Diagnostic Observation Schedule (ADOS module 3) found that L. minimally uses gestures, facial expression and look for social purposes. Therefore, at the age of about 12 years, there was the diagnosis of "mild mental retardation and autism spectrum disorder (generalized disorder not otherwise specified) in patient with dysmorphic features and minor neurological signs". Two years later, at the age of about 14 years, the clinical and neuropsychological reevaluation confirmed the diagnosis of intellectual disability (IQ = 46; VIQ = 48; PIQ = 56). As for the features of social interaction and communication skills, while pointing out an improvement, the examinations confirmed the previous diagnosis of Pervasive developmental disorder not otherwise specified. At present, in order to investigate the problems of L.'s behaviour and social sphere, psychological interviews were conducted and tests were administered designed to further delineate the cognitive profile and to assess the adaptive behaviour of the patient. The Coloured Progressive Matrices (CPM) detected the presence of the analogical-deductive reasoning ability and of abstraction ability below the norm; the evaluation through Figure of Rey has documented both a visual-spatial organization and a recovery mnemonic ability below the norm, and the administration of the Token test has revealed

an ability of seriation and discrimination below the norm. In addition, the assessment by Wechsler Adult Intelligence Scales-Revised (WAIS-R) showed the presence of mild Intellectual Disability (IQ = 57; VIQ = 61; PIQ = 62). Therefore, there is no clear discrepancy between scores for verbal subtests and those of performance. The firsts showed deficit in the quality of the relationships that the subject has abstracted from her environment, in the ability to use abstract concepts and in learning skills. Instead, from performance tests shortcomings in the planning of sequences and causal events, in the organization of perception and in motor coordination emerged. Scores of different items are shown in Table 1.

As regards the adaptive behaviour, the Vineland Adaptive Behavior Scales (VABS), compiled thanks to the information provided by the mother, found that L. has reached a level of development lower than that the expected for normally developed peers of the same chronological age in the investigated areas (daily skills and socialization). The patient, in fact, knows the value of coins and notes but needs supervision and assistance in the use of money and in savings management. It emerges lack of autonomy in transfers. There are also difficulties in the area of socialization. L., in fact, has not built deep and satisfying personal relationships outside the family. From the comparison with subjects of the same intellectual disability, however, the area most affected appears to be only the one linked to the daily abilities. Scores of different items are shown in Table 2.

The evaluation by Rating Scale behaviour (Aman-Singh) was possible by the information provided by the patient's mother and it has documented the presence of disorders, in particular a tendency to lethargy and a presence of inappropriate language with a tendency to coprolalia. In addition, the score (17/40) from the compilation, by the mother of L., of the Social Communication Questionnaire (QSC), suggested the presence of a disorder in the autism spectrum.

Table 1. WAIS results in our patient

Verbal subtest	Raw score	Weighted score
Information	4	3
Digit span	5	2
Vocabulary	17	3
Arithmetic	3	2
Comprehension	6	3
Similarities	12	5
Verbal score		18
Performance subtest	Raw score	Weighted score
Picture completion	10	6
Picture arrangement	2	3
Block design	9	2
Object assembly	21	6
Digit symbol	39	5
Performance score		22

Table 2. VABS results in our patient

Subtest	Raw score	Age equivalent score
Receptive	45	7 - 10
Expressive	180	9 - 8
Written	40	8 - 0
COMMUNICATION	253	12 - 0
Personal	140	6 - 0
Domestic	21	5 - 2
Community	50	5 - 8
DAILY LIVING SKILLS	211	7 - 9
Interpersonal relationships	80	10 - 6
Play and leisure time	80	
Coping skills	60	
SOCIALIZATION	220	11 - 11
Gross		5 - 6
Fine		5 - 6
MOTOR SKILLS		

However, as regards the assessment of symptoms and food related issues such as body image (Vocks, 2007) and dissatisfaction with their body (Stice, 2002), at the administration of the Eating Attitudes Test (EAT) the presence of altered food attitude did not emerge; the evaluation by Body Attitudes Questionnaire (BAQ) has not detected an altered perception of the body image while the administration of the Eating Disorder Inventory-2 (EDI-2) showed the presence of dissatisfaction with her own body.

These results confirm the fundamental difference between the patients with anorexia nervosa or bulimia nervosa and those with selective feeding, which is the absence in the latter of morbid concerns to the weight and shape of their body and of an altered body image (Lask and Bryant-Waugh, 2000). L. has, therefore, a mild intellectual disability, analogical-deductive reasoning skills and of abstraction, seriation and discrimination skills, visuospatial organization skills and of mnemonic recovery below the norm, behavioural disorders, autism spectrum disorder and selective feeding. Due to the Marfan like phenotype specific investigations were carried out. The echocardiographic evaluation showed mild dilatation of aortic root, with rectilinear sinotubular junction and enlarged ascending aorta (z score=2). Z-scores of the aortic root at the level of the aortic annulus, sinuses of Valsalva, sinotubular junction, and ascending aorta measured from the parasternal long axis in the diastolic phase using leading-edge-to-leading-edge technique. The eye examination was normal. The molecular genetics detected the presence of a mutation of fibrillins 1, a feature of Marfan syndrome, namely: exon 60: c.7501G> A (p.Val2501Ile).

The clinical and genetic study was extended to the family members of the patient. The same mutation was present in the mother. Our patient at age 18 years, finally received the diagnosis of Marfan syndrome associated with mild intellectual disabilities, analogical-deductive reasoning skills and abstraction, seriation and discrimination skills, organization

skills and visuospatial mnestic recovery below the standard, behavior disorders, autism spectrum disorder and selective feeding.

CONCLUSION

A recent review (Baeza-Velasco, 2014) revealed an association between heritable disorder of connective tissue, like Marfan syndrome, and a wide range of psychiatric disorders. Examples of these disorders are neurodevelopmental disorders (particularly autism spectrum disorder and developmental coordination disorder) and eating disorders.

During infancy our patient presented a developmental delay (onset of independent walking after 16 months) associated with clumsiness and Autism Spectrum Disorder (ASD). There is an association between joint hypermobility and motor delay in children (Baeza-Velasco, 2014). Hypotonia is a feature frequently seen in children with Autism Spectrum Disorder (Baeza-Velasco, 2014); a link between ASD and connective tissue disorders has been suggested. The latest version of the Diagnostic and Statistical Manual of Mental Disorders (DSM-5; APA, 2013) stressed functional difficulties in Developmental Coordination Disorder (DCD) and "clumsy" children with joint hypermobility. A selective eating disorder was also described since the first years of life. This is frequently associated with an autism spectrum disorder.

Behaviour problems related to food have been described in some genetic syndromes of neurodevelopment, particularly in Prader-Willi syndrome, characterized, in fact, by obesity and hyperphagia. As of today, eating disorders have not been specifically investigated in other genetic syndromes, such as: Angelman syndrome, the "sister imprinted disorder" of Prader-Willi syndrome, Cornelia de Lange, Fragile X, and 1p36 deletion syndrome (Welham, 2014). The clinical case of L. therefore, allows confirming and extending the results in the literature regarding the association between Marfan syndrome and psychiatric disorders. L. is affected by mild intellectual disabilities, analogical-deductive reasoning skills and abstraction, seriation and discrimination skills, organization skills and visuospatial mnestic recovery below the standard, behaviour disorders, autism spectrum disorder in a subject with selective feeding and Marfan syndrome.

In addition, the clinical case of L., highlights the need for a multidisciplinary approach to diagnosis, as emphasized by the new criteria of Ghent and by the consequent guidelines (Loeys, 2010). The importance of a multidisciplinary treatment depends on the great number of problems associated with Marfan syndrome and the possible presence of comorbid disorders, like in L. clinical case. We hope that further studies will increase knowledge on genetic syndromes and their possible association with psychiatric disorders, noting, if it possible, the causal link between the different problems.

REFERENCES

American Psychiatric Association (2013). *Diagnostic and statistical manual of mental disorders*: 5th ed. Arlington: VA American Psychiatric Publishing: 2013.

- Baeza-Velasco, C., Pailhez, G., Bulbena, A. and Baghdadli, A. Joint hypermobility and the heritable disorders of connective tissue: clinical and empirical evidence of links with psychiatry, *Gen. Hosp. Psychiatry*, 2014.
- Bandini, L. G., Anderson, S. E., Curtin, C., Cermak, S., Evans, E. W., Scampini, R., Maslin, M. and Must, A. (2010). Food selectivity in children with Autism Spectrum Disorders and typically developing children. *J. Pediatr.*, 157(2), 259-264.
- Dietz, H. C., Cutting, G. R., Pyeritz, R. E., Maslen, C. L., Sakai, L. Y., Corson, G. M., Puffenberger, E. G., Hamosh, A., Nanthakumar, E. J. and Curristin, S. M. (1991). Marfan syndrome caused by a recurrent de novo missense mutation in the fibrillin gene. *Nature*, 352(6333), 337-339.
- Goh, M., Olver, J., Huang, C., Millard, M. and O'Callaghan (2013). Prevalence and familial patterns of gastrointestinal symptoms, joint hypermobility and diurnal blood pressure variations in patients with anorexia nervosa. *J. Eat. Disord.*, 1(Suppl. 1), O45.
- Kaplan, A. S. and Katz, M. (1992). Eating disorders and connective tissue disease. Etiologic and treatment considerations. *Psychosomatics*, 33, 105-8.
- Lask, B., and Bryant-Waugh, R. (2000). *Anorexia nervosa and related eating disorders in children and adolescence*: 2nd Edition, Hove, UK: Psychology Press.
- Lipscomb, K. J., Clayton-Smith, J. and Harris, R. (1997). Evolving phenotype of Marfan's syndrome. *Arch. Dis. Child.*, 76(1), 41-46.
- Loeys, B. L., Dietz, H. C., Braverman, A. C., Callewaert, B. L., De Backer, J., Devereux, R. B., Hilhorst-Hofstee, Y., Jondeau, G., Faivre, L., Millewicz, D. M., Pyeritz, R. E., Sponseller, P. D., Wordsworth, P. and De Paepe, A. M. (2010). The revised Ghent nosology for the Marfan syndrome. *J. Med. Genet.*, 47, 476-85.
- Miles, S. C., Robinson, D. P. and Miles, J. L. (2007). Ehlers-Danlos syndrome and anorexia nervosa: a dangerous combination? *Pediatr. Dermatol.*, 24, E1-4.
- Shetreat-Klein, M., Shinnar, S. and Rapin, I. (2014). Abnormalities of joint mobility and gait in children with autism spectrum disorders. *Brain Dev.*, 36, 91-6.
- Stice, E. (2002). Risk and maintenance factors for eating pathology: a meta-analytic review. *Psychological Bulletin*, 128(5), 825-848.
- Tantam, D., Evered, C., Hersov, L. (1990). Asperger's syndrome and ligamentous laxity. *J. Am. Acad. Child. Adolesc. Psychiatry*, 29, 892-6.
- Vocks, S., Legenbauer, T., Watcher, A., Wucherer, M. and Kosfelder, J. (2007). What happens in the course of body exposure? *Journal of Psychosomatic Research*, 62, 231-239.
- Welham, A., Lau, J., Moss, J., Cullen, J., Higgs, S., Warren, G., Wilde, L., Marr, A., Cook, F. and Oliver, C. (2014). Are Angelman and Prader-Willi syndrome more similar than we thought? Food-related behavior problems in Angelman, Cornelia de Lange, Fragile X, Prader-Willi and 1p36 deletion syndrome. *Am. J. Med. Genet. Part A*, 167, 572-578.
- Willis, L., Roosvelt, G. E. and Yetman, A. T. (2009). Comparison of clinical characteristics and frequency of adverse outcomes in patients with Marfan syndrome diagnosed in adulthood versus childhood. *Pediatr. Cardiol.*, 30(3), 289-292.

Chapter 30

**OPTIMIZING OIL PRODUCTION IN *B. NAPUS* BY
GENE STACKING: TRANSGENIC CO-EXPRESSION
OF *DGAT1* AND PARTIALLY-SUPPRESSED *MTPDCK*
CUMULATIVELY ENHANCE SEED LIPID DEPOSITION**

Elizabeth-France Marillia, Tammy Francis and David C. Taylor*

National Research Council of Canada, Saskatoon,
Saskatchewan, Canada

ABSTRACT

We have previously shown that introduction of single genes encoding diacylglycerol acyltransferases (*DGAT1s*) or partially-silenced mitochondrial pyruvate dehydrogenase kinase (*mtPDCK*), each had the capacity to enhance seed oil content in *Arabidopsis*. In the current study, we report the cumulative effects of expressing a two-gene stack: a site-directed mutagenized *DGAT1* from *Tropaeolum majus* (Tm*DGAT1* Ser¹⁹⁷-to-Ala¹⁹⁷) and an anti-sense *mtPDCK* from *B. napus* (*A/S mtPDCK*) introduced into *B. napus* cultivar DH12075 to alter seed oil content. Compared to plasmid-only controls, the best lines of the two-gene construct showed, on average, a 23.6% proportional increase (11.6% net increase as % of DW) in oil content which is near-additive to the best results obtained in transgenic experiments with the single genes. These findings demonstrate the utility of stacking two specific transgenes controlling the key steps in two very different metabolic streams-mitochondrial carbon flux (*mtPDCK*; “push”) and triacylglycerol assembly via the Kennedy pathway (*DGAT1*; “pull”), to bring about significant increases in oil content in *B. napus*. This approach holds promise for similar use in other oilseed crops.

Keywords: Seed oil content, triacylglycerol synthesis, carbon flux, diacylglycerol acyltransferase, mitochondrial pyruvate dehydrogenase kinase, gene stacking

* Corresponding Author's Email: David.Taylor@nrc-cnrc.gc.ca.

INTRODUCTION

The global demand for vegetable oils for food, bio-diesel and bio-products is increasing at a rapid pace. It is estimated that a 75% increase in canola oil production will be required to meet the growing markets projected for 2020 (Weselake et al., 2009). In addition to increasing the land area devoted to the canola crop, an improvement in oil content is another approach to enhance oil production using current acreage. Plant breeders have applied a quantitative trait loci (QTL) approach to define the link between embryo and maternal genetics, cytoplasmic effects and genotype-by-environment interactions as partial determinants of seed oil content. Beyond these studies the use of genetic engineering, more specifically, the over-expression or suppression of genes encoding enzymes involved in the channeling of carbon into seed oil, has shown utility for significantly improving seed triacylglycerol (TAG) content. Salient to the current study, it has been shown that over-expression of genes encoding acyltransferases of the Kennedy pathway and repressing those controlling respiratory carbon flux can be used to alter the accumulation of oil in the developing seed (Taylor et al, 2011).

The interactions of pathways involved in fatty acid biosynthesis and seed oil (triacylglycerol, TAG) bioassembly, and the physiological and developmental regulatory factors implicated in seed oil production, have been reviewed recently by Bates et al., (2013) and Baud and Lepiniec (2010), respectively, and the reader is referred to these excellent treatises. While significantly more complex, for the purposes of this paper, we will focus on TAG synthesis in its most basic form, as catalyzed by the membrane-bound enzymes of the Kennedy pathway (Kennedy, 1961) that operate in the endoplasmic reticulum (Stymne and Stobart, 1987). The process begins with *sn*-glycerol-3-phosphate (G-3-P) undergoing two acylations catalyzed by the acyltransferases glycerol-3-phosphate acyltransferase (GPAT; EC 2.3.1.15) and *lyso*-phosphatidic acid acyltransferase (LPAAT; EC 2.3.1.51). The final acylation of *sn*-1,2-DAG by diacylglycerol acyltransferase (DGAT; EC 3.2.1.20) to produce TAG, occurs after removal of the phosphate group from the *sn*-3 position of the glycerol backbone by phosphatidic acid phosphatase (PAPase; EC 3.1.3.4).

In the Kennedy pathway, DGAT is the only enzyme that is exclusively committed to TAG biosynthesis using acyl-CoA as its acyl donor. It has been suggested that DGAT may be one of the rate-limiting steps in plant storage lipid accumulation (Ichihara et al., 1988; Perry and Harwood, 1993a & b; Perry et al., 1999; Jako et al., 2001; Weselake, 2005; Lung and Weselake, 2006), and thus a potential target in the genetic modification of plant lipid biosynthesis in oilseeds for economic benefit (Figure 1).

Accordingly, transgenic manipulation of *DGAT1* expression has resulted in enhancement of oil accumulation in the developing embryo during seed development, suggesting that activity of the corresponding indigenous enzyme is indeed, in part, rate-limiting for oil synthesis. Nowhere is this rate-limitation evidence more clear than in our studies of the *Arabidopsis* mutant AS11 having a lesion in *DGAT1* (an exon 2 repeat), which results in a knockout of DGAT activity and greatly reduced oil content. Transformation of AS11 with the wild type *A. thaliana* *DGAT1* was shown to rescue the low oil phenotype of the mutant (Zou et al., 1999; Jako et al., 2001; Xu et al., 2008). Subsequently, in our lab, lines with enhanced oil content were created by seed-specific over-expression of an *Arabidopsis* *DGAT1* in wild type *Arabidopsis* (Jako et al., 2001) and in *B. napus* cv Quantum (Taylor et al, 2009), suggesting that enhancement of DGAT1 activity may provide the metabolic “pull” needed to more

efficiently channel acyl moieties into TAGs and up-regulate the efficiency of the Kennedy pathway in general.

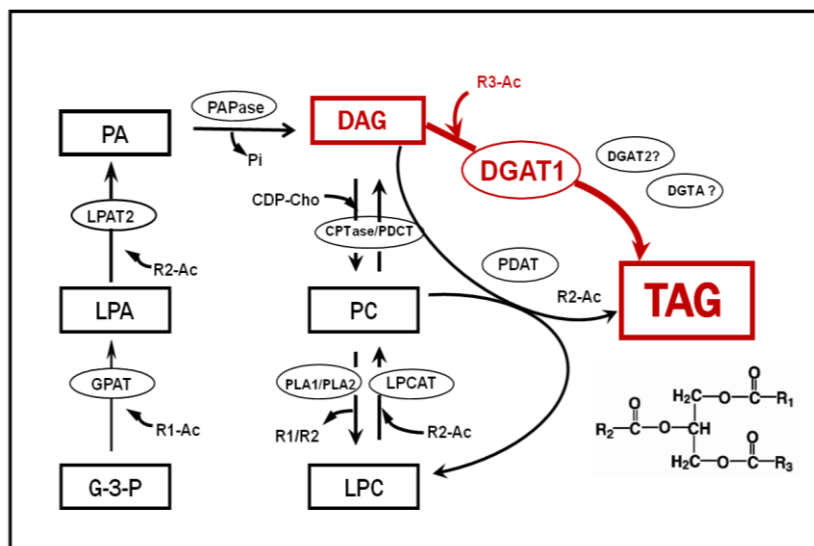


Figure 1. Increasing oil content by over-expression of DGAT1. Diagram showing the intersecting biochemical steps for glycerolipid biosynthesis in oilseed cotyledons and key enzymes (circled): GPAT, glycerol-3-phosphate acyltransferase (EC 2.3.1.15); LPAAT, lyso-phosphatidic acid acyltransferase (EC 2.3.1.51); PAPase, phosphatidic acid phosphatase (EC 3.1.3.4) and DGAT1, diacylglycerol acyltransferase 1 (EC 3.2.1.20) are the enzymes of the traditional Kennedy pathway (Kennedy, 1961). R1-Ac, R2-Ac and R3-Ac are fatty acyl-CoA donors for GPAT, LPAAT and DGAT1, for esterification at the respective *sn*-1, *sn*-2 and *sn*-3 positions on the glycerol backbone. DAG, *sn*-1, 2-diacylglycerol is the acyl acceptor for DGAT1 (as well as PDAT and DGAT2). The relative contributions of DGAT1 (the focus of this study, shown in red), diacylglycerol acyltransferase 2 (DGAT2), diacylglycerol transacylase (DGTA) and phospholipid:diacylglycerol acyltransferase (PDAT, which donates an R2-Ac from phosphatidylcholine to the *sn*-3 position of DAG) in conversion of DAG to TAG, probably vary among crop species. In canola, evidence suggests that DGAT1 is the predominant contributor. LPCAT is acyl-CoA: lyso-phosphatidylcholine acyltransferase (EC 2.3.1.23) which catalyzes the acyl-CoA-dependent exchange of oleoyl moieties (R2-Ac) at the *sn*-2 position of PC with the polyunsaturated *sn*-2 fatty acids 18:2 and 18:3, (produced by the action of FAD2 and FAD3 on phosphatidylcholine respectively; not shown). CPTase (EC 2.7.8.2) is diacylglycerol-choline phosphotransferase and PDCT is phosphatidylcholine diacylglycerol cholinephosphotransferase. These two enzymes mediate interconversions between phosphatidylcholine (PC) and diacylglycerol (DAG), thus enriching PC-modified fatty acids in the DAG pool prior to forming TAG. Often the result is polyunsaturated fatty acid or hydroxyl/epoxy fatty acid enrichment in the DAG and ultimately, the TAG pool. Thus, plants can use two main pathways to produce diacylglycerol (DAG), the immediate precursor molecule to TAG: (1) *de novo* DAG synthesis, and (2) conversion of the membrane lipid phosphatidylcholine (PC) to DAG (Bates and Browse, 2012). PLA1 and PLA2 are phospholipase A1 and phospholipase A2, respectively. Additional abbreviations: G3P, *sn*-glycerol-3-phosphate; LPA, lysophosphatidic acid; LPC, lysophosphatidylcholine; PA, phosphatidic acid; PC, phosphatidylcholine; TAG, triacylglycerol, containing *sn*-1 R1, *sn*-2 R2 and *sn*-3 R3 acyl moieties. For a more detailed treatment the reader is referred to Weselake (2005) and Bates et al., (2013).

In the developing embryo, energy is largely generated during dark respiration by conversion of photosynthate to pyruvate (via glycolysis) which is imported into the mitochondrion, where it is converted to acetyl-CoA which, in turn, feeds into the TCA cycle, producing ATP and reducing equivalents. These are required for various anabolic reactions, including those which contribute to the synthesis and accumulation of seed oil. The key link

between glycolytic and respiratory carbon flux, the conversion of pyruvate to acetyl-CoA, is catalyzed by the mitochondrial pyruvate dehydrogenase (mtPDH) complex. The enzyme mitochondrial pyruvate dehydrogenase kinase (mtPDHK, EC 2.7.1.99), is the negative regulator of mitochondrial pyruvate dehydrogenase (mtPDH, EC 1.2.4.1) (Rubin and Randall, 1977; Luethy et al., 1996; Zou et al., 1999; Buchanan et al., 2000; Marillia et al., 2003; Tovar-Méndez et al., 2003). We found that constitutive partial suppression of *Arabidopsis* mtPDHK via anti-sense (A/S mtPDCK) resulted in transgenic *Arabidopsis* lines having greater respiratory CO₂ release, earlier floral initiation, earlier maturity, enhanced seed oil accumulation and increased harvest index. When the suppression was directed in a seed-specific manner, the effect was primarily at the developing embryo/seed level and improved oil content was the predominant result (Zou et al., 1999a; Marillia, et al., 2003; Weraduwaage, 2013; Dahal et al., 2014) (Figure 2). This suggests that the A/S mtPDHK transgenics have a greater sink strength and that the anabolic properties of dark respiration can be harnessed to more efficiently divert carbon (“push”) to pathways involved in seed oil deposition.

Given these findings, we were interested in comparing results of seed-specific single gene expression of *DGAT1* or A/S *mtPDCK* to those observed upon co-expression of both, to assess whether the latter approach delivers cumulative improvements in *B. napus* oil content.

MATERIALS AND METHODS

Plant Material and Target Genes

All two-gene transformations in this study were performed in the host *B. napus* cultivar DH12075 (kindly provided by Agriculture and Agri-Food Canada, Saskatoon, SK). Single-gene transformation studies were performed in *B. napus* or *Arabidopsis* host cultivars as delineated in Table 1.

After transformation and regeneration, all experimental *B. napus* transgenic and control lines were propagated in the greenhouse at the Kristjanson Biotechnology Complex greenhouses, Saskatoon, under natural light conditions supplemented with high-pressure sodium lamps with a 16 h photoperiod (16 h of light and 8 h of darkness) at 22°C and a relative humidity of 25 to 30%. All *Arabidopsis* plants were grown in a growth chamber at 22°C with photoperiod of 16 h light (120 $\mu\text{E}\cdot\text{m}^{-2}\cdot\text{s}^{-1}$) and 8 h darkness.

The *Arabidopsis thaliana* *DGAT1* (GenBank Accession No. AJ238008) was cloned and characterized as described by Zou et al. (1999b) and Jako et al. (2001). The *Tropaeolum majus* *DGAT1* was cloned from an EST collection (GenBank Accession No. AY084052) as described by Xu et al. (2008) and the synthesis and characterization of the enzyme after site-directed mutagenesis (SDM) of amino acid residue S¹⁹⁷ to A¹⁹⁷, was also detailed therein. The *A. thaliana* mitochondrial *PDCK* (ATCC No. 209562; reported in US Patent No 6,500,670 B1) was cloned and expressed in an anti-sense configuration in *A. thaliana* as described previously Zou et al., 1999a and Marillia et al., 2003). The *B. napus* mtPDCK homolog was cloned and its sequence reported as presented in US Patent No US 7214859 B2 (Marillia et al., 1999).

Table 1. Relative oil content increases achieved in *Brassicaceae* by over-expression of *DGAT1* genes (both WT and SDM), silencing of *mtPDCK* genes or both combined

Gene	Host: Cultivar	Promoter	Proportional Oil Content Increase*
A. thal <i>DGAT1</i>	B. napus: Quantum	napin	7.4 [†] (6)
T. majus <i>DGAT1</i>	B. napus cv DH12075	napin	6.1 ^{††} (4)
T. majus SDM S¹⁹⁷-to-A¹⁹⁷ <i>DGAT1</i> (mutated <i>DGAT1</i>)	A. thaliana	napin	11.0^{††} (5)
T. majus SDM S¹⁹⁷A <i>DGAT1</i> (mutated <i>DGAT1</i>)	B. napus DH12075	napin	11.7^{††} (5)
A/S A. thal <i>mtPDCK</i>	B. napus: Quantum	napin	12.0 ^{††} (10)
A/S B. napus <i>mtPDCK</i>	B. napus: Quantum	napin	16.0^{††} (5)
A/S B. napus <i>mtPDCK</i>	B. napus: Nex710	phaseolin	13.0^{††} (7)
A/S B. napus <i>mtPDCK</i>	B. napus: Nex710	phaseolin	12.4 [†] (4)
T. majus SDM <i>DGAT1</i> + A/S B. napus <i>mtPDCK</i>	B. napus: DH12075	napin	23.6^{††} (4)

* % relative to either Non-Transformed WT (nt WT) or Empty Plasmid (pSE) controls.

[†] Results from confined 2-location field trial.

^{††} Results from greenhouse.

(Numbers in brackets) = number of best lines averaged.

Bolded data are reported here for the first time.

Production of Transgenic Lines

A. Thaliana DGAT1 Lines

The cloned full-length *A. thaliana DGAT1* cDNA (Gen-Bank No. AJ238008; Zou et al. 1999b) was used as a template for PCR amplification with the primers *Xba*I Forward -5'CTA GTC TAG AAT GGC GAT TTT GGA 3' and *Xho*I Reverse 5'GCG CTC GAG TTT CAT GAC ATC GA 3' to provide new restriction sites at each end of the sequence. The PCR amplification program was as follows: 94 °C for 1 min; 30 cycles of 94 °C for 30 s, 55 °C for 30 s, 72 °C for 1 min; and 72 °C for 5 min. The PCR product was then ligated into the pCR-2.1 vector (Invitrogen Canada, Burlington, ON). The plant transformation vector pSE129A (ATCC 97910; Covello et al., 1997) was prepared from pRD400 (Datla et al., 1992) by introducing a *Hind*III/*Xba*I fragment containing the *B. napus* napin promoter (Josefsson et al., 1987) and a *Kpn*I/*Eco*RI fragment containing the *Agrobacterium* NOS terminator (Bevan, 1983). The 1.6 kb *DGAT1* cDNA fragment was excised from the pCR-2.1 vector by an *Xba*I/*Kpn*I digestion and ligated into the corresponding sites of the pSE129A vector, in the sense orientation. The resulting plasmid was designated napin:*DGAT1*. The construct integrity was confirmed by sequencing. Electro-competent *Agrobacterium tumefaciens* cells, strain GV3101 containing helper plasmid pMP90 (Koncz and Schell, 1986) were prepared, frozen in liquid N₂ and stored at -80 °C as described previously (Jako et al. 2001). The *Agrobacterium* cells were transformed by electroporation with 50 ng of DNA (napin:*DGAT1*), plated on a selective medium (LB broth supplemented with a final concentration of 50 mg·mL⁻¹ kanamycin), and incubated for 48 h at 28 °C. Single colonies were grown for 16 h (28 °C, 225 rpm) in 5 mL LB broth supplemented with 50 mg·mL⁻¹ of kanamycin and 25 mg·mL⁻¹ of gentamycin. DNA extraction and purification were performed with a Qiaprep Spin Miniprep kit (Qiagen, Valencia, Calif.).

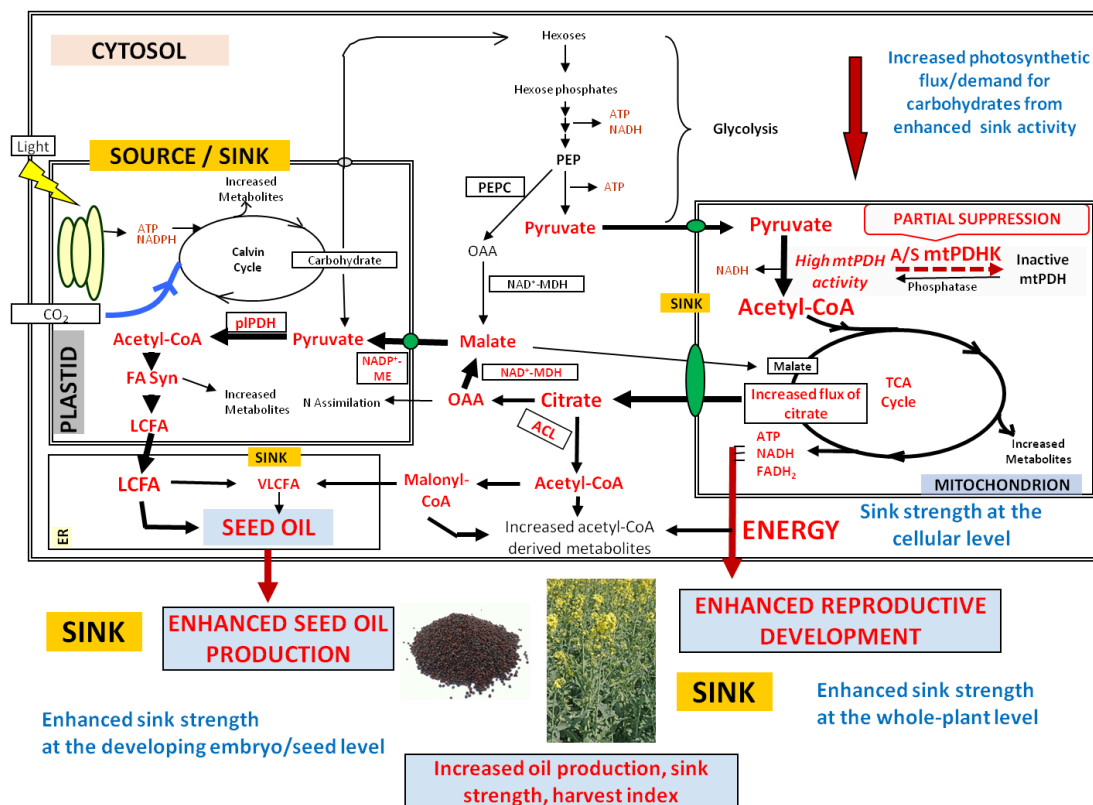


Figure 2. Metabolic implications of altered mtPDHK expression in *Arabidopsis*. Partial suppression of mtPDHK via antisense technology increases the flux of carbon through the TCA cycle both in source and sink organs by easing negative regulation of mtPDH, a key enzyme linking glycolysis and the TCA cycle. Conceptually, this may lead to increased availability of energy and TCA-cycle intermediates that are substrates for anabolic processes, e.g., amino acid and chlorophyll synthesis. Upon increased TCA cycle activity, there is increased export (“push”) of carbon, as citrate, to the cytosol. Via ATP-citrate lyase (ACL), this results in increases in two co-products: (1) cytosolic oxaloacetate which ultimately results in higher plastidial pyruvate and hence, enhanced plastidial fatty acid synthesis for oil assembly and (2) increased cytosolic acetyl-CoA, required to synthesize VLCFAs, isoprenoids, including specific plant growth regulators, aromatic amino acids, acetylated metabolites including sugars, alkaloids, flavonoids etc. necessary for plant growth and development. Partial suppression of mtPDHK provides increased sink strength, thus enhancing productivity and harvest indices. At the same time, mtPDCK suppression can reduce the sink limitation on photosynthesis, allowing higher rates of photosynthesis, providing a mechanism to divert additional photosynthate towards seed oil biosynthesis and other important biosynthetic processes. mtPDH, mitochondrial pyruvate dehydrogenase complex; pIPDH, plastidial pyruvate dehydrogenase; mtPDHK, mitochondrial pyruvate dehydrogenase kinase; MDH, malate dehydrogenase. For all other abbreviations and a more detailed description of metabolic steps impacted by partial suppression of mtPDCK expression, the reader is directed to Weraduwa (2013) from which this figure is modified. Modified with permission from Weraduwa, 2013.

The pSE129A/*DGAT1* plasmid integrity was verified by DNA sequencing following its re-isolation from *A. tumefaciens* and transformation into *E. coli*. The *B. napus* canola cultivars Quantum (Stringham et al. 1995) or DH12075 were then transformed with the *Agro*/pSE129A/*DGAT1* by the cotyledonary petiole method of Moloney et al. (1989). Kanamycin-resistant T₀ transgenic lines were selected, propagated, and segregation analyses performed to identify homozygous T₃ lines, as described previously for other *B. napus* transgenic material (Katavic et al, 2001, Taylor et al., 2001). The oil content and acyl composition of the *B. napus* cv. Quantum transgenic seed lines containing the *A. thaliana DGAT1* were determined as described below and compared to those from non-transformed WT controls.

T. majus DGAT1 and T. majus SDM DGAT1 Lines

The coding region of the *T. majus DGAT1* or *T. majus* S¹⁹⁷-to-A¹⁹⁷ SDM *DGAT1* (Xu et al., 2008) were amplified by PCR with primers: *Xba*I Forward: 5'-TATCTAGAATGGCGGTGGCAGAG -3' and *Kpn*I Reverse: 5'-ATGGTACCTCACTTTTCCTTTAGATTTATC-3' using the same amplification program as described above, and subsequently cloned behind the napin promoter in the respective sites of the pSE129A vector. The final binary vector was verified for integrity by sequencing and electroporated into *A. tumefaciens* as described above and used to transform *Arabidopsis* by the vacuum infiltration method (Bechtold and Pelletier, 1998). Kanamycin-resistant T₁ plants were selected and propagated, and the T₂ and T₃ generation selected and characterized as described by Jako et al., (2001) to identify homozygous T₃ seed lines. These were analyzed for oil content and fatty acid composition as described below. At the same time several independent plasmid-only control transgenics (pSE vector only without *DGAT1* insert) were propagated and analyzed for comparison. The same construct was used to transform *B. napus* cv. DH12075 by the cotyledonary petiole method of Moloney et al., (1989). Kanamycin-resistant T₀ transgenic lines were selected, propagated, and segregation analyses performed to identify homozygous T₃ lines, as described previously for other *B. napus* transgenic material (Katavic et al, 2001, Taylor et al., 2001). The oil content and acyl composition of these SDM *DGAT1* lines were determined as described below and compared to those from non-transformed WT (nt-WT) controls.

mtPDCK Suppression Lines

The *mtPDCK* gene was cloned from *Brassica napus* cv. Quantum by RT-PCR amplification. Total RNA was extracted from young leaves (Wang and Vodkin, 1994) and cDNA produced by reverse transcription (Life Technologies, Inc., 2002, M-MLV Reverse Transcriptase). Using this cDNA and several pairs of degenerate primers (CGGGGTACCTG GGGNNSATGAARCAR and TGCTCTAGATYANGGYAARGG YTCYTS) designed from conserved segments of known *mtPDCK* amino-acid sequences from *Arabidopsis* (CAA07447) and corn (AF038585), a fragment of about 1 kb was amplified by PCR. The fragment was cloned into the TOPO cloning vector (pCR TOPO 2.1, Invitrogen) and fully sequenced in both orientations. DNA sequence analysis revealed that this amplicon shared a high degree of homology with other known *mtPDCK* genes.

The missing termini of the gene were subsequently amplified using a 3' and 5' RACE kit (Life Technologies, Inc., 2002, 3' RACE system and 5' RACE system). The full-length gene was produced by PCR using Vent DNA polymerase (New England Biolabs) and gene specific primers (CTTTCTTCAGCTGGTTCACAC, GACTCCACATACCAATCTC TTAC

CTTCAA, CATAAGGCAGCGTCTCGAGTTCG, AGATGTGGACTTG GGAA CCGCTGAT, GTTATGGTCTGCCTATTAGTCGCTTGTA) designed from the DNA sequence information provided by the RACE-generated fragments. These primers encompassed the start and stop codons. At this stage, two *PmeI* restriction sites were also added by PCR for subsequent anti-sense insertion of the *mtPDCK* gene into expression vectors pSE129A bearing the napin promoter or pBBV-PHAS with the phaseolin promoter (courtesy of Dow AgroSciences) as described in patent CA2474906 C. The antisense (A/S) orientation of the inserted gene was verified by restriction digestions and DNA sequencing.

The final binary vector was verified and electroporated into *A. tumefaciens* as described above and used to transform *Arabidopsis* by the vacuum infiltration method (Bechtold and Pelletier, 1998). Kanamycin-resistant (pSE129A construct) or phosphinothricin-resistant (pBBV-PHAS construct) T₁ plants were selected and propagated, and the T₂ and T₃ generation selected and characterized as described by Jako et al., (2001) to identify homozygous T₃ seed lines. These were analyzed for oil content and fatty acid composition as described below. At the same time several independent plasmid-only control transgenics (pSE129A vector only without antisense *mtPDCK* insert) were propagated and analyzed for comparison. The same constructs were used to transform either *B. napus* cv. DH12075 or cv. Nex710 (Dow AgroSciences) by the cotyledonary petiole method of Moloney et al. (1989). Kanamycin-resistant (pSE129A construct) or phosphinothricin-resistant (pBBV-PHAS construct) T₀ transgenic lines were selected, propagated, and segregation analyses performed to identify homozygous T₃ lines, as described previously for other *B. napus* transgenic material (Katavic et al, 2001, Taylor et al., 2001). The oil content and acyl composition of these A/S *mtPDCK* lines were determined as described below and compared to those from nt-WT controls.

Two-Gene Stack Lines

A two gene construct was created containing the modified *T. majus* SDM *DGAT1* and the *B. napus* *mtPDCK* (in an antisense orientation) genes, both driven by the seed-specific promoter napin, and inserted in the plant transformation vector pSE129A, to produce the vector pGHI-006 (Figure 3). To do this, our original single gene expression cassettes were first inserted into the cloning vector pBluescript SK+ by a blunt ligation to create pNap: *mtPDHK* and pNap: SDM *TmDGAT1*. pNap: *mtPDHK* was then digested by *NotI* and *PspOMI* and the promoter-gene-terminator fragment was cloned into the *NotI*-digested pNap:SDM *TmDGAT1* linear vector to create pGHI-005. A *Sall* fragment containing both expression cassettes was then inserted into the transformation vector pSE129A digested with *SmaI*, by blunt ligation to create pGHI-006. This vector was subsequently used to electroporate *Agrobacterium* as described above and used to transform the *B. napus* canola cv. DH12075 using the cotyledonary petiole method of Moloney et al. (1989). Kanamycin-resistant transgenic lines were screened, selected and characterized as described previously for other *B. napus* transgenic material (Katavic et al, 2001, Taylor et al., 2001). The oil content and acyl composition of these two-gene transformant lines were determined as described below and compared to those from empty pSE129A plasmid controls.

Field Trials

Where indicated in Table 1 and in the text, confined field trials of the *A. thaliana* *DGAT1* in *B. napus* cv Quantum lines were conducted in 2006 and 2007 at Watrous, SK as described by Taylor et al., (2009); field trials of the *B. napus* *DGAT1* in *B. napus* cv DH12075 were conducted at two locations in Alberta (Ellerslie and Vegreville), as described by Weselake et al., et al., (2008). Phaseolin *mtPDCK*-suppressed Nex710 lines were tested in confined field trials at Morden, MB and Saskatoon, SK in plots designed similar to those described by Taylor et al., (2009).

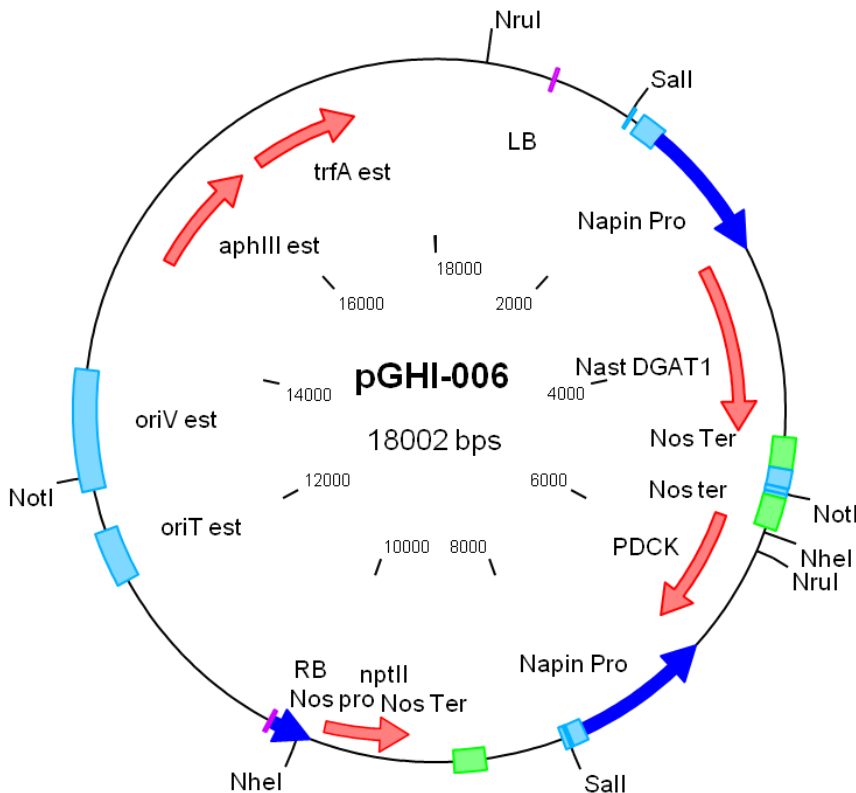


Figure 3. Map of 18 kb 2-gene construct pGHI-006, containing the SDM S¹⁹⁷-to-A¹⁹⁷ *TmDGAT1* (Nast DGAT1) and the *mtPDCK* (PDCK) in sense and antisense orientations, respectively. Both genes are driven by the napin promoter.

Analysis of Seed Oil Content and Composition in Transgenic Lines

Oil content of mature seed from *B. napus* was determined by low-resolution nuclear magnetic resonance spectroscopy (LR-NMR) using a Bruker Minispec mq20 instrument, calibrated with mature *B. napus* seed of known oil content.

As described below, quantifying the fatty acid content of a total lipid extract was the method used to determine the lipid content in pre-weighed *Arabidopsis* seed samples as this was more quantitative and accurate than the NMR method for these extremely small seeds.

The acyl composition of mature seed from *Arabidopsis* and *B. napus* transgenic and nt-WT or plasmid-only pSE129A control plants were determined as follows: Mature seed samples were weighed after being dried overnight at 80 °C, then ground with a Polytron model PT 20, in 4 mL isopropanol:CH₂Cl₂ (2:1 v/v) containing 15:0 triacylglycerol as an internal standard. The debris was washed twice with 1 mL of grinding solvent and the pooled eluants transferred to a test tube. To this was added 0.9% NaCl (1ml) and vortexed. Two ml of CH₂Cl₂ was added, the mixture re-vortexed and centrifuged at 2500 r.p.m. for 3 min. The CH₂Cl₂ layer was removed, the extraction repeated and the CH₂Cl₂ layers combined. CH₂Cl₂: Benzene: Methanol (1:1:1 v/v/v) (1 mL) was added and the sample evaporated to dryness under a stream of N₂. The dried sample was re-suspended in CHCl₃ (1 mL) to give the total lipid extract (TLE). The acyl composition and fatty acid content of the TLE were determined by trans-esterification of the neutral lipid fraction with 3N methanolic HCl, and GC of the resulting fatty acid methyl esters with 17:0 methyl ester as an external standard as previously described (Taylor *et al.*, 2001).

In order to discuss trends in oil content improvements due to different genes/constructs, it was necessary to normalize the oil content relative to the individual control plants within each experiment. This proportional increment in oil content was calculated as:

$$[(\text{Transgenic oil as \% of DW} - \text{Control oil as \% DW}) \div \text{Control oil as \% DW}] \times 100$$
, where the controls are either seed from the corresponding non-transformed WT plants or from empty plasmid transgenics within that experimental set. Data are presented $\pm$ S.D. (n=5-8).

Net oil increases as a % of dry weight were calculated as:

$$[(\text{Transgenic oil as \% of DW} - \text{Control oil as \% DW})]$$
, where the controls are either seed from the corresponding non-transformed WT plants or from empty plasmid transgenics within that experimental set. Data are presented $\pm$ S.D. (n=5-8).

Protein Analyses

Estimates of the crude protein content in seeds from selected 2-gene stack lines and from controls were obtained by the Dumas combustion method (ISO 16634-1:2008 Food products - Determination of the total nitrogen content by combustion according to the Dumas principle and calculation of the crude protein content, Part 1: Oilseeds and animal feeding stuffs. http://www.iso.org/iso/catalogue_detail.htm?csnumber=46328). Total seed nitrogen values (on the basis of % of DW) were multiplied by the conversion factor of 6.25 to arrive at the crude protein content proportions. It is important to note that, while it is one of the industry standard tests for oilseed analysis, the combustion method does not distinguish between nitrogen which is contributed by proteins vs nucleic acids, nitrates or other nitrogenous compounds.

RESULTS

Increases in *Brassica napus* seed oil content in transgenic lines expressing single *DGAT1* (WT or SDM) or silenced *mtPDCK* genes.

DGAT1 Lines

The first *DGAT1* to be cloned from plants, specifically, *A. thaliana*, was simultaneously reported by Zou et al., Hobbs et al., and Routaboul et al., in 1999. All three labs had taken advantage of an *Arabidopsis* EMS mutant, designated AS11 (*dgat1*), which we originally isolated (Katavic et al, 1995). AS11 showed reduced embryo DGAT activity and decreased seed oil with a drastically altered fatty acid composition. The lesion was shown to be an insertional event resulting in a tandem exon 2 repeat. When the WT *DGAT1* gene was over-expressed in AS11, the gene complemented the mutant phenotype, restoring oil content to WT levels and even higher (Jako et al, 2001).

Subsequently, enhancement in oil content has been achieved by over-expression of *DGAT1*s from both *A. thaliana* and *B. napus*, in two different *B. napus* cultivars, Quantum and DH12075, respectively. Seed-specific over-expression of the *A. thaliana* *DGAT1* in *B. napus* cv. Quantum produced homozygous T₃ single insert transgenic lines with absolute and proportional seed oil content increases that were on average, 4.2% and 10.0%, respectively, compared to non-transformed controls under greenhouse conditions (Weselake et al., 2007). In confined field trials over 2 years, the highest proportional oil content increases averaged 7.4% (Table 1). Transformation of *B. napus* cv. DH12075 with an expression cassette containing a modified *B. napus* *DGAT1* fitted with an ER-retention signal showed absolute oil content increases of 2.5-3.5%, which translated to 5.7–7.0% relative increases in a 2-location field trial. These field results were very similar to those observed with the same construct in the greenhouse (3-5% absolute oil content increments) (Taylor et al., 2009). In none of these transgenic experiments were there any significant differences in fatty acid profiles (data not shown). Interestingly, from an agronomic viewpoint, in a field trial in 2003 during which there was a severe drought, the over-expressed *A. thaliana* *DGAT1* was able to rescue a low oil phenotype of *B. napus* cv. Quantum with up to 4.1% absolute and 13.6% relative increases compared to the nt-WT (Weselake et al., 2008).

A *DGAT1* cloned from *Tropaeolum majus* (Xu et al., 2008) was functionally expressed in the yeast *H1246 α MAT* quadruple mutant (*dga1*, *lro1*, *are1*, *are2*) and restored the capability of the mutant host to produce TAGs. The recombinant *TmDGAT1* protein present in lysates of the transformant was capable of utilizing a range of ¹⁴C- labeled acyl-CoA donors and could synthesize ¹⁴C-trierycin from dierycin and erucyl-CoA. Collectively, these findings confirmed that the *TmDGAT1* gene encodes a protein which functions as an acyl-CoA-dependent DGAT1. In plant transformation studies, seed specific expression of the *TmDGAT1* was able to complement the low TAG and fatty acid compositional mutant phenotype of the *Arabidopsis* AS11 (*dgat1*) mutant. When the WT *TmDGAT1* was expressed in either *A. thaliana* (Figure 4), or *B. napus* cv. DH12075, (Figure 5) the proportional oil content averaged about 6% relative to the respective plasmid-only pSE control transformants.

Site-directed mutagenesis (SDM) studies were conducted on 5 putative functional regions/motifs of the *TmDGAT1* enzyme. Of note, mutagenesis of a serine¹⁹⁷ residue to an alanine residue in a putative SnRK1 target site resulted in an average 75% increase in DGAT activity when expressed in yeast (Xu et al., 2008), strong evidence that this putative Ser/Thr protein kinase (negative regulatory) site could be exploited to enhance DGAT1 activity (by preventing phosphorylation at serine¹⁹⁷). Accordingly, over-expression of the *T. majus* SDM S¹⁹⁷-to-A¹⁹⁷ *DGAT1* gene in *A. thaliana* and in *B. napus* cv. DH12075 resulted in average proportional oil content increases of 11.0% (Table 1, Figure 4) and 11.7% (Table 1, Figure 5),

respectively, relative to the plasmid-only pSE controls. Absolute oil content increments were 4.2% and 6.0%, respectively in the two plant hosts. There was no significant difference in the fatty acid profile in either set of transgenics (data not shown). Until this study, there has been no report of the effects of expression of the S¹⁹⁷-to-A¹⁹⁷ SDM *T. majus* *DGAT1*, in *Arabidopsis* or *B. napus*.

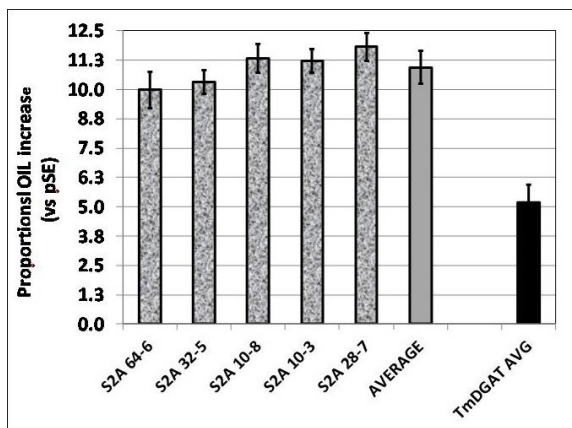


Figure 4. Effect of expression of the SDM S¹⁹⁷-to-A¹⁹⁷ *TmDGAT1* on proportional oil increases in *Arabidopsis thaliana*. Proportional oil increases in the SDM *T. majus* *DGAT1* (S2A) transgenic lines are expressed relative to controls obtained by transformation with the empty pSE plasmid alone. Results with transformation with the WT *T. majus* *DGAT1* (TmDGAT) are shown for comparison. All values have been calculated using the formula reported in Materials and Method and are from three replicates. Averages are reported $\pm$ S.D. (n = 5-8).

Repressed Mitochondrial *PDCK* Lines

The first mitochondrial *PDCK* (mt*PDCK*) gene was cloned from *Arabidopsis* by Zou et al., (1999). Transgenic *Arabidopsis* lines having either constitutive (35S) or seed-specific (napin) antisense (A/S) suppression of *mtPDHK* show elevated mtPDH activity (up to 130% 35S; 230% napin), greater respiratory CO₂ release, earlier floral initiation (by as much as 10% = 7-10 d), enhanced seed oil synthesis and accumulation (10% 35S; 30% napin) and an increased harvest index (seed weight enhanced by up 16% 35S, 23% napin) (Zou et al., 1999; Marillia, et al., 2003; Marillia and Taylor, unpublished).

Napin-driven seed-specific expression of A/S *A. thal* mt*PDCK* in *B. napus* cv. Quantum produced transgenic lines which showed, on average, a 12% proportional increase in oil content relative to the plasmid-only controls. Similar experiments conducted with the *B. napus* mt*PDCK* homolog in cv. Quantum resulted in a 3.6% absolute oil increment or a 16% proportional oil content increase (Table 1, Figure 6). Using the phaseolin promoter in *B. napus* cv. Nex710, we observed a net 5% absolute oil increase which translated to a 13% proportional increment in oil (Figure 7). In stark contrast to the case in *A. thaliana* cited above (Marillia et al., 2003), there was no significant change in average seed weights in the greenhouse. In the A/S *B. napus* mt*PDCK* cv. Quantum lines, the average seed weight was 4.38 ± 0.59 mg/seed compared to the nt-WT control at 4.00 ± 0.68 . Similarly, A/S *B. napus* mt*PDCK* transgenics in the Nex710 background had average seed weights of 4.26 ± 0.33 mg/seed vs 3.79 ± 0.31 in the

nt-WT. There were no differences in fatty acid composition between transgenics and controls (data not shown). These results were borne out in 2-site field trials where oil content was proportionally increased by about 12.4% (Table 1). Generally, there were no consistent changes in average seed weight in the field except for line 20-5-3, which showed a remarkable 18% relative increase (4.7 mg) vs the nt-WT Nex 710 (4.1 mg) at that one site. Clearly, the performance of this line merits further testing in future field trials which are beyond the scope of the current study.

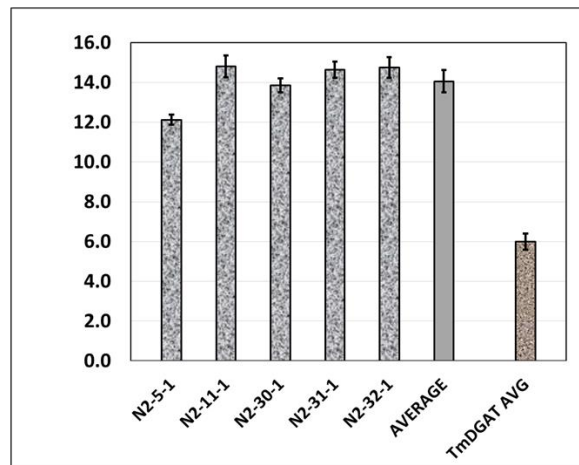


Figure 5. Effect of expression of the SDM S¹⁹⁷-to-A¹⁹⁷ *TmDGAT1* on proportional oil increases in *B. napus* cv. DH12075. Proportional oil increases in the SDM *T. majus DGAT1* (N2) transgenic lines are expressed relative to controls obtained by transformation with the empty pSE plasmid alone. Results with transformation with the WT *T. majus DGAT1* (TmDGAT) are shown for comparison. All values have been calculated using the formula reported in Materials and Method and are from three replicates. Averages are reported $\pm$ S.D. (n = 5-8).

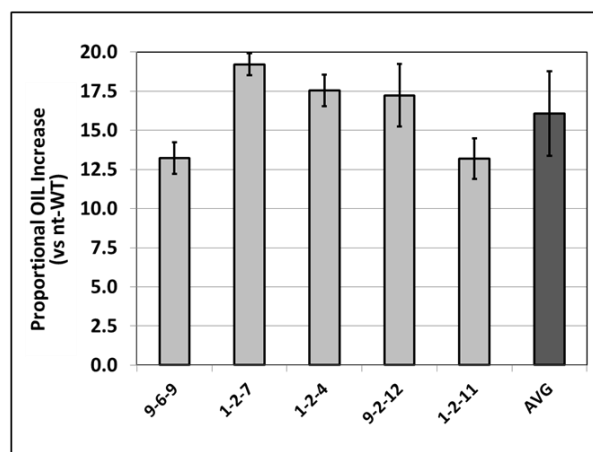


Figure 6. Effect of partial seed-specific suppression of *mtPDCK* results on proportional oil increases in *B. napus* cv. Quantum. Proportional oil increases in the anti-sense *mtPDCK* transgenic lines are expressed relative to non-transformed WT controls. All values have been calculated using the formula reported in Materials and Method and are from three replicates. The average increase is reported $\pm$ S.D. (n = 5-8).

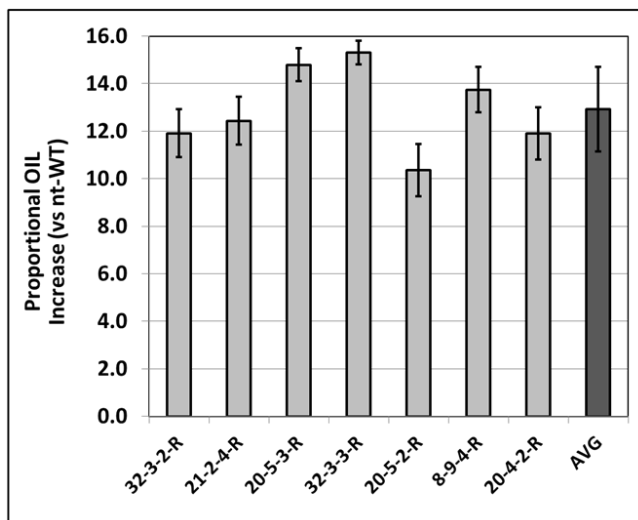


Figure 7. Effect of partial seed-specific suppression of *mtPDCK* on proportional oil increases in *B. napus* cv. Nex710. Proportional oil increases in the anti-sense *mtPDCK* transgenic lines are expressed relative to non-transformed WT controls as calculated using the formula reported in Materials and Method and are from three replicates. The average increase is reported $\pm$ S.D. ($n = 7-8$).

INCREASES IN *BRASSICA NAPUS* SEED OIL CONTENT IN TRANSGENIC LINES EXPRESSING THE TWO-GENE CONSTRUCT CONTAINING THE SDM *T. MAJUS DGAT1* PLUS A/S *B. NAPUS* *MTPDCK*

The two-gene construct driven by the napin promoter, contained the *T. majus* SDM S¹⁹⁷-to-A¹⁹⁷ *DGAT1* in a sense orientation and the *B. napus* *mtPDCK* in an anti-sense orientation. The rationale was that there would be increased carbon flux into fatty acid precursors (A/S *mtPDCK*), as well as an effect on the efficiency of oil deposition (SDM *DGAT1*), i.e., a “push-pull” effect. As shown in Figure 8, the four best two-gene transgenic lines showed consistent proportional oil increases of about 24% based on net oil increments of 11-12% compared to pSE plasmid-only transformants (Figure 9). This is very close to an additive response of those found by expression of each individual gene in a number of backgrounds (Table 1; Figure 10).

With respect to the other major seed storage component, protein, there were significant differences in protein content compared to the controls (Figure 11); for example, line 6-19M-8-1 showed more than a 5% increase in seed protein content compared to controls.

There were no significant differences in average 10-seed weights between the two-gene transgenics and controls. Comparing the seed weights of the highest and the lowest oil increase lines with the controls, line 6-19M-8-1 and 6-18P-10-4 had average 10-seed weights of 51.9 ± 5.6 mg and 48.9 ± 2.0 mg, respectively, compared to the control which had a 100-seed weight of 51.2 ± 2.9 mg (Figure 11).

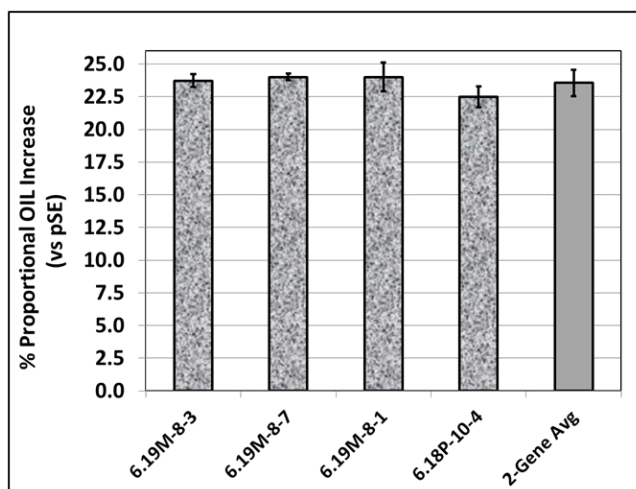


Figure 8. Effect of partial seed-specific suppression of *mtPDCK* combined with expression of the SDM S^{197} -to- A^{197} *TmDGAT1* on proportional oil increases in *B. napus* cv. DH12075. Proportional oil increases in the 2-gene transgenic lines are expressed relative to controls obtained by transformation with the empty pSE plasmid alone. Proportional oil increases have been calculated using the formula reported in Materials and Method and are from three replicates. The average increase is reported $\pm$ S.D. ($n = 4-8$).

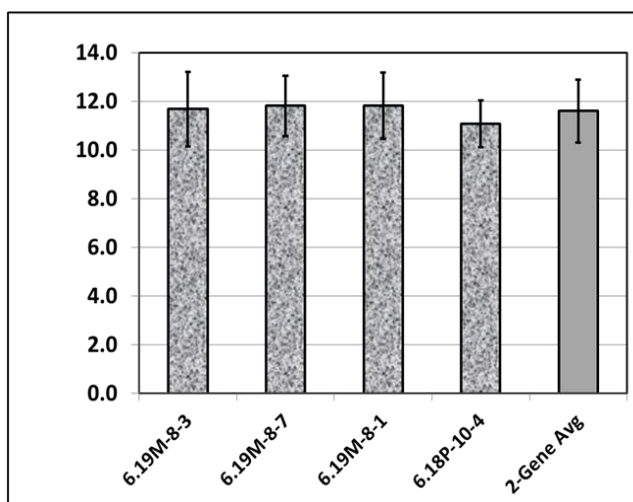


Figure 9. Effect of partial seed-specific suppression of *mtPDCK* combined with expression of the SDM S^{197} -to- A^{197} *TmDGAT1* on net oil increases (as % of dry weight) in *B. napus* cv. DH12075. Net oil increases in the 2-gene transgenic lines are expressed relative to controls obtained by transformation with the empty pSE plasmid alone. Net oil increases (% DW) have been calculated using the formula reported in Materials and Method and are from three replicates. The average increase is reported $\pm$ S.D. ($n = 4-8$).

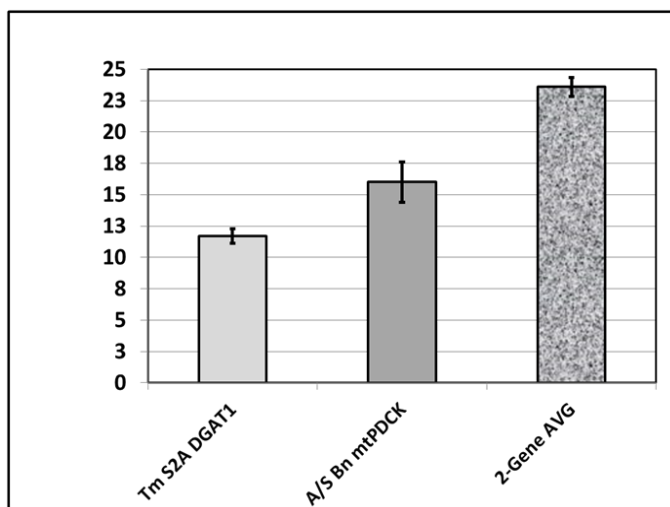


Figure 10. Comparison of effect on proportional oil increases in *B. napus* cv. DH12075 obtained by expression of the 2-gene construct (partial seed-specific suppression of *mtPDCK* combined with expression of the SDM S¹⁹⁷-to-A¹⁹⁷ *TmDGAT1*) with that obtained by transformation with the individual genes alone (*Tm S2A DGAT1* or *A/S Bn mtPDCK*). Proportional oil increases are expressed relative to controls obtained by transformation with the empty pSE plasmid alone and have been calculated using the formula reported in Materials and Methods and are from three replicates. The average increase is reported $\pm$ S.D. (n = 4-8).

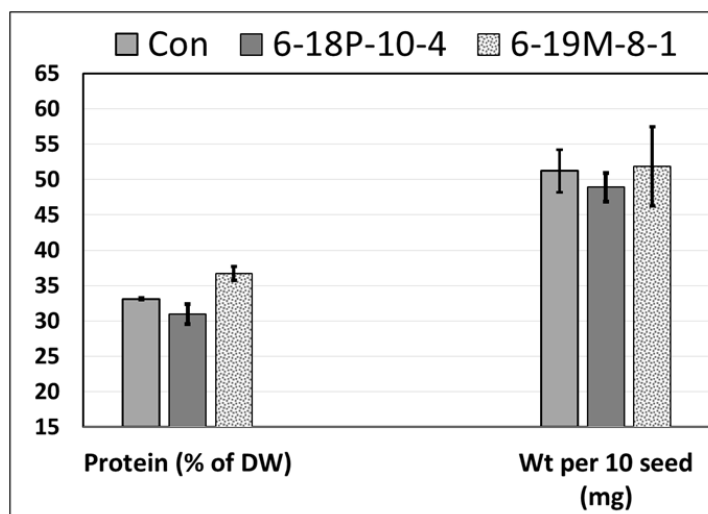


Figure 11. Effect of partial seed-specific suppression of *mtPDCK* combined with expression of the SDM S¹⁹⁷-to-A¹⁹⁷ *TmDGAT1* on average seed weight and protein content in *B. napus* cv. DH12075. Average 10-seed seed weights (mg) in the 2-gene transgenic lines are expressed relative to controls obtained by transformation with the empty pSE plasmid alone. Weights were determined on 100-125 seed samples analyzed in triplicate and averages are reported $\pm$ S.D. (n = 4-8). Crude protein content was determined using the Dumas combustion method as described in Materials and Methods (n = 4-6).

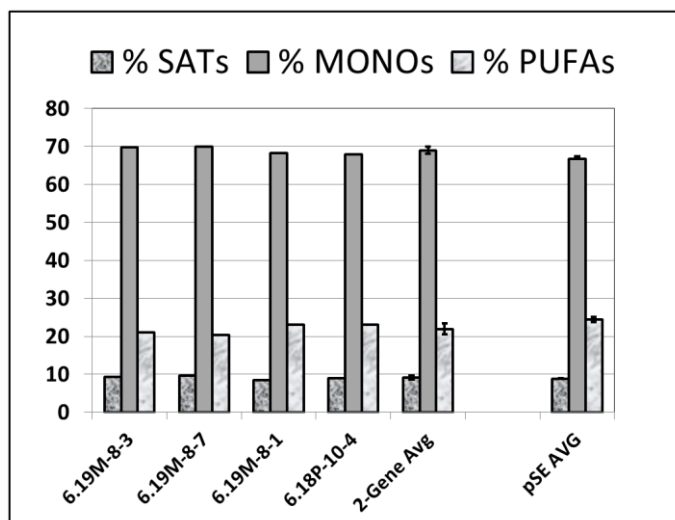


Figure 12. Effect of partial seed-specific suppression of *mtPDCK* combined with expression of the SDM S¹⁹⁷-to-A¹⁹⁷ *TmDGAT1* on fatty acid composition of *B. napus* cv. DH12075 seed oil. Fatty acid composition is reported as total saturated, mono-unsaturated or poly-unsaturated fatty acid proportions in seed oil from the 2-gene transgenic lines relative to controls obtained by transformation with the empty pSE plasmid alone. Oil analyses were performed in triplicate and averages are reported $\pm$ S.D. (n = 4-8).

With respect to the oil composition, the transgenics showed a slight increase in monounsaturated fatty acids (2.3%, 2% of this is oleic acid), a concomitant decrease (2.5%) in polyunsaturated fatty acids, and essentially identical proportions of saturated fatty acids compared to the pSE controls (Figure 12). Any increment in oleic acid is a positive change in canola germplasm improvement work. However, further study is required to see whether this small increment is observed under field conditions.

DISCUSSION

While beyond the scope of the current study, it is important to note that in breeding *B. napus*, the use of heterosis in high vs. low oil lines, and QTLs to map the relationship of specific genetic loci to oil content, continue as an effective, but not necessarily rapid, way to increase crop yield and oil content. Studies reported over the last decade on the use of QTLs in breeding for oil content improvements include those by Delourme et al., (2006), Qiu et al., (2006), Cao et al., (2010), Sun et al., (2012), Wang et al., (2013) and Jiang et al., (2014), and the reader is advised to consult these and the references therein for a comprehensive treatment of these approaches.

Single Gene Transfers for Improvements in Oil Content

Though certainly not a comprehensive list, the directed over-expression of single lipid metabolism-related genes to improve oil content in *Arabidopsis*, *B. napus* and other oilseeds

include: expression of a yeast *SLC1-1* (lyso-phosphatidic acid acyltransferase, *LPAAT*; Zou et al., 1997; Taylor et al., 2001), homeologous *B. napus* *LPATs* (Maisonneuve et al., 2010); a yeast glycerol-3-phosphate dehydrogenase (Vigeolas et al., 2007), diacylglycerol acyltransferase 1 (*DGAT1*, Bouvier-Nave et al., 2000; Jako et al., 2001; Weselake et al., 2007; Taylor et al., 2009), diacylglycerol acyltransferase-2 (*DGAT2*, Lardizabal et al., 2008; Zhang et al., 2013) and phospholipid: diacylglycerol acyltransferase1 (*PDAT1*) for leaf TAG (Fan et al., 2013). By re-targeting the Arabidopsis homomeric cytosolic form of Acetyl-CoA Carboxylase (*ACCase*) to plastids, there was a 5% increase in oil reported (Roesler et al., 1997).

The complex cascade of events/interactions among transcription factors is still unfolding as they relate to directing fatty acid synthesis and oil accumulation. Nonetheless, there have been numerous studies of targeted over-expression of transcription factors to enhance oil, including: *Zea mays Leafy cotyledon 1* (*LEC1*, Shen et al., 2006); *Castor Leafy cotyledon 2* (*LEC2*, Kim et al., 2013); soybean *GmDof4* and *GmDof11* (DNA binding with one finger) (Wang et al., 2007), *BnGRF2* (*Brassica napus growth-regulating factor 2*-like gene; Liu et al., 2012), *WRINKLED1* (Liu et al., 2010), *FUS3* (Wang et al., 2007) and *SHOOTMERISTEMLESS* (Elhiti et al., 2012).

Mutant studies have implicated pyruvate kinase (Andre et al., 2007), *FatA* thioesterases (Moreno-Perez et al., 2012), cytosolic glucose-6-phosphate dehydrogenases (Wakao et al., 2008) and *GLABRA2* (Shen et al., 2006; Shi et al., 2012) as having positive influences on seed oil accumulation in *Arabidopsis*. These are but a few examples; an examination of the extensive literature using *A. thaliana* mutants to identify (often via complementation) key genes involved in aspects of fatty acid metabolism, oil accumulation and/or seed development, is beyond the scope of this study.

In a recent report by Chen et al., (2011), a study was made of Kennedy pathway enzyme activities and their correlation with oil content in *B. napus*. Using lines with relatively high, medium or low oil, sampled over the span of 18-46 days after planting, and monitoring Kennedy pathway enzyme activities, they showed that peak values and averages of the lowest KP enzyme activity (LEA) in the lines were significantly correlated with seed oil content, indicating that KP enzyme efficiency is tightly associated with oil deposition. The LEA (which can be regarded as indicator to the TAG bioassembly efficiency) was, in most cases, *DGAT1*. This further supports over-expression data (discussed below) which confirm that *DGAT1* can be viewed as a bottleneck/rate limiter for oil deposition.

Interestingly, the first report of transgenic expression of *DGAT1* was one wherein the *A. thaliana DGAT1* was ectopically expressed not in seeds, but in leaves of tobacco where endogenous oil is typically a few percent by weight (Bouvier-Nave et al., 2000). The authors reported increases in TAG in leaf lipid droplets of 3.6-fold on average, in T₁ leaves of the best 13 transformants, with the best transformant showing a remarkable 7-fold increment in leaf TAG compared to nt-WT plants. This increase was strongly correlated with *AtDGAT1* transcript expression.

Following this, over-expression of WT *A. thal DGAT1* in *Arabidopsis* resulted in an average of 5.1% absolute and 17.5% proportional increments in seed oil content in the best 8 single-insert events, relative to plasmid-only pSE controls; multiple insert lines showed a notable 21% proportional increase (Jako et al., 2001). These results correlated well with *DGAT1* transcript levels and suggested a gene-dosage effect favored by multiple inserts. *B. napus* transgenics expressing a modified *B. napus DGAT1* (engineered with an ER-retention peptide) showed up to a 7% relative seed oil increase in a 2-location confined field trial, similar to the improvement

observed in the greenhouse (Weselake et al. 2008). This demonstrated that the capacity to improve oil content by this targeted approach was consistent, stable and able to translate exceptionally well from greenhouse to field. In contrast, silencing of *DGAT1* in tobacco caused a reduction in seed oil content (Zhang et al., 2005) which equally confirms the direct link between the level of *DGAT1* expression, *DGAT1* enzyme activity and oil deposition.

In the current study, over-expression of WT *A. thaliana* or *T. majus* *DGAT1*s in *B. napus* showed on average, 6-7% relative increments in oil content (Table 1). These results bear out past demonstrations of improvements in seed oil content using WT *DGAT1* genes.

There have also been several studies wherein *DGAT2*-like proteins from fungi or marine sources (having no significant homology to *DGAT1*s from plants) have been expressed in various plant hosts: The expression of an *Umbelopsis ramanniana* *DGAT2A* in soybean resulted in a 7.5% proportional increase in oil content (Lardizabal et al., 2008). More recently, a *TaDGAT2* cloned from the oleaginous marine protist *Thraustochytrium aureum*, was expressed in either WT or the *fad3/fae1* mutant *A. thaliana* seeds. This resulted in little change in seed oil content, but a striking increase in oleic acid proportions, up to 50% of total fatty acids in the *fad3/fae1* background (Zhang et al., 2013). It is interesting to note that neither the *Umbelopsis* nor the protist *DGAT2* transgenics exhibited striking differences in oil content in dicots. The recent transgenic expression of the *Umbelopsis ramanniana* *DGAT2* in maize using an embryo-enhanced promoter resulted in small but significant increases in kernel oil of up to 0.7% by weight (a proportional increase of 19%), while overexpression of a *Neurospora crassa* *DGAT2* gene resulted in slightly higher oil increases up to 0.9% by weight (a proportional increase of 26%) (Oakes et al., 2011). It is critical to remember that in corn, the embryo is only a small part of the kernel, and therefore the overall gain in oil is small. Thus, the capacity to strongly improve seed oil content by over-expression of individual non-plant *DGAT2* genes is questionable.

The SnRK1 (Sucrose non-fermenting (SNF)-related protein kinase-1) proteins are a class of Ser/Thr protein kinases that have been implicated in the global regulation of carbon metabolism in plants (Halford and Hardie, 1998). It has been hypothesized that the SnRK1 complex can be regulated by different metabolites according to the organs or tissues involved, (source or sink) and that SnRK1s are at the very center of energy metabolism control (Polge and Thomas, 2006). In the developing seed (sink), we conjectured that such metabolite regulation may be controlled by the size of the DAG pool (DAG is implicated as an initiator in a number of signal transduction cascades) or the relative ratio of key Kennedy pathway intermediates, such as e.g., the PC/DAG or TAG/DAG ratios. If, for example, the DAG available for synthesis of membrane and other polar lipids (e.g., phosphatidyl-cholines, -ethanolamines, -serines and -inositols) becomes limiting during the exponential phase of seed development, then a regulatory mechanism for diverting DAG to membrane lipids rather than TAG synthesis (i.e., down-regulation of *DGAT* activity), might be necessary (Xu et al, 2008). *DGAT1* could join a growing list of key enzymes in different plant metabolic pathways, including nitrate reductase, 3-hydroxymethyl -3-methylglutaryl-CoA reductase, sucrose phosphate synthase and trehalose phosphate synthase 5, which are coordinately regulated by SnRK1-catalyzed post-translational enzyme phosphorylation (Xu et al., 2008).

The current study is the first to report the effects of expressing a *DGAT1* possessing a mutation (S¹⁹⁷-to-A¹⁹⁷) in a putative SnRK1 site, with respect to seed oil deposition. The expression of the mutated *TmDGAT1* in *B. napus* resulted in a proportional increase in oil

content roughly twice that observed in experiments over-expressing the WT *T. majus* *DGAT1* (Figure 5). This mirrored the results of expression of the WT *Tm* *DGAT1* vs the SDM S¹⁹⁷-to-A¹⁹⁷ *DGAT1* in yeast, where the microsomal fractions from transformant lysates showed that the SDM *DGAT1* activity averaged 75% higher than that observed from expression of the WT *T. majus* gene (Xu et al., 2008). Collectively, this provided strong justification for selecting the SnRK1 site-mutated, rather than the WT, form of the *T. majus* *DGAT1* for our gene stacking experiments in *B. napus*, as described below.

At about the same time that our work on the significance of the S¹⁹⁷ residue in *DGAT1* was published, a study in corn showed that a high-oil QTL (*qHO6*) which positively correlated with maize seed oil and oleic-acid contents, encodes an acyl-CoA:diacylglycerol acyltransferase (*DGAT1-2*), which catalyzes the final step of oil synthesis. The study demonstrated that a phenylalanine insertion in *DGAT1-2* at amino acid residue 469 (F⁴⁶⁹) is responsible for the enhancement in these two traits. The *DGAT1-2* allele with F⁴⁶⁹ was shown to be ancestral, whereas the allele without F⁴⁶⁹ is a more recent mutant selected by domestication or breeding. Ectopic expression of the high-oil *DGAT1-2* allele increased oil content from 3-4%, typically found in domestic maize cultivars, to 6-7%, an increase of up to 41% (Zheng et al., 2008).

While counter-intuitive, rather than increasing carbon loss, the enhanced respiration (increased mitochondrial PDH and TCA cycle activity) afforded by transgenic silencing the negative regulator of the PDH complex (*mtPDCK*) in *Arabidopsis* led to enhanced oil accumulation and a higher yield (seed weight) as reported in section 3.1.2, (Zou et al., 1999a; Marillia et al., 2003). Subsequent experimentation wherein, under control of the napin promoter, we silenced the *B. napus* *mtPDCK* in *B. napus* cvs. Quantum and Nex710, showed strong proportional increases in oil content (13-15%) compared to nt-WT plants (Figure 6). This confirmed that a useful approach to effect improvements in oil content is the suppression of seed *mtPDCK* to limit the negative regulation of the PDH complex activity, the entryway of photosynthetic carbon into the respiratory TCA cycle for the production of ATP and reducing equivalents. We have recently reported preliminary findings that transgenic *Arabidopsis* having increased respiratory flux through partial suppression of *mtPDHK* show 2-3 times more plant and oil productivity and an enhanced harvest index at high CO₂ (700 ppm vs 380 ppm) (Grodzinski et al., 2011; Weraduwege, 2013; Dahal et al., 2014). Further work is in progress, but these initial findings suggest that the anabolic properties of dark respiration can be harnessed, allowing additional photosynthate produced at higher atmospheric CO₂ to be more efficiently diverted to the production of plant oils.

Gene Stacking

Que et al. (2010) have recently reviewed the current status of trait stacking in crops like canola, cotton and corn. They have focused on the myriad of transgene combinations (formed by conventional breeding, i.e., the crossing of a transgenic plant (event) containing individual transgenes with other event(s) containing single or double transgenic traits) having to do with stacked trait cultivars with increasing numbers of genes for insect- and herbicide- tolerance to expand the scope of insect pests and weeds that can be combatted (see Que et al., (2010), their Table 1). As reported by James (2012), stacked traits are an important feature of biotech crops: 13 countries planted biotech crops with two or more traits in 2012. Globally, around 33 million ha, equivalent to 26% of the 170 million ha devoted to biotech crops, were stacked trait

plantings. The need to stack multiple gene events is becoming more complex than ever in this realm; for example, in corn, at least 8 trait genes have been combined to provide for combined weed control and at least two modes of action for controlling four major pests (Genuity/SmartStax™ by Monsanto and Dow AgroSciences).

With respect to gene stacking, Taverniers et al., in a 2008 review have said: “....*there are many types, definitions, and perceptions of stacking. These include: (1) stacking of traits and (2) stacking of events, which are the most widely accepted perceptions of stacking, and (3) stacking of genes, which from the analytical and traceability point of view may be a more appropriate perception.*”, and herein, we have adopted definition (3) for purposes of discussion and refer to it as “gene stacking”.

Multi-gene expression cassette transformations where a single construct harbours two or more genes can be performed relatively easily. Commercial examples include Herculex™ I, Herculex™ RW and Agrisure™ CB/LL in corn and Vistive™ Gold in soybean (International Service for the Acquisition of Agri-biotech Applications, 2013). A new approach: trait stacking via targeted genome editing, has been reported in maize by Ainley et al. (2013). They demonstrated that the combination of high-efficiency targeted genome editing driven by engineered zinc finger nucleases (ZFNs) with modular ‘trait landing pads’ (TLPs) would allow ‘mix-and-match’, on-demand transgene integration and trait stacking in crop plants. They illustrated the utility of nuclease-driven TLP technology by applying it to the stacking of two herbicide resistance traits. Both herbicide resistance traits co-segregated, thereby demonstrating a TLP linkage of the two independently transformed transgenes.

Gene stacking is especially applicable in metabolic engineering of plants since most metabolic processes and biochemical pathways involve numerous genes interacting with each other. Multi-gene transfers have enabled the import of entire metabolic pathways, the expression of entire protein complexes and the development of transgenic crops simultaneously engineered to produce a spectrum of added-value compounds (Naqvi et al., 2009). Commercial successes include the engineering of the entire pathway for provitamin A (beta carotene) biosynthesis in rice endosperm by stacking three carotenoid genes (“Golden Rice”, Ye et al., 2000) and the “Blue Rose”, possessing modified flower color, produced by stacking two genes in the anthocyanin biosynthetic pathway to alter flower pigmentation. This gives the biotech rose flowers the capacity to accumulate delphinidin-based anthocyanins, resulting in novel shades of blue rose petals (Tanaka et al., 2009). The transgenic expression of a multi-gene construct has resulted in the reconstitution of the polyhydroxybutyrate biosynthetic pathway in *Arabidopsis* (Kourtz et al., 2007) and in the high biomass crop, switchgrass (Somleva et al., 2008).

Gene Stacking for Improvements in Oil Composition

In the area of seed lipid biosynthesis, there are numerous reports of stacking genes for oil quality/fatty acid composition. This topic has been extensively reviewed of late (Haslam et al., 2013) and so, only a few selected examples follow:

In *B. napus*, simultaneous over-expression of an acyl-ACP thioesterase (*FatB*) and silencing of eight putative acyl-acyl carrier protein desaturases (SADs) using artificial microRNAs, was shown to increase saturated fatty acid proportions in the seed oil (Sun et al.,

2014). Total saturated fatty acid content was increased from 7.4% in the control to 37.3–45.6% in the transformed lines.

Upon expression in canola, the stacking of $\Delta 12$ and $\Delta 6$ desaturases from the fungus *Mortierella alpina* led to seeds accumulating of over 40% γ -linolenic acid (GLA) and 2% stearidonic acid (SDA) (Liu et al., 2001). To increase metabolic flux toward SDA synthesis, Ursin (2003) stacked the *M. alpina* *FAD2* and *FAD6* with the *B. napus* $\Delta 15$ desaturase (*FAD3*) genes, which resulted in GLA and SDA accumulations of 23% and 15%, respectively.

Value-added seed oil compositional changes have been made in soybean by gene stacking: suppressing the expression of all three *GmFAD3* genes with a single RNA interference construct generated an ultra-low linolenic phenotype. Simultaneous down-regulation of soybean *FAD2-1A* and *-1B* resulted in lines with oleic acid contents exceeding 80% (Clemente and Cahoon, 2009). The coordinated expression of the *FAD6* from borage with the *FAD3* from *Arabidopsis* produced an oil with GLA, linolenic acid, and SDA proportions of 5.8%, 33.5%, and 21.6%, respectively (Eckert et al., 2006).

The most successful examples in the extreme have been in engineering the production of long-chain PUFAs in members of the *Brassicaceae*. Wu et al., (2005). produced transgenics containing significant amounts of arachidonic acid (AA) and eicosapentaenoic acid (EPA) by a step-wise metabolic engineering strategy: By a series of transformations in *B. juncea* using constructs containing increasing numbers of transgenes, they achieved incremental production of AA up to 25% and EPA up to 15%. More remarkably, they re-constituted the entire DHA biosynthetic pathway in *B. juncea* using a single construct containing 9 transgenes, leading to the production of EPA levels up to 15% and DHA levels up to 1.5% in seeds. More recently, the engineering of *Arabidopsis* to produce long-chain PUFAs, involved the introduction of a completely new pathway comprised of five additional enzymatic steps, plus enhancement of endogenous precursor pathways (Petrie et al., 2010, 2012; Ruiz-Lo'pez et al., 2012), perhaps the most complex metabolic engineering result achieved in plants to date. The amount of DHA produced exceeded 12% which is the level at which DHA is generally found in bulk fish oil. By projection, if transferred to canola, the authors reported that one hectare of a crop containing 12% DHA in the seed oil would produce as much DHA as approximately 10,000 fish. Clearly, this is a breakthrough in the development of sustainable alternative sources of DHA, and this was demonstrated by the subsequent transfer of this DHA pathway to the new oilseed crop, *Camelina sativa* (Petrie et al., 2014) which has yields as high as canola and mustards.

Gene Stacking for Improvements in Oil Content

There have been few reports of oil content enhancement through stacking transgenes in oilseeds. In *Arabidopsis*, *LEAFY COTYLEDON1* (*LEC1*) and *LEC1-LIKE* (*L1L*) are key regulators of fatty acid biosynthesis. Over-expression of *AtLEC1* orthologs from *B. napus* (*BnLEC1* and *BnL1L*), caused an increased fatty acid level in transgenic *Arabidopsis* plants, but these exhibited severe developmental abnormalities. Using a truncated napin A promoter (which retains the seed-specific expression pattern but with a reduced potency) to drive the co-expression of *BnLEC1* and *BnL1L* in transgenic canola resulted in proportional increases of 2–20% in the seed oil content without any detrimental effects on major agronomic traits (Tan et al., 2011). Co-expression of *WRINKLED 1* and *DGAT1* in *Arabidopsis* under control of either

the *A. thal* FAE1 or *B. napus* napin promoter gave 8-11% relative seed oil content increases compared to nt-WT (Kim et al., 2012).

Recently, it has been shown that Phospholipid: Diacylglycerol Acyltransferase1 (PDAT1) is critical for TAG biosynthesis in leaves. Over-expression of *PDAT1* in *Arabidopsis* leaves led to oil droplet fusion and thus over-expansion, while ectopic expression of oleosin was found to promote the clustering of small oil bodies. When ectopically co-expressed, *PDAT1* plus *oleosin* genes enhanced *Arabidopsis* leaf TAG content by up to 6.4% of the dry weight without affecting membrane lipid composition and plant growth (Fan et al., 2013). The extreme Xerophyte *Tetraena mongolica* Maxim has high oil content in the stems. In a similar vein to the Fan et al., study, two *DGAT1*s from this species were cloned and overexpressed. In a 2-gene construct containing both *TmDGAT1a* and *TmDGAT1b*, increases in oil content in soybean hairy roots ranged from 37% to 85%, while co-expression in *T. mongolica* calli resulted in increases in oil from 59% to 108%. Accompanying altered fatty acid profiles had an increase in linoleic and palmitic acids with a dramatic decrease in 18:3. The authors conclude that combined over-expression of these two genes may both increase oil content and alter fatty acid composition (Li et al., 2013). However, it is difficult to compare these examples of stacking results with ours, as in the former, the tissues chosen for expression are not cotyledons/seeds but vegetative (leaves or roots), or undifferentiated calli. Nonetheless, these gene combinations hold promise as a means to enhance the oil content of vegetative biomass for the energy and animal feed sectors.

In multi-step biosynthetic pathways there are usually rate-limiting steps that control the flux through the whole pathway. In its simplest form, this can be overcome by either increasing the amount of the rate-limiting enzyme or by expressing a variant enzyme that is more active (Moeller and Wang, 2008).

We made use of both principles in the current study: our combined gene transfer approach to enhance *B. napus* oil content included over-expression of a variant gene encoding one of the key rate-limiting steps in the Kennedy pathway for oil assembly (SDM *T. majus* *DGAT1*) with anti-sense suppression of *mtPDCK*, resulting in up-regulation of the activity of *mtPDH*, which universally controls the flux of photosynthetic carbon through the mitochondrial respiration pathway. Our stacking results are superior to the best individual transgenic effects using single gene constructs to over-express either the *A. thaliana* *DGAT1*, the *TmDGAT1*, or its variant SDM *S*¹⁹⁷-to-*A*¹⁹⁷ *TmDGAT1* in *Arabidopsis* or canola, or to effect suppression of *mtPDCK* in canola. Indeed, the enhancement of oil content (almost 24%) was near-cumulative in the 2-gene stack lines compared to the best single gene events (Figure 10). Furthermore, the proportional improvement in seed oil content in our best 2-gene transgenic lines are consistently high (>22%), and therefore somewhat superior to the best effect afforded by the co-expression of the *BnLEC1* and *BnLec1*-like genes in canola (2-20%; Tan et al., 2011) or *wrinkled 1* + *DGAT1* in *Arabidopsis* seed (Kim et al., 2012).

Effects on Seed Weight

There do not appear to be any directed studies of the effects of gene stacking to specifically address seed weights in oilseeds.

To our knowledge, the most dramatic increase in seed size in canola has been by a single gene transfer- the seed-specific expression of the *Rev* gene (plant growth and/or development -

associated gene, *Revoluta*TM) which resulted in 5.6 to 59.8% increments relative to their null siblings in field trials (Derocher and Nugyen, 2007; Patent application CA2633988A1).

In our studies, there were significant increases in average seed weight upon expression of the *A. thaliana* *DGAT1* in WT *Arabidopsis* (averaging 20.6% in multiple and 5% in single insert lines) compared to non-transformed pSE controls (Jako et al., 2001). However, there were little or no consistent differences in this trait in *AtDGAT1* *B. napus* transgenics in the greenhouse. Similarly, the *A/S mtPDCK Arabidopsis* transgenics showed strong average seed weight increments of up to 16% upon constitutive expression and up to 23% under napin control (Marillia et al., 2003), but the trait did not translate well to *B. napus* (data not shown).

In our *B. napus* two-gene transgenic lines, there was no significant difference with respect to average seed weight compared to the controls, despite the fact that oil content was markedly improved.

Why the individual or stacked transgenes do not have the same stimulating effect on seed weight in *B. napus* as they do in *Arabidopsis* may be partially due to the fact that the two hosts are propagated in a dissimilar way. *B. napus* transgenics are propagated, partially pruned and bagged (to ensure “selfing” and not out-crossing) during seed development and through to maturity in the greenhouse. This more than likely affects the overall “yield”. This may be the most important difference; in *Arabidopsis*, even though plants are coned at the bolting stage until seed harvest, there is no pruning during the seed development phase. Recently, it has been suggested that in *Arabidopsis*, starch turnover is functionally linked to cell division and differentiation rather than to developmental or storage functions specific to embryos; the situation is more complicated than previously thought. The pathways of embryo starch metabolism are similar in several respects to those in *Arabidopsis* leaves an entirely new way of looking at the role of embryonic starch (Vasilios et al., 2010). Thus, while the reasons for the differential effects of *DGAT1* over-expression and *mtPDCK* suppression on seed weight between *Arabidopsis* and *B. napus* require further investigation, it may, to some extent, have to do with differences between the two species in the degree to which these genetic modifications affect how starch is partitioned and channeled between oil synthesis and cell division/proliferation in the developing seed. This requires further comparative study.

Effects of the 2-Gene Stack on Protein Content

We compared the protein contents of the highest and the lowest oil increase lines with the control. Line 6-19M-8-1 and 6-18P-10-4 had average protein contents of $36.7 \pm 0.9\%$ and $31.0 \pm 1.4\%$, respectively, compared to the control, which had a protein content of $33.2 \pm 0.2\%$. Thus, there were interesting differences compared to controls: In the case of the high oil increase line 6-19M-8-1, protein was also significantly increased compared to the control, while in the lowest oil increase line 6-18P-10-4, the protein content was slightly lower than the control.

What is not clear, for example, is why line 6-19M-8-1 has proportional increments in both oil and protein compared to the controls and yet average seed weights are not significantly different. This would suggest that other factors contributing to seed weight such as hull and integument thicknesses, and the proportions of secondary seed components like phytate and sinapine, may be differentially affected in the 2-gene transgenics compared to the controls.

FUTURE STUDIES AND CONCLUSION

We anticipate that proper assessment of the trends in seed yield in our 2-gene stack *B. napus* transgenics will require large-scale field trails which are planned for the future. This will also give us a better opportunity to assess the effects of the combined genes on the collective proportions of oil, protein and other seed components in a field environment. This will be the true test of the value of combining the two technologies. Additionally, top-down metabolic flux analyses (Ramli et al., 2002; Weselake et al., 2008) of our best T₃ two-gene stack lines are currently underway to examine the relative contributions of up-regulation of fatty acid synthesis activity (Block A) vs. TAG assembly activity (Block B), with respect to the enhanced oil phenotypes.

In conclusion, our results show that at the greenhouse level, stacking *DGAT1* with suppression of *mtPDCK* is a means to effect strong improvements in oil content in canola. While rigorous field testing is still required, this approach holds promise for testing in other oilseed crops as well.

ACKNOWLEDGMENTS

The authors acknowledge the contributions of past collaborators to some of the single-gene work reviewed in this manuscript, including the labs of Dr. R.J. Weselake (Universities of Lethbridge and Alberta), Dr. J. Harwood (Cardiff University) and Drs. B. Micallef and B. Grodzinski (University of Guelph). We also thank Dr. T. Fawcett for the seed protein analysis of the 2-gene transgenic lines. This is NRCC manuscript # 56088.

REFERENCES

- Ainley, W.M., Sastry-Dent, L., Welter, M.E., Murray, M.G., Zeitler, B., Amora, R., Corbin, D.R., Miles, R.R., Arnold, N.L., Strange, T.L., Simpson, M.A., Cao, Z., Carroll, C., Pawelczak, K.S., Blue, R., West, K., Rowland, L.M., Perkins, D., Samuel, P., Dewes, C.M., Shen, L., S., Sriram, Evans, S.L., Rebar, E.J., Zhang, L., Gregory, P.D., Urnov, F.D., Webb, S.R. & Petolino, J.F. (2013) Trait stacking by targeted genome editing. *Plant Biotechnol. J.* 11, 1126-1134.
- Andre, C., Froehlich, J.E., Moll, M.R. & Benning, C. (2007). A heteromeric plastidic pyruvate kinase complex involved in seed oil biosynthesis in *Arabidopsis*. *Plant Cell*, 19, 2006-2022.
- Bates, P.D. & Browse, J. (2012). The significance of different diacylglycerol synthesis pathways on plant oil composition and bioengineering. *Front. Plant Sci.*, 3:14 doi: 10.3389/fpls.2012.0014.
- Bates, P.D., Stymne, S. & Ohlrogge, J. (2013). Biochemical pathways in seed oil synthesis. *Current Opinion in Plant Biology*, 16, 358–364.
- Baud, S. & Lepiniec, L. (2010). Physiological and developmental regulation of seed oil production. *Progr. Lipid Res.*, 49, 235-249.

- Bechtold, N. & Pelletier, G. (1998). In *Planta Agrobacterium*-mediated transformation of adult *Arabidopsis thaliana* plants by vacuum infiltration. *Arabidopsis Protocols, Methods in Molecular Biology*, Volume 82, 259-266. New York, NY: Humana Press.
- Bevan, M. (1983) Binary *Agrobacterium* vectors for plant transformation. *Nucleic Acid Res.*, 12, 8711–8721.
- Bouvier-Nave, P.P., Benveniste, P., Oelkers, S.L., Sturley, S.L. & Schaller, H. (2000). Expression in yeast and tobacco of plant cDNAs encoding acyl CoA:diacylglycerol acyltransferase. *Eur. J. Biochem.*, 267, 85-96.
- Buchanan, B. B., Gruissem, W. & Joner, R. (2000). *Biochemistry and Molecular Biology of Plants*, Hoboken, NJ: John Wiley and Sons Inc.
- Cao, Z., Tian, F., Wang, N., Jiang, C., Lin, B., Xia, W., Shi, J., Long, Y., Zhang, C. & Meng, J. (2010). Analysis of QTLs for erucic acid and oil content in seeds on A8 chromosome and the linkage drag between the alleles for the two traits in *Brassica napus*. *J. Genet. Genomics*, 37, 231-40.
- Chen, M., Qi, W.C., Wang, S.Y., Guan R.Z. & Zhang, H.S. (2011). Correlation of Kennedy pathway efficiency with seed oil content of canola (*Brassica napus* L.) lines. *Can. J. of Plant Sci.*, 91, 251-259.
- Clemente, T.E. & Cahoon, E.B. (2009). Soybean oil: Genetic approaches for modification of functionality and total content. *Plant Physiol.*, 151, 1030-1040.
- Covello, P.S., Reaney, M.J.T. & MacKenzie, S.L. (2000) *Nucleotide sequences coding an enzymatic protein; for use in the production of a colourless oil used in the cosmetics and health related fields*. US patent: US 6153815 A.
- Dahal, K., Weraduwege, S.M., Kane, K., Rauf, S.A., Leonardos, E.D., Gadapati, W., Savitch, L., Singh, J., Marillia, E-F., Taylor, D.C., Micallef, M.C., Knowles, V., Plaxton, W., Barron, J., Sarhan, F., Hüner, N., Grodzinski, B. & Micallef, B.J. (2014). Enhancing biomass production and yield by maintaining enhanced capacity for CO₂ uptake in response to elevated CO₂. *Can. J. Plant Sci.*, 94, 075–1083.
- Datta, R.S.S., Hammerlindl, J.K., Panchuk, B., Pelcher, L.E. & Keller, W.A. (1992). *Modified binary plant transformation vectors with the wild-type gene encoding NPTIII*. *Gene*, 211, 383–384.
- Delourme, R., Falentin, C., Huteau, V., Clouet, V., Horvais, R., Gandon, B., Specel, S., Hanneton, L., Dheu, J.E., Deschamps, M., Margale, E., Vincourt, P. & Renard, M. (2006). Genetic control of oil content in oilseed rape (*Brassica napus* L.). *Theor. Appl. Genet.*, 113, 1331-45.
- Derocher, J. & Nguyen, T. (2007). *Increased seed size and seed number through transgenic over expression of revoluta during early embryo development*. Patent Application: CA 2633988 A1, published July 12, 2007.
- Eckert, H., LaVallee, B.J., Schweiger B.J., Kinney, A.J., Cahoon, E.B. & Clemente, T. (2006). Co-expression of the borage $\Delta 6$ desaturase and the *Arabidopsis* $\Delta 15$ desaturase results in high accumulation of stearidonic acid in the seeds of transgenic soybean. *Planta*, 224, 1050–1057.
- Elhiti, M., Yang, C., Chan, A., Durnin, D.C., Belmonte, M.F., Ayele, B.T., Tahir, M. & Stasolla, C. (2012). Altered seed oil and glucosinolate levels in transgenic plants overexpressing the *Brassica napus* SHOOTMERISTEMLESS gene. *J. Exp. Bot.*, 63, 4447-4461.

- Fan, J., Yan, C., Zhang, X. & Xu, C. (2013). Dual role for phospholipid:diacylglycerol acyltransferase: enhancing fatty acid synthesis and diverting fatty acids from membrane lipids to triacylglycerol in *Arabidopsis* leaves. *The Plant Cell*, 9, 3506-3518.
- Grodzinski, B., Weraduwege, S., Taylor, D.C., Marillia, E-F. & Micallef, B. (2011). Roles of dark respiration in plant growth and productivity. In: A. Moo-Young (Ed.), *Comprehensive Biotechnology 2-Vol. 4: Agricultural and Related Biotechnologies*, Section 1: *Plant Systems*, pp 197-207. ScienceDirect.
- Halford, N.G. & Hardie, D.G. (1998). SNF1-related protein kinases: global regulators of carbon metabolism in plants? *Plant Mol. Biol.*, 37, 735-748.
- Haslam, R.P., Ruiz-Lopez, N., Eastmond, P., Moloney, M., Sayanova, O. & Napier, J.A. (2013). The modification of plant oil composition via metabolic engineering—better nutrition by design. *Plant Biotechnol. J.*, 11, 157–168.
- Hobbs, D.H., Lu, C. & Hills, M.J. (1999). Cloning of a cDNA encoding diacylglycerol acyltransferase from *Arabidopsis thaliana* and its functional expression. *FEBS Letts.*, 452, 145-149.
- Ichihara, K., Takahashi, T. & Fujii, S. (1988). Diacylglycerol acyltransferase in maturing safflower seeds: its influences on the fatty acid composition of triacylglycerol and on the rate of triacylglycerol synthesis. *Biochim. Biophys. Acta.*, 958, 125-129.
- International Service for the Acquisition of Agri-biotech Applications (2013), *ISAAA Brief 46-2013: Executive summary, Global status of commercialized biotech/GM crops:2013*. <http://www.isaaa.org/resources/publications/briefs/46/executivesummary/>.
- Jako, C., Kumar, A., Wei, Y., Zou, J., Barton, D.L., Giblin, E.M., Covello, P.S. & Taylor, D.C. (2001). Seed-specific over-expression of an *Arabidopsis* cDNA encoding a diacylglycerol acyltransferase enhances seed oil content and seed weight. *Plant Physiol.*, 126, 861-874.
- James, C. (2012). Brief 44 - Executive Summary Global Status of Commercialized Biotech/GM Crops: 2012. *ISAAA International Service for the Acquisition of Agri-biotech Applications*. Pp 1-12. <http://www.isaaa.org>.
- Jiang, C., Shi, J., Li, R., Long, Y., Wang, H., Li, D., Zhao, J. & Meng, J. (2014). Quantitative trait loci that control the oil content variation of rapeseed (*Brassica napus* L.). *Theor. Appl. Genet.*, 127, 957-68.
- Josefsson, L-G., Lenman, M., Ericson, M.L. & Rask, L. (1987). Structure of a gene encoding the 1.7S storage protein, napin, from *Brassica napus*. *J. Biol. Chem.*, 262, 12196–12201.
- Katavic, V., Reed, D.W., Taylor, D.C., Giblin, E.M., Barton, D.L., Zou, J., MacKenzie, S.L., Covello, P.S. & Kunst, L. (1995). Alteration of seed fatty acid composition by an ethyl methanesulfonate-induced mutation in *Arabidopsis thaliana* affecting diacylglycerol acyltransferase activity. *Plant Physiol.*, 108, 399-409.
- Katavic, V., Friesen, W., Barton, D.L., Gossen, K.K., Giblin, E.M., Luciw, T. An, J., Zou, J., MacKenzie, S.L., Keller W.A., Males, D. & Taylor, D.C. (2001). Improving erucic acid content in rapeseed through biotechnology: What can the *Arabidopsis* *FAE1* and the Yeast *SLC1-1* genes contribute? *Crop Science*, 41, 739-747.
- Kennedy, E.P. (1961). Biosynthesis of Complex Lipids. *Fed. Proc. Amer. Soc. Exp. Biol.*, 20, 934-940.
- Kim, H., Kim, Hyun, U. & Suh, M.C. (2012). Efficiency for increasing seed oil content using *WRINKLED1* and *DGAT1* under the control of two seed-specific promoters, *FAE1* and *Napin*. *J. Plant Biotechnol.*, 4, 242-252.

- Kim, H.U., Jung, S.-J., Lee, K.R., Kim, E.H., Lee, S.M., Roh, K.H. & Kim, J.B. (2013). Ectopic overexpression of castor bean *LEAFY COTYLEDON2* (*LEC2*) in *Arabidopsis* triggers the expression of genes that encode regulators of seed maturation and oil body proteins in vegetative tissues. *FEBS Open Bio.*, 4, 25-32.
- Koncz, C. & Schell, J. (1986). The promoter of TL-DNA gene 5 controls the tissue-specific expression of chimaeric genes carried by a novel type of *Agrobacterium* binary vector. *Mol. Gen. Genet.* 204, 383-396.
- Kourtz, L., Dillon, K., Daughtry, S., Peoples, O.P. & Snell, K.D. (2007). Chemically-inducible expression of the PHB biosynthetic pathway in *Arabidopsis*. *Transgenic Res.*, 16, 759–769.
- Lardizabal, K., Effertz, R., Levering, C., Mai, J., Pedroso, M.C., Jury, T., Aasen, E., Gruys, K. & Bennett, K. (2008) Expression of *Umbelopsis ramanniana* *DGAT2A* in seed increases oil in soybean. *Plant Physiol.*, 148, 89–96.
- Li, M., Zhao, M., Wu, H., Wu, W., & Xu, Y. (2013). Cloning, characterization and functional analysis of two type 1 diacylglycerol acyltransferases (DGAT1s) from *Tetraena mongolica*. *J. Integr. Plant. Biol.* 55, 490-503.
- Liu, J.W., DeMichele, S., Bergana, M., Bobik, E. Jr., Hastilow C., Chuang, L.T., Mukerji, P. & Huang, Y.S. (2001). Characterization of oil exhibiting high γ -linolenic acid from a genetically transformed canola strain. *J. Am. Oil Chem. Soc.*, 78, 489–493.
- Liu, J., Hua, W., Yang, H.-L., Zhan, G.-M., Li, R., Deng, L.-B., Wang, X.-F., Liu, G.-H. & Wang, H.-Z. (2012). The *BnGRF2* gene (*GRF2*-like gene from *Brassica napus*) enhances seed oil production through regulating cell number and plant photosynthesis. *J. Exp. Bot.*, 63, 3727-3740.
- Liu, J., Hua, W., Zhan, G., Wei, F., Wang, X., Liu, G. & Wang, H. (2010). Increasing seed mass and oil content in transgenic *Arabidopsis* by the overexpression of *wri1*-like gene from *Brassica napus*. *Plant Physiol. Biochem.*, 48 9-15.
- Luethy, M. H., Miernyk, J.A., David, N.R. & Randall, D.D. (1996) Plant pyruvate dehydrogenase complexes, In: M.S. Patel, T.E. Roche & R.A. Harris (Eds.), *α -Keto Acid Dehydrogenase Complexes*. pp. 71-92. Basel, Switzerland: Birkhäuser-Verlag.
- Lung, A.-C. & Weselake, R.J. (2006). Diacylglycerol acyltransferase: A key mediator of plant triacylglycerol synthesis. *Lipids*, 41, 1073-1088.
- Maisonneuve, S., Bessoule, J.J., Lessire, R., Delseny, M. & Roscoe, T.J. (2010). Expression of rapeseed microsomal lysophosphatidic acid acyltransferase isozymes enhances seed oil content in *Arabidopsis*. *Plant Physiol.* 152, 670-84.
- Marillia, E.F., Micallef, B.J., Micallef, M., Weninger, A., Pederson, K.K., Zou, J. & Taylor, D.C. (2003). Biochemical and physiological studies of *Arabidopsis thaliana* transgenic lines with repressed expression of the mitochondrial pyruvate dehydrogenase kinase. *J. Exp. Bot.*, 54, 259-270.
- Marillia, E.-F., Zou, J. & Taylor, D.C. (1999). *Brassica* Pyruvate Dehydrogenase Kinase. US patent: US7214859 B2.
- Moeller, L. & Wang, K. (2008). Engineering with Pprecision: Tools for the new generation of transgenic crops. *Science*, 58, 391-401.
- Moloney, M.M., Walker, J.M. & Sharma, K.K. (1989). High efficiency transformation of *Brassica napus* using *Agrobacterium* vectors. *Plant Cell Reports*, 8, 238-242.
- Moreno-Pérez, A.J., Venegas-Calderón, M., Vaistij, F.E., Salas, J.J., Larson, T.R., Garcés, R., Graham, I.A. & Martínez-Force, E. (2012). Reduced expression of *FatA* thioesterases in

- Arabidopsis* affects the oil content and fatty acid composition of the seeds. *Planta*, 235, 629-639.
- Naqvi, S., Farre, G., Sanahuja, G., Capell, T., Zhu, C. & Christou, P. (2009). When more is better: multigene engineering in plants. *Trends in Plant Sci.*, 15, 48-56.
- Oakes, J., Brackenridge, D., Colletti, R., Daley, M., Hawkins D.J., Xiong, H., Mai, J., Screen, S.E., Val, D., Lardizabal, K., Gruys, K. & Deikman, J. (2011). Expression of fungal diacylglycerol acyltransferase2 genes to increase kernel oil in maize. *Plant Physiol.*, 155, 1146-1157.
- Perry H.J., Bligny, R., Gout, E. & Harwood, J.L. (1999). Changes in Kennedy pathway intermediates associated with increased triacylglycerol synthesis in oilseed rape. *Phytochemistry*, 52, 799-804.
- Perry, H.J. & Harwood, J.L. (1993a). Changes in the lipid content of developing seeds of *Brassica napus*. *Phytochemistry*, 32, 1411-1415.
- Perry, H.L. & Harwood, J.L. (1993b). Radiolabelling studies of acyl lipids in developing seeds of *Brassica napus*: use of [$1\text{-}^{14}\text{C}$] acetate precursor. *Phytochemistry*, 33, 329-333.
- Petrie, J.R., Shrestha, P., Mansour, M.P., Nichols, P.D., Liu, Q. & Singh, S.P. (2010). Metabolic engineering of omega-3 long-chain polyunsaturated fatty acids in plants using an acyl-CoA $\Delta 6$ -desaturase with $\omega 3$ -preference from the marine microalga *Micromonas pusilla*. *Metabolic Eng.* 12, 233-240.
- Petrie JR, Shrestha P, Zhou X-R, Mansour MP, Liu Q, et al. (2012) Metabolic engineering plant seeds with fish oil-like levels of DHA. *PLOS One* 7: e49165. doi: 10.1371/journal.pone.0049165.
- Petrie, J.R., Shrestha, P., Belide, S., Kennedy, Y., Lester, G., Liu, Q., Divi, U.K., Mulder, R.L., Mansour, M.P., Nichols, P.D., & Singh, S.P. (2014) Metabolic engineering of *Camelina sativa* with fish oil-like levels of DHA. *PLOS One* 9: 9(1): e85061. doi:10.1371/journal.pone.0085061.
- Polge, C. & Thomas, M. (2006). SNF1/AMPK/SnRK1 kinases, global regulators at the heart of energy control. *Trends in Plant Sci.*, 12, 20-28.
- Qiu, D., Morgan, C., Shi, J., Long, Y., Liu, J., Li, R., Zhuang, X., Wang, Y., Tan, X., Dietrich, E., Weihmann, T., Everett, C., Vanstraelen, S., Beckett, P., Fraser, F., Trick, M., Barnes, S., Wilmer, J., Schmidt, R., Li, J., Li, D., Meng, J. & Bancroft, I. (2006). A comparative linkage map of oilseed rape and its use for QTL analysis of seed oil and erucic acid content. *Theor. Appl. Genet.*, 114, 67-80.
- Que, Q., Mary-Dell, M., Chilton, C., Fontes, M., He, C., Nuccio, M., Zhu, T., Wu, Y., Chen, J.S. & Shi, L. (2010). Trait stacking in transgenic crops. Challenges and opportunities, *GM Crops*, 1, 220-229.
- Ramli, U.S., Baker, D.S., Quant, P.A. & Harwood, J.L. (2002). Control analysis of lipid biosynthesis in tissue cultures from oil crops shows that flux control is shared between fatty acid synthesis and lipid assembly. *Biochem. J.*, 364, 393-401.
- Roesler, K., Shintani, D., Savage, L. Boddupalli, S. & Ohlrogge, J. (1997). Targeting of the *Arabidopsis* homomeric acetyl-Coenzyme A carboxylase to plastids of rapeseeds. *Plant Physiol*, 113, 75-81.
- Routaboul, J.M., Benning, C., Bechtold, N., Caboche, M. & Lepiniec, L. (1999). The *TAG1* locus of *Arabidopsis* encodes for a diacylglycerol acyltransferase. *Plant Physiol. and Biochem.*, 37, 831-840.

- Rubin, P. M., & Randall, D. D. (1977). Regulation of plant pyruvate dehydrogenase complex by phosphorylation. *Plant Physiol.*, 60, 34-39.
- Ruiz-Lopez, N., Haslam, R.P., Usher, S.L., Napier, J.A. & Sayanova, O. (2013) Reconstitution of EPA and DHA biosynthesis in Arabidopsis: Iterative metabolic engineering for the synthesis of n-3 LC-PUFAs in transgenic plants. *Metabolic Eng.* 17, 30-41.
- Shen, B., Sinkevicius, K.W., Selinger, D.A. & Tarczynski, M.C. (2006). The homeobox gene *GLABRA2* affects seed oil content in Arabidopsis. *Plant Mol. Biol.*, 60, 377-87.
- Shi, L., Katavic, V., Yu, Y., Kunst, L. & Haughn, G. (2012) *Arabidopsis glabra2* mutant seeds deficient in mucilage biosynthesis produce more oil. *Plant J.*, 69, 37-46.
- Somleva, M.N., Snell, K.D., Beaulieu, J.J., Peoples, O.P., Garrison, B.R. & Patterson, N.A. (2008). Production of polyhydroxybutyrate in switchgrass, value-added co-product in an important lignocellulosic biomass crop. *Plant Biotechnol. J.* 6, 663-678.
- Stringham, G.R., Degenhardt, D.F., Thiagarajak, M.R. & Bonsal, V.K. (1995). Quantum summer rape. *Can. J. Plant Sci.*, 75, 403-404.
- Stymne, S. & Stobart, A.K. (1987). Triacylglycerol biosynthesis. In P.K. Stumpf (Ed.), *The Biochemistry of Plants, Vol. 9, Lipids: Structure and Function*. pp 175-214. New York, NY: Academic Press.
- Sun, J-Y., Hammerlindl, J., Forseille, L., Zhang, H. & Smith, M. A. (2014) Simultaneous over-expressing of an acyl ACP thioesterase (FatB) and silencing of acyl-acyl carrier protein desaturases by microRNAs increases saturated fatty acid levels in *Brassica napus* seeds. *Plant Biotechnol. J.* 12, 624-637.
- Sun, M., Hua, W., Liu, J., Huang, S., Wang, X., Liu, G. & Wang, H. (2012). Design of new genome- and gene-sourced primers and identification of QTL for seed oil content in a specially high-oil *Brassica napus* cultivar. *PLoS One*. 7 (10):e47037. doi: 10.1371/journal.pone.0047037.
- Tan, H., Yang, X., Zhang, F., Zheng, X., Qu, C., Mu, J., Fu, F., Li, J., Guan, R., Zhang, H., Wang, G. & Zuo, J. (2011). Enhanced seed oil production in canola by conditional expression of *Brassica napus* *LEAFY COTYLEDON1* and *LEC1-LIKE* in developing seed., *Plant Physiol.*, 156, 1577-1588.
- Tanaka, Y., Brugliera, F. & Chandler, S. (2009). Recent progress of flower colour modification by biotechnology. *Int. J. Mol. Sci.*, 10, 5350-5369.
- Taverniers, I., Papazova, N., Bertheau, Y., De Loose, M. & Holst-Jensen, A. (2008). Gene stacking in transgenic plants: towards compliance between definitions, terminology, and detection within the EU regulatory framework. *Environ. Biosafety Res.*, 7, 197-218.
- Taylor, D.C., Katavic, V., Zou, J., MacKenzie, S.L., Keller, W.A., An, J., Friesen, W., Barton, D.L., Gossen, K.K., Giblin, E.M., Ge, Y., Dauk, M., Luciw, T. & Males, D. (2001). Field-testing of transgenic rapeseed cv. Hero transformed with a yeast *sn-2* acyltransferase results in increased oil content, erucic acid content and seed yield. *Mol. Breeding*, 8, 317-322.
- Taylor, D.C., Zhang, Y., Kumar, A., Francis, T., Giblin, E.M., Barton, D.L., Ferrie, J.R., Laroche, A., Shah, S., Zhu, W., Snyder, C.L., Hall, L., Rakow, G., Harwood, J.L. & Weslake, R.J. (2009). Molecular modification of triacylglycerol accumulation under field conditions to produce canola with increased seed oil content. *Botany*, 87, 533-543.
- Taylor, D.C., Smith, M.A., Fobert, P., Meitkiewska, E. & Weslake, R.J. (2011). Metabolic Engineering of Higher Plants to Produce Bio-industrial Oils. In: A. Moo-Young (Ed.), *Comprehensive Biotechnology 2-Vol. 4: Agricultural and Related Biotechnologies, Section 1: Plant Systems*, pp 67-85. ScienceDirect.

- Tovar-Méndez, A., Miernyk, J.A. & Randall, D.D. (2003). Regulation of pyruvate dehydrogenase complex activity in plant cells. *Eur. J. Biochem.* 270, 1043-1049.
- Ursin, V.M. (2003). Modification of plant lipids for human health: development of functional land-based omega-3 fatty acids. *J. Nutr.*, 133, 4271–4274.
- Vasilios, M., Andriotis, E., Pike, M.J., Kular, B., Rawsthorne, S. & Smith, A.M. (2010). Starch turnover in developing oilseed embryos. *New Phytologist*, 187, 791–804.
- Vigeolas H., Waldeck, P., Zank, T. & Geigenberger, P. (2007). Increasing seed oil content in oil-seed rape (*Brassica napus* L.) by over-expression of a yeast glycerol-3-phosphate dehydrogenase under the control of a seed-specific promoter. *Plant Biotechnol. J.*, 5, 431–441.
- Wakao, S., Andre, C. & Benning, C. (2008). Functional analyses of cytosolic glucose-6-phosphate dehydrogenases and their contribution to seed oil accumulation in *Arabidopsis*. *Plant Physiol.*, 146, 277-288.
- Wang, C.S. & Vodkin L.O. (1994) Extraction of RNA from tissues containing high levels of procyanidins that bind RNA. *Plant Mol. Biol. Rep.*, 12, 132-145.
- Wang, X., Wang, H., Long, Y., Li, D., Yin, Y., Tian, J., Chen, L., Liu, L., Zhao, W., Zhao, Y., Yu, L. & Li, M. (2013). Identification of QTLs associated with oil content in a high-oil *Brassica napus* cultivar and construction of a high-density consensus map for QTLs comparison in *B. napus*. *PLOS ONE* Open access; <http://www.plosone.org/article/info%3Adoi%2F10.1371%2Fjournal.pone.0080569>.
- Wang, H., Guo, J. Lambert, K.N. & Lin, Y. (2007). Developmental control of *Arabidopsis* seed oil biosynthesis. *Planta*, 226, 773-783.
- Weraduwege, S.M. (2013). *Harnessing the anabolic properties of dark respiration to enhance sink activity at elevated CO₂ using Arabidopsis thaliana with partially-suppressed mitochondrial pyruvate dehydrogenase kinase*. Ph.D. Thesis. University of Guelph, Guelph, ON, Canada. https://atrium.lib.uoguelph.ca/xmlui/bitstream/handle/10214/6759/Weraduwege_Sarathi_201305_PhD.pdf?sequence.
- Weslake, R.J. (2005). Storage lipids. In D.J. Murphy (Ed.), *Plant Lipids, Biology, Utilisation and Manipulation*. Pp 162-225. Cardiff, UK: Blackwell Publishing Ltd.
- Weslake, R.J., Shah, S., Taylor, D.C., Zou, J., Laroche, A., Moloney, M., Rakow, G., Raney, J.P. & Harwood, J. (2007). Transformation of *Brassica napus* with diacylglycerol acyltransferase-1 results in increased seed oil content. In: C. Benning & J Ohlrogge (Eds.), *Current Advances in the Biochemistry and Cell Biology of Plant Lipids. Proceedings of the 17th International Symposium on Plant Lipids*, pp 232-234. Salt Lake City, UT: Aardvark Global Publishing Co., LLC.
- Weslake, R.J., Shah, S., Tang, M., Quant, P.A., Snyder, C.L., Furukawa-Stoffer, T.L., Zhu, W., Taylor, D.C., Zou, J., Kumar, A., Hall, L., Laroche, A., Rakow, G., Raney, P., Moloney, M.M. & Harwood, J.L. (2008). Metabolic control analysis is helpful for informed genetic manipulation of oilseed rape (*Brassica napus*) to increase seed oil content. *J. Exp. Bot.*, 59, 3543-3549.
- Weslake, R.J. Taylor, D.C., Rahman, H., Shah, S., Laroche, A., McVetty P.B.E. & Harwood, J.L. (2009). Increasing the flow of carbon into seed oil. *Biotechnol. Adv.*, 27, 866-878.
- Wu, G., Truksa, M., Datla, N., Vrinten, P., Bauer, J., Zank, T., Cirpus, P., Heinz, E. & Qiu, X. (2005). Stepwise engineering to produce high yields of very long-chain polyunsaturated fatty acids in plants. *Nature Biotechnol.*, 23, 1013–1017.

- Xu, J. Francis, T., Mietkiewska, E. Giblin, E.M., Barton, D.L., Zhang, Y., Zhang, M. & Taylor, D.C. (2008). Cloning and characterization of an acyl-CoA-dependent diacylglycerol acyltransferase 1 (*DGAT1*) gene from *Tropaeolum majus* and a study of the functional motifs of the DGAT protein using site-directed mutagenesis to modify enzyme activity and oil content. *Plant Biotechnol. J.*, 6, 799-818.
- Ye, X.D., Al Babili, S., Kloti, A., Zhang, J., Lucca, P., Beyer, P. & Potrykus, I. (2000). Engineering the provitamin A (beta-carotene) biosynthetic pathway into (carotenoid-free) rice endosperm. *Science*, 287, 303-305.
- Zhang, F.-Y., Yang, M-F. & Xu, Y-N. (2005). Silencing of DGAT1 in tobacco causes a reduction in seed oil content. *Plant Sci.*, 169, 689-694.
- Zhang, C., Iskandarov, U., Klotz, E.T. & Stevens, R.L. (2013). A *thraustochytrid* diacylglycerol acyltransferase 2 with broad substrate specificity strongly increases oleic acid content in engineered *Arabidopsis thaliana* seeds. *J. Exp. Bot.*, 64, 3189-3200.
- Zheng, P., Allen, W.B., Roesler, K., Williams, M.E., Zhang, S., Li, J., Glassman, K., Ranch, J., Nubel, D., Solawetz, W., Bhattaramakki, D., Llaca, V., Deschamps, S., Zhong, G-Y., Tarczynski, M.C. & Shen, B. (2008) A phenylalanine in DGAT is a key determinant of oil content and composition in maize. *Nat. Genet.* 40, 367-372.
- Zou, J., Katavic, V., Giblin, E.M., Barton, D.L., MacKenzie, S.L., Keller, W.A., Hu, X. & Taylor, D.C. (1997). Modification of seed oil content and acyl composition in *Brassicaceae* by expression of a yeast *sn-2* acyltransferase gene. *The Plant Cell*, 9, 909-923.
- Zou, J., Qi, Q., Katavic, V., Marillia, E-F & Taylor, D.C. (1999a). Effects of antisense repression of an *Arabidopsis thaliana* pyruvate dehydrogenase kinase cDNA on plant development. *Plant Mol. Biol.*, 41, 837-849.
- Zou, J., Wei, Y., Jako, C., Kumar, A., Selvaraj, G. & Taylor, D.C (1999b). The *Arabidopsis thaliana* *TAG1* mutant has a mutation in a diacylglycerol acyltransferase gene. *Plant J.*, 19, 645-653.

Chapter 31

PERIODONTITIS: A COMMON ORAL DISEASE THE GENETIC SUSCEPTIBILITY

Ying Zheng¹, You-Qiang Song² and W. Keung Leung^{1,*}

¹Faculty of Dentistry

²Department of Biochemistry,

Li Ka Shing Faculty of Medicine,

The University of Hong Kong, Hong Kong SAR, China

ABSTRACT

It is well known that genetic variations can in part affect human oral health. Periodontitis is a common dental noncommunicable disease (NCD). According to the World Health Organization, periodontitis affects 20% of the world population. Periodontal inflammation could eventually induce alveolar bone resorption, causing tooth mobility and tooth loss, which ultimately affects oral function, oral health and the individual's quality of life. Family study, twins study and linkage analysis in the early years revealed that genetic influence contributes to periodontal disease. Since the first single nucleotide polymorphisms (SNP) study on chronic periodontitis in 1997, many genetic loci were found to be associated with periodontitis, including IL-1, Fc gamma receptor and complement component 5 genes. Population structure or earlier investigations employing a less desirable sample size may lead to various limitations or bias and therefore diverse results. Meta-analysis reports have proved the association between periodontitis and SNPs in some regions, such as *IL1* and *HLA*. With development of the technologies, multiple genes can be analyzed in a single assay, and even whole genome SNPs can be screened. Seven genome-wide association studies (GWASs) investigated the potential genetic risk locus for periodontitis and found *GLT6D1* is significantly associated with aggressive periodontitis. In the future, whole genome sequencing, together with replication in multiple populations and functional studies, could potentially disclose the nature of periodontitis as an NCD.

* Corresponding Author's Email: ewkleung@hkucc.hku.hk (Professor W. K. Leung, Oral Diagnosis and Polyclinics, Room 1B25, 34 Hospital Road, Prince Philip Dental Hospital, Faculty of Dentistry, The University of Hong Kong, Hong Kong SAR, China; Tel: +852-2859-0417; Fax: +852-2858-2532).

Keywords: genetic predisposition to disease, genome-wide association study, periodontitis, polymorphism, single nucleotide

1. PERIODONTITIS - A BRIEF INTRODUCTION

Periodontitis is characterized by periodontal attachment destruction and alveolar bone loss. The common clinical symptoms include gum swelling, increased tooth mobility and excessive gingival recession, which reduce the chewing function, lead to aesthetic problems and affect the patient's quality of life.

It is one of the most common oral diseases and one of the major causes of tooth loss. The WHO reported in 2003 that up to 20% of the global human population has suffered from periodontal diseases (Petersen, 2003). The prevalence of 40-66.5% of subjects with community periodontal index of more than 2 in any one surveyed sextant, indicative of mild to severe periodontal inflammation, was reported in different adult populations of more than 30 years old. Approximately 8-11% of the total population is affected by severe chronic periodontitis (Hugoson et al., 2008; Eke et al., 2012).

Chronic periodontitis (CP) and aggressive periodontitis (AgP) are two of the major subtypes of periodontitis (Figure 1). CP is the most prevalent form of periodontitis in middle-aged to elderly adults and often has slow to moderate disease progression rate. It is caused by inadequate oral hygiene and may be modified by systemic diseases, such as diabetes, or other risk factors, like smoking or stress (Armitage, 2004). Different from CP, AgP is a periodontitis with prevalence ranging from 0.1% to 3.0% in adolescents and young adults, characterized by rapid progression and familial aggression (Eres et al., 2009).



Figure 1. Clinical cases of periodontitis: A) Chronic periodontitis – 42-year-old female suffering from severe form of the disease started approximately two years before presentation; B) Aggressive periodontitis – 20-year-old female suffering from generalized form of the disease. Human periodontitis in general is characterized by gum inflammation, tooth attachment apparatus breakdown including periodontal attachment loss, drifting and even tooth loss, which is observable from clinical pictures and radiographs of both cases shown. Note that case A lost the lower right second molar with radiographic sign of combined periodontal-endodontic lesion in lower left lateral incisor. On top of what was described earlier, both individuals have poor oral hygiene and calculus deposition.

Dental plaque or the oral biofilm, inhabited by various pathogens, is regarded as the etiological factor of periodontal disease. The subgingival bacterial biofilm, through interactions with periodontal lining cells, triggers the host inflammation and induces local innate and adaptive immune responses. However, periodontal destruction is not caused by the bacteria-host interaction *per se*, and not all individuals are equally susceptible to periodontitis despite the presence of plaque bacteria. People who never learn tooth brushing as a proper healthy lifestyle practice, however, are at higher risk to develop gingivitis or periodontitis but could remain periodontally healthy throughout life (Loe et al., 1986).

Either hyper- or hypo-responsiveness of the immune system, reflective of inappropriate host response toward periodontal disease pathogens, especially the overreaction, was believed to be the driver for destruction of the tooth-supporting tissue and eventually tooth loss, provided the disease remains untreated (Kinane et al., 2007). The variation of host susceptibility is determined by a multitude of factors. Smoking, diabetes, stress, gender, specific pathogenic bacteria and other risk factors have been reported to affect the pathogenesis and the severity of periodontitis (Grossi et al., 1994; Ng et al., 2006). Nevertheless, individual differences cannot be explained exclusively by the exposure to any one or a combination of these factors alone. Along such lines, genetic risks of periodontal disease are investigated to map out the individual periodontitis susceptibility variations.

2. REVIEW OF GENETIC STUDIES ON PERIODONTITIS

The genetic risks for periodontitis have been studied for decades. Periodontitis was once regarded as simple Mendelian disease in the early years. However, except for “periodontitis as a manifestation of systemic disease,” most periodontitis cases have been realized later as a complex common disease affected by multiple genes and environmental factors. The candidate gene approach and genome-wide study without pre-investigation hypothesis helped build some genetic background on periodontitis. Here we review the history of periodontitis genetic studies and summarize the appreciation of the situation so far.

2.1. Periodontitis Associated with Other Genetic Disorders

Genetic factors play an important role in some aggressive forms of periodontitis secondary to particular bodily conditions, which are termed “periodontitis as a manifestation of systemic disease.” Many of these systemic early-onset diseases are monogenic disease following Mendel’s Law, such as Papillon-Lefevre disease, Chediak-Higashi disease, hypophosphatasia, congenital and cyclic neutropenia, leukocyte adhesion deficiency, glycogen storage disease and Ehlers-Danlos disease. Mutation on Cathepsin C, lysosomal trafficking regulator, ALPL, ELANE, Beta-2 integrin chain, GDP-fucose transporter-1, SLC37A4 and Collagen alpha-1/2(V) genes are identified (Khoht & Albandor, 2014).

With such a profound genetic influence, these diseases often manifest in prepubescent and express early impact on the permanent and even primary teeth periodontium. It is not uncommon that some AgP cases have undetected or partial penetrant systemic conditions like Papillon-Lefevre disease (Hewitt et al., 2004).

2.2. Inheritability of Periodontal Disease: Family Aggregation and Twin Studies

“Periodontitis as a manifestation of systemic disease” is merely a subclass in the classification of periodontitis, while the majority of the periodontitis cases belong to CP, followed by AgP (Armitage, 1999). AgP used to be called “periodontosis, juvenile periodontitis, early onset periodontitis, or rapidly progressive periodontitis,” while CP is used to be called as “adult periodontitis.”

Although only AgP is characterized by familial aggregation, both periodontitis classes could present within families (Armitage, 2004). Family members of the patients with common noncommunicable diseases like periodontitis normally have a higher possibility to develop the disease concerned, as they are more likely to share the same genetic or environmental risk factors compared to the general population, more so if the genetic factor plays a key role in the etiology. A familial pattern of periodontitis was reported in twins and families for the first time in 1967 (Benjamin & Baer, 1967). A high occurrence of periodontal disease was found in the relatives of the patients with an “early onset” form of periodontitis (Butler, 1969; Hoffman, 1983; Beaty et al., 1987). The prevalence of moderate to severe periodontitis in the family of the proband ranged from 40-50% or even higher (Beaty et al., 1987; Boughman et al., 1992; Lopez, 1992).

All these family studies implied the possibilities of genetic background concerning periodontal disease. However, as mentioned earlier, familial aggregation of the disease may be the result of shared environment or genetic background, as both factors can contribute to progression of human periodontitis.

In the family-based studies above, some shared environmental factors are suggested to be one significant reason, such as transmission of periodontopathogens between family members (Vandervelden et al., 1993; Petit et al., 1994).

The proportion of the phenotypic variance determined by the genetic factors is referred to as the heritability of the disease concerned. In order to differentiate the genetic and environmental influences and to evaluate the heritability of periodontitis, twin studies were carried out until the late 1990s.

Theoretically, monozygotic twins share more similarities than dizygotic twins, while dizygotic twins are no more similar than normal siblings. Assuming the twins share the same environment, when periodontal disease occurs more often or more severely in monozygotic than dizygotic twins, it means that some genetic risk factors related to the disease may be in operation. In CP, the concordance rate of monozygotic twins experiencing periodontitis (0.23 to 0.38) is significantly higher than dizygotic twins (0.08 to 0.16) according to a self-reported twin study, which indicated that the similarity of disease experience is consistent with the genetic similarity (Corey et al., 1993). After adjustment of environmental and behavioral factors, similar results were repeated in other studies regarding both severity and extent of the disease. By comparing twins reared together and apart, a study with 110 pairs of adult twins estimated that heritability and, hence, genetic factors may contribute to 38~82% of the individual variance for periodontal status (Michalowicz et al., 1991). A heritability of 50% was reported in aggressive periodontitis as well. These early twin studies verified the genetic influence on periodontitis and provided some evidence of the disease heritability for future studies.

2.3. Heredity Mode and Heterogeneity: Segregation Analysis and Early Linkage Study

At the beginning, it was assumed that there was a major-effect gene for periodontal disease based on Mendel's law, so segregation studies and pedigree analysis were used to investigate its proposed inheritance modes. The first paper published in 1980, analyzing 129 first-degree relatives of 31 juvenile periodontitis or AgP probands, suggested an autosomal recessive model of inheritance (Saxen, 1980). This inheritance mode of periodontitis in Caucasians was advocated in some other studies (Fourel, 1972; Jorgenson et al., 1975; Beaty et al., 1987). In contrast, a study with more than 100 AgP families reported an autosomal-dominant transmission in both blacks and non-blacks, with a penetrance of 70% (Marazita et al., 1994). Besides autosomes, genders present an association with AgP as well. The excess of female proportion in the affected subjects was explained as an X-linked dominant form of AgP (Melnick et al., 1976; Page et al., 1985). It is surprising to see all these various hereditary modes suggested for periodontitis if the disease is considered to be passed on in a simple Mendelian fashion. The debates and conflicts on hereditary modes are actually caused by the wrong primary assumption. The heterogeneity of this complex disease could not be explained by one gene with major effect, so traditional approaches analyzing Mendelian diseases are not suitable for periodontitis. Moreover, neither twin studies nor these early segregation studies can locate the risk genes on the chromosomes.

In order to map the chromosomal location of disease-associated gene(s), linkage study protocols were adapted for periodontitis. They are based on the co-segregation of genetic markers and the disease-predisposing/-associating gene, because they are in physical proximity within the same chromosome and therefore have a high likelihood to pass on together in a pedigree. The genetic markers could be microsatellite, single nucleotide polymorphisms (SNPs) or a known disease-associated gene for another condition.

In 1986, a high co-occurrence of localized AgP and type III dentinogenesis imperfecta was reported in a family. A linkage analysis revealed that a known risk locus for the type III dentinogenesis imperfecta on chromosome 4q is closely linked with this localized AgP (Boughman et al., 1986). However, another study failed to verify this result, and chromosome 1 was suspected to be linked with AgP (Hart et al., 1993; Li et al., 2004). The controversies of all the early studies above forced people to reconsider the genetic nature of periodontitis and the genetic model of simple Mendelian inheritance regarding periodontitis was doubted. In fact all the results from previous studies suggested that the condition is influenced by complex multigenetic factors.

2.4. Complex Common Disease Associated with Single Nucleotide Polymorphisms

Since the previous studies have failed in identifying a major-effect risk gene, periodontitis is now realized to be a complex common disease. Based on the concept of "common disease common variances hypothesis," periodontitis should be influenced by multiple genetic polymorphisms. The hypothesis predicts that the genetic polymorphisms causing common disease should be frequent in the human population. Common diseases, such as periodontitis, hypertension and diabetes, are usually not immediately fatal at the early stage, so the allele

carriers can survive and possibly pass the genes to their offspring. Therefore, the frequencies of these genetic polymorphisms are stable in the population, and we refer to them as “common variance.” SNPs are the most common genetic variants. In periodontal diseases, especially CP, the effect of the each genetic variant is relatively small. Association study is a classical genetic method used to detect risk alleles with small to moderate effect in common diseases, while linkage study may not have enough power for this instance (Hodge, 1994; Cardon & Bell, 2001; Tabor et al., 2002).

2.4.1. Candidate Gene Association Study

Many studies adopted candidate gene association study in the last decades for periodontitis. In a candidate gene association study, the targeted locus is selected based on the current knowledge of the molecular etiology or plausibility for the disease. Genes regulating the pathogenic pathway are common candidates. The first paper of this kind concerning the periodontitis association study was published in 1997 and revealed the association between IL-1 polymorphisms and CP (Kornman et al., 1997). After that, hundreds of SNP association studies were done on periodontal disease. Candidate genes selected were mainly on the aspects of immuno-regulation or bone/collagen metabolism, such as genes encoding cytokines, cell surface receptors, chemokines and enzymes (Chai et al., 2012). The more intensively investigated polymorphisms of this sort are discussed as follows.

2.4.2. Cytokine

2.4.2.1. Interleukins (IL)

IL-1

IL-1 is an important inflammatory cytokine consisting of 13 subfamilies produced by monocytes, macrophages and dendritic cells. IL-1 α and IL-1 β are the two major subtypes, while the former mainly works as an intracellular regulator affecting the local inflammation and the latter is a released extracellular protein. A significant increase on the transcript levels of *IL1A*, *IL1B* and *IL1RN* genes was found in the gingival tissue of periodontitis patients (Braosi et al., 2012). The study on genetic polymorphism of *IL1* was one of the first SNP studies on periodontitis (Kornman et al., 1997). The polymorphism composition of IL-1 α -889 (rs1800587) and IL-1 β +3954 (rs1143634) was significantly associated with the severity of periodontitis in Caucasian nonsmokers with a reported high odds ratio of 18.9. After that, numerous studies were carried out on periodontitis and *IL1* polymorphisms, including IL-1 α -889/+4845 (rs1800587), IL-1 β +3953/4 (rs1143634), IL-1 β -511/-31 (rs16944) and IL1RN VNTR +2018 (rs2234677) (Laine et al., 2012).

A meta-analysis of 53 studies including 4178 cases and 4590 controls demonstrated a significant association of IL-1 α -889 and IL-1 β +3953/4 polymorphisms and a weak but positive association of IL-1 β -511 with CP but not AgP (Nikolopoulos et al., 2008). These two SNPs are verified to be associated with CP in Caucasians in a second meta-analysis (OR=1.48-1.54, MAF=0.22~0.28) (Karimbux et al., 2012), another systematic review also reported consistent and significant results of IL-1 α -889 in both Caucasians and Asians (Mao et al., 2013).

However, the frequencies of these SNPs are much lower in Asians (MAF of rs1800587=0.08~0.09 and MAF of rs1143634=0.02~0.04), though the systemic review above reported a positive result (Armitage et al., 2000; T. Kobayashi et al., 2009). A recent study in 2015 tried to ascribe it to the linkage disequilibrium between these SNPs with the functional SNPs. The working hypothesis was that these previously tagging alleles are variable among populations with different ethnic backgrounds but the true functional alleles will be consistent. By genotyping four SNPs in the *IL1B* promoter region (rs16944, rs1143623, rs1143633 and rs4848306), Caucasians, Africans, Hispanics and the Chinese population all presented a significant association with the IL-1 genes concerned (Wu et al., 2015).

IL-4 and IL-13

IL-4 is a cytokine, which triggers the differentiation of a naïve T cell to a T helper 2 cell and promotes the B lymphocyte-mediated immunity. Inflamed gingival tissues are enriched in macrophages but lack IL-4. Topical application of IL-4 increased the macrophage apoptosis in adult periodontitis and down-regulate CD14 expression in monocytes, a major receptor of lipopolysaccharides (LPS) from periodontopathogens (Shapira et al., 1992; Yamamoto et al., 1996).

Therefore, it was suggested that the absence of IL-4 contributes to periodontal tissue destruction. The low IL-4 concentrations in the gingival crevicular fluid (GCF) are also associated with periodontal disease progression (Kabashima et al., 1996; Tsai et al., 2007).

IL-13 is a cytokine that confers similar function on human monocytes as IL-4 (Yamazaki et al., 1997), which is not detectable in periodontitis lesions. The cytokine can activate TGF- β production in macrophages while acting directly or indirectly on fibroblasts, which activates collagen production and inhibits matrix metalloproteinase-3 synthesis in the latter, hence attenuating tissue destruction in periodontitis (Wynn, 2003; Fukuda et al., 2006).

Several studies have reported SNPs of IL-4 and IL-13 associated with periodontitis. In a recent study in the Chinese population, -590 variation of *IL4* was found to be significantly associated with CP (Loo et al., 2012). A correlation was found between IL-13 -1112 (rs1800925) and the susceptibility of AgP, with odd ratios ranging from 4.48 to 6.45 (Wu et al., 2010). All the SNPs studied were located in the promoter region of the genes concerned.

A recent meta-analysis summarized the association between periodontitis and IL-4 polymorphisms, namely -590 C/T (rs2243250), -33 C/T (rs2070874) and 70-bp polymorphisms. After analyzing 23 case-control studies (with 1,220 cases and 2,039 controls for -590 C/T, 715 cases and 967 controls for -33 C/T, 426 cases and 506 controls for 70-bp), the author concluded that IL-4 -590 T allele and TT genotype were significantly associated with an increased risk of periodontitis in whites but not in Asians (Yan et al., 2014). It may imply the SNP confers different influences on diverse ethnic groups, but the result should be interpreted with caution, as there was evidence indicating publication bias, which was mentioned as one of the study limitations by the authors.

IL-6

Produced by T cells, macrophages and even osteoclasts, IL-6 is an anti-inflammatory cytokine through its inhibitory effect on IL-1 and tumor necrosis factor- α (TNF- α). Salivary IL-6 showed a strong ability to predict an individual's likelihood to develop gingivitis (Lee et al., 2012). A few studies investigated the association between the genetic polymorphisms of IL-6 gene and periodontal disease. Multiple loci on the region of *IL6* were studied, including

IL-6 -174 (rs1800795), -190, -373, -572 (rs1800796), -597 (rs1800797), -1363 (rs2069827), -1480 and -6106 (Laine et al., 2012). A meta-analysis in 2013 accessed the association between periodontitis and two SNPs in the promoter region of the IL-6 gene. The review contained 16 studies on IL-6 -174 polymorphisms and 10 studies on IL-6 -572 polymorphisms and concluded that the significant result of -174 was detected only in the Brazilian population and that the polymorphisms on -572 may be associated with the increased risk of CP in Europeans (Song et al., 2013a).

IL-10

IL-10 is produced by both T cells and B cells. As a multi-functional cytokine, it can inhibit the expression of cytokines from T helper 1 cell, MHC class II antigens, and co-stimulatory molecules on macrophages. Meanwhile, it facilitates the differentiation, development and proliferation of B cells as well. A few loci of *IL10* have been reported to be associated with periodontal disease, such as polymorphisms on -1087/-1082(rs1800896), -627, -819/-824(rs1800871) and -592/-597(rs1800872) (Laine et al., 2012). Two meta-analyses, both published in 2012, presented the same observations, i.e., the polymorphisms of -819 and -592 of the IL-10 gene are associated with the susceptibility of periodontitis (Albuquerque et al., 2012; Zhong et al., 2012).

2.4.2.2. Tumor Necrosis Factor Alpha

TNF- α is a pro-inflammatory cytokine produced mainly by macrophages and a variety of other cells in the acute phase reaction to systemic or localized inflammation. The cytokine may induce the differentiation of osteoclast and contribute to bone resorption alongside IL-1 (Kobayashi et al., 2000; Loos et al., 2005). As an overreacted host response to the periodontopathogens, the increased local TNF- α leads to destruction of connective tissue and apoptosis of fibroblasts in the deeper part of gingival tissue (Graves & Cochran, 2003).

One SNP in the intron region (-489) and many in the promoter region [-238 (rs361525), -308 (rs1800629), -367, -857, -863 (rs1800630), -1031] of *TNFA* were investigated regarding their association with periodontitis (Laine et al., 2012). A recent meta-analysis studied the influence of TNF- α -308G/A, -863C/A and -238G/A polymorphisms on CP or AgP. The A allele at -308 and -863 was found to be associated with both kinds of periodontitis (Ding et al., 2014).

2.4.2.3. Cell Surface Receptors

The Fc Gamma Receptor Genes

The Fc γ receptor is a receptor protein on the surface of various cells, including B cells, dendritic cells, macrophages, neutrophil and natural killer cells. It specifically binds to the Fc part of the antibodies, such as IgG, and leads to phagocytosis and cytotoxicity of the IgG bound antigen/pathogen. There are nine subclasses in the family of Fc γ receptors (CD64: Fc γ RIa, Ib and Ic; CD32: Fc γ RIIa, IIb and IIc; CD16: Fc γ RIIIa and IIIb and Fc γ RIV). Strong IgG mediated responses to periodontopathogens in gingival tissue were observed along with Fc γ receptors commonly found in both gingival and pocket epithelium, indicating that the Fc γ receptor may play a crucial role in immune responses against bacteria causing periodontitis (Chai et al., 2012).

The polymorphisms of *FCGR1A*, *IIB*, *IIIA* and *IIIB* are investigated in both AgP and CP in various ethnic groups (Laine et al., 2012). A meta-analysis of 17 studies, consisting of 1685 cases and 1570 controls, investigated three SNPs, including *FcγRIIIa* H131R (rs1801274), *FcγRIIIa* F158V (rs396991) and *FcγRIIIb* NA1/NA2. Only the *FcγRIIIb* NA1/NA2 polymorphism presented a significant association with both kinds of periodontitis with the odds ratio ranging from 1.397 to 2.005 (Dimou et al., 2010). Many of the studies included in this review have a relatively small sample size. Our recent study on *Fcγ* receptor gene polymorphisms in 359 Hong Kong subjects reported that the C allele of rs445509 of *FCGR3A* provides a protective effect against chronic periodontitis, with an OR=0.3 after adjustment for age, gender and smoking (Chai et al., 2010).

Toll-Like Receptor

Toll-like receptors (TLR) are transmembrane proteins on host cells and play important roles in recognizing microbes or dangerous molecules. There are more than ten subclasses in the TLR family, of which TLR2 and TLR4 are the most extensively studied in periodontitis. The TLR4-dependant pathway can be activated by LPS of periodontopathic bacteria, like *Porphyromonas gingivalis* and *Capnocytophaga ochracea* (Yoshimura et al., 2002). The association between periodontitis and polymorphisms of *TLR2* and *TLR4* are investigated, including TLR2 Arg753Gln (rs57437087), Arg677Trp, Asp299Gly, Thr399Ile, -183, -148, -146, +1350 (rs3804100), +2343 and TLR4 Asp299Gly (rs4986790), Thr399Ile (rs498671), +3528, +3525, +4022, +4529. Three of these polymorphisms [TLR2 Arg753Gln (G2408A), TLR4 Asp299Gly (A896G) and Thr399Ile (C1196)] were investigated in a meta-analysis in 2013. After analyzing 14 studies, the author reported a lack of evidence regarding the association between these SNPs and periodontitis (Song et al., 2013b). However, an earlier meta-analysis in 2009 reported a different conclusion, that the TLR4 Asp299Gly allele A increases the risk of chronic periodontitis (OR=1.43, after analyzing seven studies), and the allele C of TLR4 Thr399Ile polymorphism shows a protective effect against aggressive periodontitis (OR=0.29, after analyzing four studies). The results were tested again using random effect models, which were more appropriate when the genetic heterogeneity was present. The significant result of TLR4 Asp299Gly was retained after the test, but TLR4 Thr399Ile was not (Ozturk & Vieira, 2009). The review in 2009 assessed the CP and AgP data separately, while the one in 2013 combined the two diseases. The inconsistency of these two reviews could be explained by the different individual studies included or by the fact that the TLR4 Asp299Gly is associated with CP but not AgP, so this polymorphism can present a positive result even in the analysis with a smaller sample size in the earlier review. Yet, more qualified studies with bigger sample sizes and a clear disease definition are required in the future.

Vitamin D Receptor

The Vitamin D receptor (VDR) is related to periodontitis based on the mechanism of the former influencing the bone mineral density, turnover and bone loss (Uitterlinden et al., 2006). Women with active or prior history of periodontitis have higher plasma vitamin D (25-hydroxyvitamin D-3) (Jabbar et al., 2011). There is a start codon polymorphism in *VDR*, consisting of three codons before a second start site (ATG). The corresponding genotype is determined by the restriction enzyme *FokI* and leads to some functional consequences in immune cells. This polymorphism has either a long form (f-VDR) or a shorter form (F-VDR)

protein. The difference between “f” and “F” is that the restriction site for *FokI* and the first ATG is present in f-VDR but is absent in the F-VDR. The presence of the shorter F-VDR results in higher NF-kappa B and NFAT-driven transcription, higher IL-12 p40 promoter-driven transcription and T lymphocytes with higher proliferation, which may therefore affect immune cell behavior and functions in immune-mediated diseases (van Etten et al., 2007).

Two meta-analyses were carried out but showed different results. The one published in 2011 summarized a total of 15 studies with 1338 cases and 1302 controls and assessed the SNPs of VDR *TaqI* (rs731236), VDR *BsmI* (rs1544410), VDR *FokI* (rs2228570) and VDR *Apal* (rs7975232). The latter analysis in 2012 included four more studies while focusing on the same SNPs. Both of the studies failed to detect any significant result in their overall analysis. However, in the subgroup of Asian subjects, the AA genotype of *Apal* was significantly associated with CP after Bonferreni’s correction for multiple comparisons in the first paper but not the second one. On the contrary, *TaqI* may decrease the risk of CP, and *FokI* presents a marginal protective effect against AgP under some genetic models in Asians (Deng et al., 2011; Chen et al., 2012). The difference may be caused by the unequal number of studies included, different inclusion and exclusion criteria, or diverse genetic models. The significant result in Asians but not Caucasians may be explained by the sample size, as there were twice as many Asian subjects as white subjects.

2.4.2.4. Matrix Metalloproteinases

Matrix metalloproteinases (MMPs) degrade extracellular matrix (ECM) components in connective tissue, including that of human periodontium. There are more than twenty members in this family. Most of them, functioning as enzymes that degrade collagens, have been detected in periodontal tissue and are related to gingival inflammation (Ingman et al., 1996). MMPs’ expressions are higher in the pathogen-induced inflamed periodontal tissue than normal healthy tissue. Clinicians use chemical MMP inhibitors including doxycycline as adjuncts to periodontal mechanical therapy (Sorsa et al., 2006). The polymorphisms of MMP1, MMP2, MMP3, MMP9 and MMP12 genes have been studied in periodontitis subjects (Laine et al., 2012).

MMP1 is the most generally expressed MMP against fibrillar collagens (collagen types I, II, and III). A meta-analysis containing 11 studies with 1580 periodontitis cases and 1386 controls indicated MMP1-1607dupG polymorphism associated with increased disease susceptibility (Li et al., 2013). Published in the same year, a similar result was reported in another meta-analysis considering 10 studies, with 8 of them also included in the Li et al. (2013) meta-analysis. Additionally, the Song et al. (2013b) group reported a negative association between *MMP9* and CP (Song, Kim, et al., 2013). Polymorphisms of the *MMP3* gene were reported to be significantly associated with CP in individual case-control studies in various populations, but no meta-analysis so far has been carried out (Astolfi et al., 2006; Letra et al., 2012; Li et al., 2012). Nevertheless, neither Chinese AgP nor Japanese CP patients appeared to have their disease associated with *MMP3* SNPs (Itagaki et al., 2004; Chen et al., 2007).

2.4.2.5. Summary of Candidate Gene Study

Besides the genes discussed above, there are many other polymorphisms studied on periodontal disease (Kinane & Hart, 2003; Kauppila, 2009; Laine et al., 2012). These candidate gene studies portray a rough sketch regarding the landscape of periodontitis genetic studies and inspire future directions for periodontal research. However, inconsistent results from various

studies, including meta-analyses, reveal the problem that candidate gene association studies suffered from insufficient sample size and, hence, low power, ethnic heterogeneity, and diverse inclusion and exclusion criteria based on different or loose disease classifications (Kinane & Hart, 2003; Yoshie et al., 2007; Nikolopoulos et al., 2008; Noack et al., 2008; Dimou et al., 2010; Ryder, 2010).

Ambiguous phenotype classification and various attention levels applied to confounding factors, such as smoking status and age, increased the difficulty of, and at the same time reduced the reliability of, meta-analyses. Repeated low-quality studies with less than ideal study design, limited sample size or ill-defined underlying pathogenic mechanisms of some candidate genes selected all turned out to be of low intellectual value. A large-scale replication study containing more than two thousand Caucasians in an attempt to verify the relevance of 23 SNPs previously reported to be periodontitis associated revealed that the initial positive results repeated on most of the candidate gene association studies were due to type I error (Vaithilingam et al., 2014). A sample size with a minimum of 1,000 well-defined cases is suggested to be an indispensable prerequisite for identification of a risk allele with an odds ratio of more than 1.3. Obviously, most of the previous candidate gene studies regarding periodontitis are underpowered (Schäfer et al., 2011). Instead of a single locus, a complete set of all haplotypes of the candidate gene shall be analyzed to cover all the potential causal alleles. Otherwise, follow-up fine mapping needs to be performed to evaluate the SNPs within linkage. Detailed information of confounding factors should be recorded and adjusted, whereas standard phenotype classification of either CP or AgP should be utilized in the candidate association studies in the future.

2.4.3. Genome-Wide Association Study (GWAS)

Genome-wide association study (GWAS) is based on the HapMap project that revealed more than 10 million common SNPs in 6 populations. By comparing the whole-genome SNPs between controls and cases, which potentially display some particular trait for disease, GWAS could be used to identify the disease or trait-associated alleles. Since the first GWAS paper on age-related macular degeneration in 2005, there have been more than 2000 GWASs published, and data upon more than 15000 SNPs were reported up to 20/2/2015 (<https://www.genome.gov/26525384>).

In contrast to candidate gene directed association study, GWAS is non-hypothesis directed, which avoids the bias and risk when choosing candidate genes. The unified classification of cases and controls with a large sample size in a single study enhances the reliability of the study result. The first GWAS on periodontitis published in 2010 reported a risk locus for AgP in Caucasians (Schaefer et al., 2010). By comparing 283 cases and 972 controls, rs1537415 in the intron region of the GLT6D1 gene was identified as a risk locus (OR=1.59). It was successfully verified by the replication study with another 602 cases and 577 controls. GLT6D1 is named glycosyltransferase 6 domain containing 1, with the function uncharacterized hitherto. This association was replicated in a Sudanese population, and the study reported an odd ratio of 1.5 (Hashim et al., 2015). The polymorphism of rs1537415 may alter the affinity of GATA-3, which is a transcription factor mainly expressed in T helper cells. There is no clear linkage from GLT6D1 to periodontal disease so far, but the risk locus is shared with patients with cardiovascular disease, which may indicate the mutual physiological inflammatory background (Schaefer et al., 2009).

After Schaefer and coworkers, there are six GWAS papers published on periodontitis to date (Divaris et al., 2013; Teumer et al., 2013; Feng et al., 2014; Freitag-Wolf et al., 2014; Shaffer et al., 2014; Shimizu et al., 2015). Four of them were on CP, one on AgP and the other one was not defined. None of them detected any SNPs with significant association, except rs1537415 (Table 1). SNPs with a marginal *p* value were selected in these papers and reported as suggestive genes. Some of the suggestive genes may relate to periodontal inflammation and have been studied in the previous research, such as *LAMA2*, *HAS2*, *HAS2AS*, *CDH2*, *ESR1* and *NPY*. *LAMA2*, the laminin alpha 2, is an extracellular protein that functions as a major component in the basement membrane. Its expression is increased in inflamed periodontal ligaments (PDL) (Han & Amar, 2002). *HAS2* (hyaluronan synthase 2 gene) produces hyaluronan synthase 2, a membrane-bound enzyme that may facilitate tissue repair, promote proliferation of PDL cells and even inhibit pathogens (Rodrigues et al., 2010; Takeda et al., 2011). *CDH2* (gene of Cadherin-2, which is a calcium-dependent cell adhesion glycoprotein) and *ESR1* (estrogen receptor 1 gene) are both reported to be relevant for differentiation of PDL cells (Lin et al., 1999; Pan et al., 2011). *ESR1* is also involved in the regulation of bone modeling (Zhang et al., 2011). Neuropeptide Y (NPY) is highlighted in two of these GWAS papers, on both Caucasian CP and AgP individuals. NPY works as a neurotransmitter in the brain and in the human autonomic nervous system (Colmers & El Bahh, 2003). NPY Y1 receptors were detected in the gingival tissue, and a lower NPY protein level was found in human GCF from periodontitis patients (Lundy et al., 2009). It is abundant in bone and may help maintain homeostasis of the tissue. Concerning T helper type 1/2 cell balance, NPY also enhances the Th2 cell-mediated anti-inflammation response (Freitag-Wolf et al., 2014). Therefore, *NPY* could be considered an important gene related to periodontitis susceptibility and could warrant further investigation. However, there is limited evidence linking the rest of the genes suggested with periodontitis. On the other hand, functional studies are also essential to confirm the pathophysiological contributions of these potential risk-indicating genes.

So far, no significant association was identified from these GWASs except *GLT6D1* for AgP. It is probably because of the insufficient sample sizes of the currently available GWASs, which were not enough for small-effect SNP risk detection regarding the complex human disease of periodontitis. Meta-analysis should be performed by inputting different genotyping platforms, and it may assist in achieving a higher power and eventually a positive result if possible. More international cooperation is required to identify the genes of periodontal disease susceptibility across multiple ethnic groups.

Some limitations of GWAS should be noticed. A recent systematic review compared the cancer-associated genes found by GWAS and candidate gene methods. The study analyzed a database with research since 2000 and found GWAS reported 269 significantly cancer-associated genes, while candidate gene methods reported 349. Only 7.1% of the genes were found significantly associated with cancer in both methods and with similar effect sizes (Chang et al., 2014). Additionally, compared to the estimation by traditional genetic methods, like familial and twin studies, the heritability found by GWAS seems limited. For instance, like human height, the genetic factor is regarded as the main contributor because traditional genetic studies advocated a heritability of 80~90% regarding the phenotype.

However, according to three GWASs, the trait-associated single SNPs contribute only 5% of the height heritability. When such a phenomenon was translated into disease heritability, i.e., causation apparently proven by traditional genetic studies but which cannot be explained by GWAS, a new term “missing heritability” was defined (Maher, 2008). The missing heritability

may be caused by rare variants, structural variations, gene-environmental interactions, parent of origin effects, and errors in narrow-sense heritability estimates and epigenetics, which means the effected molecules may be modified by other factors on a transcriptional or translational level. One of the problems concerning GWAS is that as SNPs are defined as the common variation with the prevalence of more than 1%, it is difficult to detect an uncommon disease caused by rare disease alleles (Zaitlen & Kraft, 2012).

Table 1. Summary of Periodontitis GWAS

Author (year)	Periodontitis type	Initial Sample Size	Replication Sample Size	Reported Gene(s)	SNPs	Context	Risk Allele Frequency	p-Value*	OR	95% CI	Platform [SNPs; passing QC]	Functional study
Schaefer et al. 2010	Aggressive	283 German cases, 972 German controls	Dutch: 602 cases, 577 controls	GLTSD1	rs1537415	intron	0.38	6.0E-09*	1.59	[1.36-1.86]	Affymetrix [345,646]	Fine mapping and annotation
Teumer et al. 2013	Chronic	4032 German	NA	NA	NA	NA	NA	NA	NA	NA	Affymetrix/Illumina [17,533,349] (imputed)	None
Divaris et al. 2013	Chronic (Severe)	4,504 European ancestry Americans	656 European ancestry individuals	NPY	rs2521634		NA	4.0E-07	1.49	[1.28-1.73]	Affymetrix [2,135,236] (imputed)	None
	(Moderate)			NCR2	rs7762544		NR	8.0E-08	1.4	[1.24-1.59]		
	(Moderate)			EMR1, VAV1	rs3826782	intron	NR	8.0E-07	2.01	[1.52-2.65]		
Shaffer et al. 2014	Chronic PD1 (Assuming missing teeth were not affected)	673 non-Hispanic American Caucasians (18-49y)	NA	Gene(s) within 400 kb							Illumina Human610-QuadV1_B [1.4 million]	None
				HSP90AB2P, RAB28, BOD1L, and NKG2-2	rs733048	NA	0.22	1.0E-06	2.4	NA		
				LAMA2 and ARHGAP18	rs10457525	NA	0.8	3.5E-06	2.33	NA		
					rs7749983	NA	0.19	2.4E-06	2.39	NA		
					rs10457526	NA	0.78	6.0E-06	2.26	NA		
					rs7816221	NA	0.32	9.2E-06	2.12	NA		
				HAS2 and HAS2AS	rs3870371	NA	0.32	5.6E-06	2.15	NA		
					rs920455	NA	0.68	9.2E-06	2.11	NA		
				Many genes	rs12799172	NA	0.64	5.1E-06	2.12	NA		
				CDH2	rs11659841	NA	0.13	9.4E-06	2.48	NA		
Feng et al. 2014	Chronic	866 Americans with both African and European ancestry (70% white, 21.9% black, 8.1% others)	1819 Brazilians (81.1% White, 8.9% Black)	uncharacterized non-coding RNA (LOC100506172)	rs1477403	Intergenic	0.46	6.2E-05	1.2	NA	Illumina Human610-QuadV1_B [620,901]	None
Freitag-Wolf et al. 2014	Aggressive	329 German cases and 983 controls	382 Australian/German cases and 489 controls	NPY	rs198712	upstream	Cases: 0.48 in male and 0.29 in female	5.4E-06	2.36	[1.63-3.42]	Affymetrix 500K [2,041 sex related SNPs]	None
Shimizu et al. 2015	Not defined	2,760 cases and 15,158 controls Japanese	1167 cases and 7,178 controls	KCNQ5	rs9446777		0.14	4.8E-06	0.82	[0.74-0.88]	Illumina HumanOmni Express [597,434]	None
				GPR141-NME8	rs2392510		0.41	4.2E-06	0.87	[0.82-0.92]		

2.5. Future Direction

The lessons we learned from these earlier candidate gene studies and GWASs inspired us to consider improving our current study design, appreciate the genetic background of different periodontitis and stimulate exploration for the future research direction.

Case selection should be considered as a key or major issue in future study design. Cases with earlier onset and more severe clinical symptoms may reduce the phenotypic heterogeneity and increase the study power. According to the disease feature and previous genetic studies, we may assume that some of the periodontitis cases, such as some CP cases, may be mainly

influenced by the lifestyle (smoking, diet, stress) or the etiological factors (periodontopathogens), while certain AgP cases are determined by some rare genetic variances, and these rare genetic risk factors have a major effect on pathogenesis (Stabholz et al., 2010). Therefore, issues that have been argued for decades should be revisited again: Is the current disease classification suitable for periodontitis genetic study? Do CP and AgP exhibit fundamentally different genetic backgrounds concerning pathogenicity? In other terms, shall we establish a new classification based on the appreciation regarding different levels of genetic influence? How does that affect our current clinical diagnosis? In some AgP cases, the disease is affected by rare genetic variances, and association studies, either population-based candidate gene study or GWAS, do not have enough power for proper investigation. Family-based study design and large pedigree analysis should be utilized in this AgP category. Distant relatives are supposed to have less common alleles than close relatives. The disease status and the theoretical degree of genetic resemblance can be used to predict the locus of the disease-associated genes using linkage analysis, i.e., the “identity by descent” approach. As the next generation of sequencing develops, we are now capable of sequencing more genes not only limited to the SNPs concerned, with lower cost. Whole exome sequencing and even whole genome sequencing for every suspected locus becomes possible these days. Together with “identity by descent,” such approaches may give researchers a better chance to identify the genes for periodontitis susceptibility, especially for rare AgP. Additionally, because of the pleiotropy of some shared-risk genetic factors between the complex diseases, genes previously identified as susceptible genes for other systemic chronic inflammatory diseases, such as cardiovascular diseases, diabetes or even cancers, could be considered as candidate genes in future periodontitis studies (Vaithilingam et al., 2014).

3. FUNCTIONAL STUDY

Ideally, the allele of disease susceptibility would be located in the protein-coding region, so the disease mechanisms could be easily identified, provided that the function of the underlying molecule(s) of concern could be revealed. However, most of the loci that were found to be associated with periodontitis, similar to that of many other complex common diseases, were located in the non-coding region of the gene(s) concerned, such as the regulatory region. How to link these risk alleles to the pathogenesis and causality of periodontitis remained a major problem after the initial genetic screening/detection.

3.1. Observation at the mRNA and Protein Level

Observation at the mRNA and protein level of concern in the subjects with different genotypes is the initial step of the functional study on periodontitis. The differences in expression level of the genes could be tested directly from samples of saliva, serum, GCF or gingival tissue.

As the first SNPs ever studied in periodontitis, there are many functional studies testing the impact of the IL-1 polymorphisms on the disease concerned. The cytokine levels of IL-1 α , IL-1 β and TNF- α in the GCF of subjects with or without the particular IL-1 α +4845 and IL-1 β

+3953 polymorphisms were soon assessed after the first association study. The concentration of IL-1 β is higher in the patients with the periodontitis-associated genotype, and the difference sustained even after periodontal treatment (Engebretson et al., 1999). Another study reported the patients with T allele on the -889 of *IL1A* have increased IL-1 α level in the GCF by fourfold compared to controls (Shapira et al., 2005). The mRNA and protein levels of other genes were studied as well. Significantly more CP patients and individuals with higher *P. gingivalis* HmuY-induced IL-6 have the G allele at position -174 (Trindade et al., 2013). A study with 122 CP patients and 532 healthy controls assayed the serum levels of MMP-1, MMP-3, MMP-9, IL-2, IL-8 and COX-2 of their subjects using ELISA. Concentrations of all the cytokines followed were higher in the CP patients, but only the -1562 T allele of the MMP-9 gene was found to be associated with higher serum level of MMP-9 compared with CC homozygote (Li et al., 2012). Negative results regarding target genes' expression are not uncommon, even though the gene concern is significantly associated with periodontitis. Although an *IL8* haplotype appeared susceptible to CP [-251(T/A) (rs4073), +396(T/G) (rs2227307) and +781(C/T) (rs2227306)] and IL-8 concentration showed an association with the volume of GCF, *IL8* haplotype was not associated with IL-8 cytokine levels (Corbi et al., 2012). The insignificant results may be caused by many reasons, regardless of the potential type I error that could occur, as mentioned previously in SNP association studies. We can group the disease status, exclude the subjects with some systemic disease, and match the age and genders, but it is impossible to control other factors, which may compensate the effect of the risk loci. The samples are usually collected at one time point from various subjects, and the negative results may be due to epigenetic modification, expression variation by time or other genetic variances which influence the same target.

3.2. Annotation

As negative results on the expression of mRNA or protein are commonly seen in the samples directly collected from subjects with different genotypes, further annotation of the risk loci is required to reveal the function of the genetic polymorphisms *in vitro*.

Annotation means analysis and interpretation of the DNA sequence for the consequent biological significance and process. It consists of mainly two parts. One is structural annotation, which is the basic annotation to recognize the protein-coding gene, RNA gene and other functional elements, like regulatory elements concerning the region or gene of susceptibility. By comparing the sequence similarities with the known functional sequences, these elements, including exons and introns, could be identified manually or computationally by some sophisticated software algorithms. The second part of annotation is the functional annotation, which explains the biological function of the gene or the corresponding protein, the regulation mechanisms and the interaction with other molecules.

As the risk alleles are usually found in the regulatory region of the genes, it is essential to characterize the function of these loci. The gene expression can be regulated in many stages, like transcription initiation, elongation and mRNA processing, transport, translation and stability (Maston et al., 2006). Most of the regulation of the transcription occurred in the initiation stage. The genetic polymorphisms could be found in the region of any regulatory elements, such as enhancers, promoters, insulators and silencers.

Both *de novo* computational analysis and experimental laboratory work contribute to the structural and functional annotation of the periodontitis-associated genes. The putative protein product of the expression could be predicted *in silico* by searching for proteins with similar sequences.

3.2.1. Computational Annotation

By integrating the knowledge of these regulatory elements and the risk-associated SNPs, software algorithms and online databases based on the literature or studies of other species can be used to study the influence of the risk alleles identified, such as transcription factor binding and activity. For instance, in the first AgP GWAS, the G allele of rs1537415 in the regulatory region of *GLT6D1* was identified as a risk allele associated with AgP.

More than five hundred binding models of vertebrate transcription factors were obtained from the TRANSFAC database in order to find the transcription factor, which is most likely to be affected by this SNP. The TRANSFAC database focuses on eukaryotic transcriptional regulation. It contains data on transcription factors, the corresponding target genes and regulatory binding sites.

The transcription factor binding affinities were evaluated by the TRAP method instead of experimental binding data, which is a biophysical model based on the genome-wide binding data from yeast (Schaefer et al., 2010). The C to G exchange of rs1537415 reduced the predicted binding affinity of GATA-3 significantly, which was later verified experimentally by electrophoretic mobility shift analysis. Computational annotation can indicate the potential function of the genetic variance effectively and efficiently, but experimental verification is required afterward.

After a set of genes were identified by previous genetic studies like GWAS, pathway analysis could be carried out to map the genes concerned with certain pathways and therefore reveal the pathogenesis of the disease. In one of the GWASs on CP, although it did not find significantly disease-associated SNPs, it reported eleven ingenuity pathway analysis canonical signaling pathways were significantly enriched in the moderate and severe CP cases.

Most of these pathways are related to neurotransmitters and nervous signaling pathways, which indicate the potential pathogenesis of CP and a new research direction in the future (Rhodin et al., 2014).

3.2.2. Experimental Annotation

Experimentally testing on cell or tissue models can exclude confounding factors and could become a fundamental part of functional study in the post-SNPs era. By establishing a cell or tissue model with the wild/mutated genotype by bimolecular methods, studies that control the other variants could focus on investigating the underlying pathogenic effect of the target SNPs. The function of rs1800972 in the 5' untranslated region of the human β -defensin 1 gene, an antimicrobial peptide that may be related to periodontitis, was investigated in 2009. The study constructed a cell model with the transfected chloramphenicol acetyltransferase (CAT) reporter gene and the 5' untranslated region and found that the gene with G allele produced much more CAT protein compared to other haplotypes. The SNP concerned was indicated as a cis regulatory element effecting transcription or translation of Human β -defensin 1 (Kalus et al., 2009). Another study in 2012 reported that the G allele of rs11536889 in 3'-UTR of TLR4 suppressed the luciferase activity induced by LPS or IL-6 (Sato et al., 2012). However, there

are very limited studies available regarding either computational or experimental annotation on risk SNPs relevant to periodontitis.

3.3. Epigenetic Regulation of Gene Expression and Gene-Environment Interactions

The epigenetic process chemically modifies the DNA and the associated proteins, so the genes will be selectively activated or inhibited and the expression levels will be affected as well. DNA methylation and histone deacetylation are two of the most extensively investigated epigenetic alterations. Methylation adds a methyl group to adenine or cytosine DNA nucleotides within a CpG dinucleotide context typically, while histone deacetylase removes an acetyl group from the histone and leads to a tightly histone-wrapped DNA. Both of them block the binding of transcription factors and are associated with transcription silencing. There are studies on these two epigenetic modifications on periodontitis, but limited studies elucidate the effect of the epigenetic modifications upon periodontitis-associated SNPs.

The impact of DNA methylation and histone modifications was studied on IL-10, via comparing the IL-10 gene expression in subjects with different genotypes of IL-10 -1087. In the B cells stimulated with LPS, the increase of the DNA methylation of histone H3 and acetylation of histone H4 of the genotype GG cells appeared higher than that of the AA genotype cells. In the unstimulated cells, a larger increase of the acetylation and methylation of histone H3 was detected in GG than AA genotype cells, while acetylation of histone H4 appeared higher in AA genotype cells (Larsson et al., 2012).

Gene-environment interactions and epigenetics should also be considered in the future. Gene-environment interaction in chronic inflammatory disease includes intrinsic and extrinsic mechanisms (Renz et al., 2011). Regarding intrinsic mechanisms, epigenetic variables link the disease phenotype and environment with the human genome in the disease. Environmental triggers like smoking may lead to epigenetic changes. The interaction between Vitamin D receptor gene polymorphism and smoking presented some evidence of association upon the severe periodontitis progression (Nibali et al., 2008a). Significant additive interaction shows the effect of VDR risk allele and smoking upon periodontitis appeared more than the effect of the sum of these two risk factors working separately (Tanaka et al., 2013). Concerning extrinsic mechanisms, periodontopathogens, for instance, could be regarded as a main factor. As oral microbiota is the cause for the inflammation in gingivitis or periodontitis, the prevalence of some key oral pathogens like *Aggregatibacter actinomycetemcomitans*, *P. gingivalis* and *Tannerella forsythensis* was found to be associated with IL-6 and Fcγ receptor gene polymorphisms (Nibali et al., 2007 & 2008b). The mutual interaction between host genetic factors and the pathogens has not been thoroughly investigated. With the development of next-generation sequencing, the relationship between human genome-wide information and the pathogen profile could also be a new study direction. A GWAS published in 2012 concerning 1020 Caucasians reported a negative association between the human whole genome SNPs and eight periodontopathogens detectable by DNA-DNA hybridization (Divaris et al., 2012). A larger sample size and perhaps more pathogens should be investigated in future studies.

CONCLUSION AND FUTURE DIRECTION

It has been proven that SNPs on various genes, such as IL-1, IL-6 and MMP-9 genes, confer influence regarding the corresponding gene's expression. Individuals with various genotypes associated with the corresponding mRNA and protein levels confirmed the potential effects of many SNPs on periodontitis. Whole exome sequencing and even whole genome sequencing become possible to detect rare variants with a relatively large effect. In order to elaborate particular periodontal pathogenic mechanisms, more computational or experimental annotations and epigenetic investigations should be performed in the future. Cell, tissue or even animal models with the normal/risk genotypes should be established to test the function and pathophysiology of the periodontitis-associated SNPs in order to decipher the mechanism of the disease development with all other factors controlled. Epigenetic modifications on the genes of interest by environmental factors, e.g., by a particular periodontopathogen or smoking, should also be characterized. Periodontitis is a dynamic and complex disease whose etiology cannot be explained or predicted by any known risk indicators. Inspired by the concept of "systems biology," multiple etiological and/or risk factors of the complex disease periodontitis could be integrated into a network and be regarded as one system with emergent properties, based on the complicated interaction among the factors included. By establishing a new model with more comprehensive information on the genomic, transcriptomic, proteomic and metabolomic levels, the environment and their interactions, people may stand a better chance of understanding periodontitis in the future.

ACKNOWLEDGMENTS

The work described in this chapter was partially supported by grants from the Health and Medical Research Fund (02132216), the Research Grants Council of the Hong Kong Special Administrative Region, China (HKU 772110M), the University of Hong Kong Small Project Fund (200607176088) and the Seed Funding Programme for Basic Research (200411159024).

REFERENCES

- Albuquerque, C. M., Cortinhas, A. J., Morinha, F. J., Leitao, J. C., Viegas, C. A. & Bastos, E. M. (2012). Association of the IL-10 polymorphisms and periodontitis: a meta-analysis. *Mol. Biol. Rep.* 39, 9319-29.
- Armitage, G. C. (1999). Development of a classification system for periodontal diseases and conditions. *Ann. Periodontol.* 4, 1-6.
- Armitage, G. C. (2004). Periodontal diagnoses and classification of periodontal diseases. *Periodontol.* 2000, 34, 9-21.
- Armitage, G. C., Wu, Y., Wang, H. Y., Sorrell, J., di Giovine, F. S. & Duff, G. W. (2000). Low prevalence of a periodontitis-associated interleukin-1 composite genotype in individuals of Chinese heritage. *J. Periodontol.* 71, 164-71.

- Astolfi, C. M., Shinohara, A. L., da Silva, R. A., Santos, M. C. L. G., Line, S. R. P. & de Souza, A. P. (2006). Genetic polymorphisms in the MMP-1 and MMP-3 gene may contribute to chronic periodontitis in a Brazilian population. *J. Clin. Periodontol.* 33, 699-703.
- Beaty, T. H., Boughman, J. A., Yang, P., Astemborski, J. A. & Suzuki, J. B. (1987). Genetic-analysis of juvenile periodontitis in families ascertained through an affected proband. *Am. J. Hum. Genet.* 40, 443-52.
- Benjamin, S. D. & Baer, P. N. (1967). Familial patterns of advanced alveolar bone loss in adolescence (periodontosis). *Periodontics* 5, 82-88.
- Boughman, J. A., Astemborski, J. A. & Suzuki, J. B. (1992). Phenotypic assessment of early onset periodontitis in sibships. *J. Clin. Periodontol.* 19, 233-9.
- Boughman, J. A., Halloran, S. L., Roulston, D., Schwartz, S., Suzuki, J. B., Weitkamp, L. R., et al. (1986). An autosomal-dominant form of juvenile periodontitis: its localization to chromosome 4 and linkage to dentinogenesis imperfecta and Gc. *J. Craniofac. Genet. Dev. Biol.* 6, 341-50.
- Braosi, A. P. R., de Souza, C. M., Luczynszyn, S. M., Dirschnabel, A. J., Claudino, M., Olandoski, M., et al. (2012). Analysis of IL-1 gene polymorphisms and transcript levels in periodontal and chronic kidney disease. *Cytokine* 60, 76-82.
- Butler, J. H. (1969). A familial pattern of juvenile periodontitis (periodontosis). *J. Periodontol.* 40, 115-8.
- Cardon, L. R. & Bell, J. I. (2001). Association study designs for complex diseases. *Nat. Rev. Genet.* 2, 91-9.
- Chai, L., Song, Y. Q. & Leung, W. K. (2012). Genetic polymorphism studies in periodontitis and Fcgamma receptors. *J. Periodontal. Res.* 47, 273-85.
- Chai, L., Song, Y. Q., Zee, K. Y. & Leung, W. K. (2010). SNPs of Fc-gamma receptor genes and chronic periodontitis. *J. Dent. Res.* 89, 705-10.
- Chang, C. Q., Yesupriya, A., Rowell, J. L., Pimentel, C. B., Clyne, M., Gwinn, M., et al. (2014). A systematic review of cancer GWAS and candidate gene meta-analyses reveals limited overlap but similar effect sizes. *Eur. J. Hum Genet* 22, 402-8.
- Chen, D., Wang, Q., Ma, Z. W., Chen, F. M., Chen, Y., Xie, G. Y., et al. (2007). MMP-2, MMP-9 and TIMP-2 gene polymorphisms in Chinese patients with generalized aggressive periodontitis. *J. Clin. Periodontol.* 34, 384-9.
- Chen, L. L., Li, H., Zhang, P. P. & Wang, S. M. (2012). Association between vitamin D receptor polymorphisms and periodontitis: A meta-analysis. *J. Periodontol.* 83, 1095-103.
- Colmers, W. F. & El Bahh, B. (2003). Neuropeptide Y and epilepsy. *Epilepsy Curr.* 3, 53-8.
- Corbi, S. C., Anovazzi, G., Finoti, L. S., Kim, Y. J., Capela, M. V., Secolin, R., et al. (2012). Haplotypes of susceptibility to chronic periodontitis in the interleukin 8 gene do not influence protein level in the gingival crevicular fluid. *Arch. Oral. Biol.* 57, 1355-61.
- Corey, L. A., Nance, W. E., Hofstede, P. & Schenkein, H. A. (1993). Self-reported periodontal disease in a Virginia twin population. *J. Periodontol.* 64, 1205-8.
- Deng, H., Liu, F., Pan, Y. H., Jin, X. Q., Wang, H. N. & Cao, J. F. (2011). BsmI, TaqI, ApaI, and FokI polymorphisms in the vitamin D receptor gene and periodontitis: a meta-analysis of 15 studies including 1338 cases and 1302 controls. *J. Clin. Periodontol.* 38, 199-207.
- Dimou, N. L., Nikolopoulos, G. K., Hamodrakas, S. J. & Bagos, P. G. (2010). Fc gamma receptor polymorphisms and their association with periodontal disease: a meta-analysis. *J. Clin. Periodontol.* 37, 255-65.

- Ding, C., Ji, X., Chen, X., Xu, Y. & Zhong, L. (2014). TNF-alpha gene promoter polymorphisms contribute to periodontitis susceptibility: evidence from 46 studies. *J. Clin. Periodontol.* 41, 748-59.
- Divaris, K., Monda, K. L., North, K. E., Olshan, A. F., Lange, E. M., Moss, K., et al. (2012). Genome-wide association study of periodontal pathogen colonization. *J. Dent. Res.* 91, 21S-8S.
- Divaris, K., Monda, K. L., North, K. E., Olshan, A. F., Reynolds, L. M., Hsueh, W. C., et al. (2013a). Exploring the genetic basis of chronic periodontitis: a genome-wide association study. *Hum. Mol. Genet.* 22, 2312-24.
- Eke, P. I., Dye, B. A., Wei, L., Thornton-Evans, G. O., Genco, R. J., & CDC Periodontal Disease Surveillance workgroup. (2012). Prevalence of periodontitis in adults in the United States: 2009 and 2010. *J. Dent. Res.* 91, 914-20.
- Engelbreton, S. P., Lamster, I. B., Herrera-Abreu, M., Celenti, R. S., Timms, J. M., Chaudhary, A. G., et al. (1999). The influence of interleukin gene polymorphism on expression of interleukin-1beta and tumor necrosis factor-alpha in periodontal tissue and gingival crevicular fluid. *J. Periodontol.* 70, 567-73.
- Eres, G., Saribay, A. & Akkaya, M. (2009). Periodontal treatment needs and prevalence of localized aggressive periodontitis in a young Turkish population. *J. Periodontol.* 80, 940-4.
- Feng, P., Wang, X., Casado, P. L., Kuchler, E. C., Deeley, K., Noel, J., et al. (2014). Genome wide association scan for chronic periodontitis implicates novel locus. *BMC Oral Health* 14, 84.
- Fourel, J. (1972). Periodontosis: a periodontal syndrome. *J. Periodontol.* 43, 240-55.
- Freitag-Wolf, S., Dommisch, H., Graetz, C., Jockel-Schneider, Y., Harks, I., Staufenbiel, I., et al. (2014). Genome-wide exploration identifies sex-specific genetic effects of alleles upstream NPY to increase the risk of severe periodontitis in men. *J. Clin. Periodontol.* 41, 1115-21.
- Fukuda, K., Fujitsu, Y., Kumagai, N. & Nishida, T. (2006). Inhibition of matrix metalloproteinase-3 synthesis in human conjunctival fibroblasts by interleukin-4 or interleukin-13. *Invest. Ophthalmol. Vis. Sci.* 47, 2857-64.
- Graves, D. T. & Cochran, D. (2003). The contribution of interleukin-1 and tumor necrosis factor to periodontal tissue destruction. *J. Periodontol.* 74, 391-401.
- Grossi, S. G., Zambon, J. J., Ho, A. W., Koch, G., Dunford, R. G., Machtei, E. E., et al. (1994). Assessment of risk for periodontal-disease. 1. Risk indicators for attachment loss. *J. Periodontol.* 65, 260-7.
- Han, X. & Amar, S. (2002). Identification of genes differentially expressed in cultured human periodontal ligament fibroblasts vs. human gingival fibroblasts by DNA microarray analysis. *J. Dent. Res.* 81, 399-405.
- Hart, T. C., Marazita, M. L., McCanna, K. M., Schenkein, H. A. & Diehl, S. R. (1993). Reevaluation of the chromosome 4q candidate region for early onset periodontitis. *Hum. Genet.* 91, 416-22.
- Hashim, N. T., Linden, G. J., Ibrahim, M. E., Gismalla, B. G., Lundy, F. T., Hughes, F. J., et al. (2015). Replication of the association of *GLT6D1* with aggressive periodontitis in a Sudanese population. *J. Clin. Periodontol.* 42, 319-24.

- Hewitt, C., McCormick, D., Linden, G., Turk, D., Stern, I., Wallace, I., et al. (2004). The role of cathepsin C in Papillon-Lefevre syndrome, prepubertal periodontitis, and aggressive periodontitis. *Hum. Mutat.* 23, 222-8.
- Hodge, S. E. (1994). What association analysis can and cannot tell us about the genetics of complex disease. *Am. J. Med. Genet.* 54, 318-23.
- Hoffman, I. D. (1983). Familial occurrence of juvenile periodontitis with varied treatment of one of the siblings with five-year follow-up. Case report. *J. Periodontol.* 54, 44-9.
- Hugoson, A. & Norderyd, O. (2008). Has the prevalence of periodontitis changed during the last 30 years? *J. Clin. Periodontol.* 35, 338-45.
- Ingman, T., Tervahartiala, T., Ding, Y. L., Tschesche, H., Haerian, A., Kinane, D. F., et al. (1996). Matrix metalloproteinases and their inhibitors in gingival crevicular fluid and saliva of periodontitis patients. *J. Clin. Periodontol.* 23, 1127-32.
- Itagaki, M., Kubota, T., Tai, H., Shimada, Y., Morozumi, T. & Yamazaki, K. (2004). Matrix metalloproteinase-1 and-3 gene promoter polymorphisms in Japanese patients with periodontitis. *J. Clin. Periodontol.* 31, 764-9.
- Jabbar, S., Drury, J., Fordham, J., Datta, H. K., Francis, R. M. & Tuck, S. P. (2011). Plasma vitamin D and cytokines in periodontal disease and postmenopausal osteoporosis. *J. Periodontol. Res.* 46, 97-104.
- Jorgenson, R. J., Levin, L. S., Hutcherson, S. T. & Salinas, C. F. (1975). Periodontosis in sibs. *Oral Surg. Oral Med. Oral Pathol.* 39, 396-402.
- Kabashima, H., Nagata, K., Hashiguchi, I., Toriya, Y., Iijima, T., Maki, K., et al. (1996). Interleukin-1 receptor antagonist and interleukin-4 in gingival crevicular fluid of patients with inflammatory periodontal disease. *J. Oral Pathol. Med.* 25, 449-55.
- Kalus, A. A., Fredericks, L. P., Hacker, B. M., Dommisch, H., Presland, R. B., Kimball, J. R., et al. (2009). Association of a genetic polymorphism (-44 C/G SNP) in the human *DEFB1* gene with expression and inducibility of multiple beta-defensins in gingival keratinocytes. *BMC Oral Health* 9, 21.
- Karimbux, N. Y., Saraiya, V. M., Elangovan, S., Allareddy, V., Kinnunen, T., Kornman, K. S., et al. (2012). Interleukin-1 gene polymorphisms and chronic periodontitis in adult whites: a systematic review and meta-analysis. *J. Periodontol.* 83, 1407-19.
- Kaupilla, L. I. (2009). Atherosclerosis and disc degeneration/low-back pain - A systematic review. *Eur. J. Vas. Endovasc. Surg.* 37, 661-70.
- Khocht, A. & Albandar, J. M. (2014). Aggressive forms of periodontitis secondary to systemic disorders. *Periodontol* 2000 65, 134-48.
- Kinane, D. F., Demuth, D. R., Gorr, S. U., Hajishengallis, G. N. & Martin, M. H. (2007). Human variability in innate immunity. *Periodontol.* 2000 45, 14-34.
- Kinane, D. F. & Hart, T. C. (2003). Genes and gene polymorphisms associated with periodontal disease. *Crit. Rev. Oral Biol. Med.* 14, 430-49.
- Kobayashi, K., Takahashi, N., Jimi, E., Udagawa, N., Takami, M., Kotake, S., et al. (2000). Tumor necrosis factor alpha stimulates osteoclast differentiation by a mechanism independent of the ODF/RANKL-RANK interaction. *J. Exp. Med.* 191, 275-86.
- Kobayashi, T., Murasawa, A., Ito, S., Yamamoto, K., Komatsu, Y., Abe, A., et al. (2009). Cytokine gene polymorphisms associated with rheumatoid arthritis and periodontitis in Japanese adults. *J. Periodontol.* 80, 792-9.

- Kornman, K. S., Crane, A., Wang, H. Y., diGiovine, F. S., Newman, M. G., Pirk, F. W., et al. (1997). The interleukin-1 genotype as a severity factor in adult periodontal disease. *J. Clin. Periodontol.* 24, 72-7.
- Laine, M. L., Crielaard, W. & Loos, B. G. (2012). Genetic susceptibility to periodontitis. *Periodontol.* 2000, 58, 37-68.
- Larsson, L., Thorbert-Mros, S., Rymo, L. & Berglundh, T. (2012). Influence of epigenetic modifications of the interleukin-10 promoter on *IL10* gene expression. *Eur. J. Oral Sci.* 120, 14-20.
- Lee, A., Ghaname, C. B., Braun, T. M., Sugai, J. V., Teles, R. P., Loesche, W. J., et al. (2012). Bacterial and salivary biomarkers predict the gingival inflammatory profile. *J. Periodontol.* 83, 79-89.
- Letra, A., Silva, R. M., Rylands, R. J., Silveira, E. M., de Souza, A. P., Wendell, S. K., et al. (2012). MMP3 and TIMP1 variants contribute to chronic periodontitis and may be implicated in disease progression. *J. Clin. Periodontol.* 39, 707-16.
- Li, D. D., Cai, Q., Ma, L., Wang, M. L., Ma, J. Q., Zhang, W. B., et al. (2013). Association between MMP-1 g.-1607dupG polymorphism and periodontitis susceptibility: A meta-analysis. *PLoS One* 8:e59513.
- Li, G. Y., Yue, Y., Tian, Y., Li, J. L., Wang, M., Liang, H., et al. (2012). Association of matrix metalloproteinase (MMP)-1, 3, 9, interleukin (IL)-2, 8 and cyclooxygenase (COX)-2 gene polymorphisms with chronic periodontitis in a Chinese population. *Cytokine* 60, 552-60.
- Li Y, Xu L, Hasturk H, Kantarci A, DePalma SR, Van Dyke TE. (2004). Localized aggressive periodontitis is linked to human chromosome 1q25. *Hum Genet* 114, 291-7.
- Lin, W. L., Chien, H. H. & Cho, M. I. (1999). N-cadherin expression during periodontal ligament cell differentiation *in vitro*. *J. Periodontol.* 70, 1039-45.
- Loe, H., Anerud, A., Boysen, H. & Morrison, E. (1986). Natural history of periodontal disease in man- rapid, moderate and no loss of attachment in Sri-Lankan laborers 14 to 46 years of age. *J. Clin. Periodontol.* 13, 431-40.
- Loo, W. T. Y., Fan, C. B., Bai, L. J., Yue, Y., Dou, Y. D., Wang, M., et al. (2012). Gene polymorphism and protein of human pro- and anti-inflammatory cytokines in Chinese healthy subjects and chronic periodontitis patients. *J. TranslMed* 10(Suppl 1): S8.
- Loos, B. G., John, R. P. & Laine, M. L. (2005). Identification of genetic risk factors for periodontitis and possible mechanisms of action. *J. Clin. Periodontol.* 32(Suppl 6), 159-79.
- Lopez, N. J. (1992). Clinical, laboratory, and immunological studies of a family with a high prevalence of generalized prepubertal and juvenile periodontitis. *J. Periodontol.* 63, 457-68.
- Lundy, F. T., El Karim, I. A. & Linden, G. J. (2009). Neuropeptide Y (NPY) and NPY Y1 receptor in periodontal health and disease. *Arch. Oral Biol.* 54, 258-62.
- Maher, B. (2008). Personal genomes: The case of the missing heritability. *Nature* 456(7218), 18-21.
- Mao, M., Zeng, X. T., Ma, T., He, W., Zhang, C. & Zhou, J. (2013). Interleukin-1alpha -899 (+4845) C-->T polymorphism increases the risk of chronic periodontitis: evidence from a meta-analysis of 23 case-control studies. *Gene* 532, 114-9.
- Marazita, M. L., Burmeister, J. A., Gunsolley, J. C., Koertge, T. E., Lake, K. & Schenkein, H. A. (1994). Evidence for autosomal dominant inheritance and race-specific heterogeneity in early-onset periodontitis. *J. Periodontol.* 65, 623-30.

- Maston, G. A., Evans, S. K. & Green, M. R. (2006). Transcriptional regulatory elements in the human genome. *Annu Rev Genomics Hum. Genet.* 7, 29-59.
- Melnick, M., Shields, E. D. & Bixler, D. (1976). Periodontosis: a phenotypic and genetic analysis. *Oral Surg. Oral Med. Oral Pathol.* 42, 32-41.
- Michalowicz, B. S., Aeppli, D., Virag, J. G., Klump, D. G., Hinrichs, J. E., Segal, N. L., et al. (1991). Periodontal findings in adult twins. *J. Periodontol* 62, 293-9.
- Ng, S. K. S. & Leung, W. K. (2006). A community study on the relationship between stress, coping, affective dispositions and periodontal attachment loss. *Community Dent. Oral Epidemiol.* 34, 252-66.
- Nibali, L., Parkar, M., D'Aiuto, F., Suvan, J. E., Brett, P. M., Griffiths, G. S., et al. (2008a). Vitamin D receptor polymorphism (-1056 Taq-I) interacts with smoking for the presence and progression of periodontitis. *J. Clin. Periodontol* 35, 561-7.
- Nibali, L., Ready, D. R., Parkar, M., Brett, P. M., Wilson, M., Tonetti, M. S., et al. (2007). Gene polymorphisms and the prevalence of key periodontal pathogens. *J. Dent Res.* 86, 416-20.
- Nibali, L., Tonetti, M. S., Ready, D., Parkar, M., Brett, P. M., Donos, N., et al. (2008b). Interleukin-6 polymorphisms are associated with pathogenic bacteria in subjects with periodontitis. *J. Periodontol.* 79, 677-83.
- Nikolopoulos, G. K., Dimou, N. L., Hamodrakas, S. J. & Bagos, P. G. (2008). Cytokine gene polymorphisms in periodontal disease: a meta-analysis of 53 studies including 4178 cases and 4590 controls. *J. Clin. Periodontol.* 35, 754-67.
- Noack, B., Gorgens, H., Lorenz, K., Ziegler, A., Hoffmann, T. & Schackert, H. K. (2008). TLR4 and IL-18 gene variants in aggressive periodontitis. *J. Clin. Periodontol.* 35, 1020-6.
- Ozturk, A. & Vieira, A. R. (2009). TLR4 as a risk factor for periodontal disease: a reappraisal. *J. Clin. Periodontol.* 36, 279-86.
- Page, R. C., Vandesteen, G. E., Ebersole, J. L., Williams, B. L., Dixon, I. L. & Altman, L. C. (1985). Clinical and laboratory studies of a family with a high prevalence of juvenile periodontitis. *J. Periodontol.* 56, 602-10.
- Pan, F., Zhang, R., Wang, G. & Ding, Y. (2011). Oestrogen receptors are involved in the osteogenic differentiation of periodontal ligament stem cells. *Biosci. Rep.* 31, 117-24.
- Petersen, P. E. (2003). The World Oral Health Report 2003: continuous improvement of oral health in the 21st century - the approach of the WHO Global Oral Health Programme. *Community Dent. Oral Epidemiol.* 31, 3-23.
- Petit, M. D. A., Vansteenberg, T. J. M., Timmerman, M. F., Degraaff, J. & Vandervelden, U. (1994). Prevalence of periodontitis and suspected periodontal pathogens in families of adult periodontitis patients. *J. Clin. Periodontol.* 21, 76-85.
- Renz, H., von Mutius, E., Brandtzaeg, P., Cookson, W. O., Autenrieth, I. B. & Haller, D. (2011). Gene-environment interactions in chronic inflammatory disease. *Nat. Immunol.* 12, 273-7.
- Rhodin, K., Divaris, K., North, K. E., Barros, S. P., Moss, K., Beck, J. D., et al. (2014). Chronic periodontitis genome-wide association studies: Gene-centric and gene set enrichment analyses. *J. Dent. Res.* 93, 882-90.
- Rodrigues, S. V., Acharya, A. B., Bhadbhade, S. & Thakur, S. L. (2010). Hyaluronan-containing mouthwash as an adjunctive plaque-control agent. *Oral Health Prev. Dent.* 8, 389-94.

- Ryder, M. I. (2010). Comparison of neutrophil functions in aggressive and chronic periodontitis. *Periodontol.* 2000, 53, 124-37.
- Sato, K., Yoshimura, A., Kaneko, T., Ukai, T., Ozaki, Y., Nakamura, H., et al. (2012). A single nucleotide polymorphism in 3'-untranslated region contributes to the regulation of Toll-like receptor 4 translation. *J. Biol. Chem* 287, 25163-72.
- Saxen, L. (1980). Heredity of juvenile periodontitis. *J. Clin. Periodontol.* 7, 276-88.
- Schaefer, A. S., Richter, G. M., Groessner-Schreiber, B., Noack, B., Nothnagel, M., El Mokhtari, N. E., et al. (2009). Identification of a shared genetic susceptibility locus for coronary heart disease and periodontitis. *PLoS Genet* 5, e1000378.
- Schaefer, A. S., Richter, G. M., Nothnagel, M., Manke, T., Dommisch, H., Jacobs, G., et al. (2010). A genome-wide association study identifies *GLT6D1* as a susceptibility locus for periodontitis. *Hum. Mol. Genet.* 19, 553-62.
- Schäfer, A. S., Jepsen, S., & Loos, B. G. (2011). Periodontal genetics: a decade of genetic association studies mandates better study designs. *J. Clin. Periodontol.* 38, 103-7.
- Shaffer, J. R., Polk, D. E., Wang, X., Feingold, E., Weeks, D. E., Lee, M. K., et al. (2014). *Genome-wide association study of periodontal health measured by probing depth in adults ages 18-49 years.* *G3* (Bethesda), 4, 307-14.
- Shapira, L., Vandyke, T. E. & Hart, T. C. (1992). A localized absence of interleukin-4 triggers periodontal disease activity: a novel hypothesis. *Med. Hypotheses* 39, 319-22.
- Shapira, L., Wilensky, A. & Kinane, D. F. (2005). Effect of genetic variability on the inflammatory response to periodontal infection. *J. Clin. Periodontol.* 32(Suppl 6), 72-86.
- Shimizu, S., Momozawa, Y., Takahashi, A., Nagasawa, T., Ashikawa, K., Terada, Y., et al. (2015). A Genome-wide association study of periodontitis in a Japanese population. *J. Dent. Res.* 94, 555-61.
- Song, G. G., Choi, S. J., Ji, J. D. & Lee, Y. H. (2013). Association between tumor necrosis factor- α promoter -308 A/G, -238 A/G, interleukin-6 -174 G/C and -572 G/C polymorphisms and periodontal disease: a meta-analysis. *Mol. Biol. Rep.* 40, 5191-203.
- Song, G. G., Kim, J. H. & Lee, Y. H. (2013). Toll-like receptor (TLR) and matrix metalloproteinase (MMP) polymorphisms and periodontitis susceptibility: a meta-analysis. *Mol. Biol. Rep.*, 40, 5129-41.
- Sorsa, T., Tjaderhane, L., Kontinen, Y. T., Lauhio, A., Salo, T., Lee, H. M., et al. (2006). Matrix metalloproteinases: Contribution to pathogenesis, diagnosis and treatment of periodontal inflammation. *Ann. Med.* 38, 306-21.
- Stabholz, A., Soskolne, W. A. & Shapira, L. (2010). Genetic and environmental risk factors for chronic periodontitis and aggressive periodontitis. *Periodontol.* 2000, 53, 138-53.
- Tabor, H. K., Risch, N. J. & Myers, R. M. (2002). Candidate-gene approaches for studying complex genetic traits: practical considerations. *Nat. Rev. Genet* 3, 391-7.
- Takeda, K., Sakai, N., Shiba, H., Nagahara, T., Fujita, T., Kajiya, M., et al. (2011). Characteristics of high-molecular-weight hyaluronic acid as a brain-derived neurotrophic factor scaffold in periodontal tissue regeneration. *Tissue Eng. Part A* 17(7-8), 955-67.
- Tanaka, K., Miyake, Y., Hanioka, T. & Arakawa, M. (2013). VDR gene polymorphisms, interaction with smoking and risk of periodontal disease in Japanese women: the Kyushu Okinawa maternal and child health study. *Scand. J. Immunol.* 78, 371-7.
- Teumer, A., Holtfreter, B., Volker, U., Petersmann, A., Nauck, M., Biffar, R., et al. (2013). Genome-wide association study of chronic periodontitis in a general German population. *J. Clin. Periodontol.* 40, 977-85.

- Trindade, S. C., Olczak, T., Gomes, I. S., de Moura-Costa, L. F., Vale, V. C., Galdino-Neto, M., et al. (2013). *Porphyromonas gingivalis* HmuY-induced production of interleukin-6 and IL-6 polymorphism in chronic periodontitis. *J. Periodontol.* 84, 650-5.
- Tsai, C. C., Ku, C. H., Ho, Y. P., Ho, K. Y., Wu, Y. M. & Hung, C. C. (2007). Changes in gingival crevicular fluid interleukin-4 and interferon-gamma in patients with chronic periodontitis before and after periodontal initial therapy *Kaohsiung J. Med. Sci.* 23, 1-7.
- Uitterlinden, A. G., Ralston, S. H., Brandi, M. L., Carey, A. H., Grinberg, D. & Langdahl, B. L. (2006). The association between common vitamin D receptor gene variations and osteoporosis: a participant-level meta-analysis. *Ann. Intern. Med.* 145, 255-64; erratum in 145, 936.
- Vaithilingam, R. D., Safii, S. H., Baharuddin, N. A., Ng, C. C., Cheong, S. C., Bartold, P. M., et al. (2014). Moving into a new era of periodontal genetic studies: relevance of large case-control samples using severe phenotypes for genome-wide association studies. *J. Periodontol. Res.* 49, 683-95.
- van Etten, E., Verlinden, L., Giulietti, A., Ramos-Lopez, E., Branisteanu, D. D., Ferreira, G. B., et al. (2007). The vitamin D receptor gene *FokI* polymorphism: Functional impact on the immune system. *Eur. J. Immunol.* 37, 395-405.
- Vandervelden, U., Abbas, F., Armand, S., Degraaff, J., Timmerman, M. F., Vanderweijden, G. A., et al. (1993). The effect of sibling relationship on the periodontal condition. *J. Clin. Periodontol.* 20, 683-90.
- Wu, X., Offenbacher, S., Lomicronpez, N. J., Chen, D., Wang, H. Y., Rogus, J., et al. (2015). Association of interleukin-1 gene variations with moderate to severe chronic periodontitis in multiple ethnicities. *J. Periodontol. Res.*, 50, 52-61.
- Wu, Y. M., Chuang, H. L., Ho, Y. P., Ho, K. Y., & Tsai, C. C. (2010). Investigation of interleukin-13 gene polymorphisms in individuals with chronic and generalized aggressive periodontitis in a Taiwanese (Chinese) population. *J. Periodontol. Res.*, 45, 695-701.
- Wynn, T. A. (2003). IL-13 effector functions. *Ann Rev Immunol* 21, 425-56.
- Yamamoto, M., Kawabata, K., Fujihashi, K., McGhee, J. R., VanDyke, T. E., Bamberg, T. V., et al. (1996). Absence of exogenous interleukin-4-induced apoptosis of gingival macrophages may contribute to chronic inflammation in periodontal diseases. *Am. J. Pathol.* 148, 331-9.
- Yamazaki, K., Nakajima, T., Kubota, Y., Gemmell, E., Seymour, G. J. & Hara, K. (1997). Cytokine messenger RNA expression in chronic inflammatory periodontal disease. *Oral Microbiol. Immunol.* 12, 281-7.
- Yan, Y., Weng, H., Shen, Z. H., Wu, L. & Zeng, X. T. (2014). Association between interleukin-4 gene -590 c/t, -33 c/t, and 70-base-pair polymorphisms and periodontitis susceptibility: a meta-analysis. *J. Periodontol* 85, e354-62.
- Yoshie, H., Kobayashi, T., Tai, H. & Galicia, J. C. (2007). The role of genetic polymorphisms in periodontitis. *Periodontol 2000* 43, 102-32.
- Yoshimura, A., Kaneko, T., Kato, Y., Golenbock, D. T. & Hara, Y. (2002). Lipopolysaccharides from periodontopathic bacteria *Porphyromonas gingivalis* and *Capnocytophaga ochracea* are antagonists for human toll-like receptor 4. *Infect Immun.* 70, 218-25.
- Zaitlen, N. & Kraft, P. (2012). Heritability in the genome-wide association era. *Hum Genet* 131, 1655-64.

- Zhang, B., Li, Y., Zhou, Q. & Ding, Y. (2011). Estrogen deficiency leads to impaired osteogenic differentiation of periodontal ligament stem cells in rats. *Tohoku J. Exp. Med.* 223, 177-86.
- Zhong, Q. F., Ding, C., Wang, M. L., Sun, Y. & Xu, Y. (2012). Interleukin-10 gene polymorphisms and chronic/aggressive periodontitis susceptibility: A meta-analysis based on 14 case-control studies. *Cytokine* 60, 47-54.

Chapter 32

GENOMIC IMPRINTING AND THE BRAIN: NEURON-SPECIFIC SWITCHING OF GENE EXPRESSION AT IMPRINTING REGIONS

Masamitsu Eitoku and Hidenori Kiyosawa*

Department of Environmental Medicine, Kochi Medical School,
Kochi University, Nankoku, Kochi, Japan

ABSTRACT

Imprinted regions of the mammalian genome are commonly composed of one or more paired paternally and maternally imprinted genes and differentially methylated regions (DMRs). These DMRs are methylated in an allele-specific manner during germ-line or early embryonic development, and they regulate allele-specific gene expression. Although genes in most imprinted regions are expressed ubiquitously among tissues, some imprinted regions manifest tissue-specific gene expression patterns. The brain is a major site of tissue-specific gene expression from imprinted regions in embryonic tissues.

We summarize several imprinted regions that show neuron-specific gene expression patterns. First, we describe the *SNRPN-UBE3A* domain, which contains three brain-specific imprinted genes, *SNORD115*, *UBE3A-ATS* and *UBE3A*, as well as one imprinted gene with a brain-specific promoter, *SNRPN*. Second, we discuss the *DLK1-DIO3* domain, which contains the brain-dominant imprinted genes *SNORD112*, *SNORD113* and *SNORD114* as well as numerous brain-specific imprinted miRNAs. Third, we provide an overview of *Grb10*, which also has a brain-specific promoter and switches the imprinted allele during early neurogenesis.

Our chapter reviews these imprinted regions and describes the similarities and differences of the neuron-specific imprinting switch in each region.

* Corresponding Author's Email: kiyosawa-fks@umin.org.

INTRODUCTION

Genomic imprinting is an intriguing epigenetic phenomenon leading to allele-specific or allele-biased gene expression. Among the estimated 25,000 mammalian genes, a few hundred are imprinted, and some of these are clustered into restricted genomic regions [1]. Most but not all of these genes are regulated by differentially methylated regions (DMRs) within the cluster. This expression system is allele-specific and plays an essential role in embryogenesis. Mammalian reproduction requires both sexes as donors of gametes; the gametes from each parent, sperm and oocyte, function as vectors to transmit the imprinting marks, the pattern of methylation, from the paternal and maternal genomes to the embryo. The DMR is modified in two steps to enable the inheritance of these imprinting marks, which are maintained in somatic cells. The first step is erasure of the methylation of the DMR during the development of primordial germinal cells. The second step is the establishment of DMRs during gametogenesis in a sex-specific manner, resulting in different imprinting patterns in oocytes and sperm. The DMRs of the gametes are maintained and shared among the somatic cells that arise from them; however, tissue-specific genomic imprinting can also occur. Most of the tissue-specific imprinted genes are imprinted in extra-embryonic tissues such as placenta and yolk sac, but several neuron-specific imprinted genes have been reported. In this chapter, we provide an overview of representative imprinted regions that contain imprinted genes with neuron-specific expression patterns, focusing on the mechanism of the neuron-specific imprinting switch.

Overview of the *SNRPN-UBE3A* Domain

The imprinted domain *SNRPN-UBE3A*, located on human 15q11-q13 and mouse 7qC [2, 3], consists of two imprinted regions (Figure 1A, B). The first region contains several paternally expressed genes (*makorin/ring finger protein 3* [*MKRN3/ZNF127*], *MAGE-like 2* [*MAGEL2*], *Necdin* [*NDN*], *small nuclear ribonucleoprotein polypeptide N* [*SNRPN*], *imprinted in Prader-Willi syndrome* [*IPW*], several small nucleolar RNA [snoRNA] genes including *SNORD116/HBII-85* [*SNORD116*] and *SNORD115/HBII-52* [*SNORD115*] and the long non-coding RNA *UBE3A antisense* [*UBE3A-ATS*]). The second region contains a maternally expressed gene, *UBE3A* [4]. In both human and mouse, this domain contains at least three brain-specific imprinted genes, paternal *SNORD115* [5-7] and *UBE3A-ATS* [8, 9] and maternal *UBE3A* [10-12], and one gene with a brain-specific promoter, *SNRPN* [13-16]. This *SNRPN-UBE3A* domain is also a critical region for two neurobehavioral disorders, Angelman syndrome (AS) and Prader-Willi syndrome (PWS). The genes responsible for AS and PWS are *UBE3A* and *SNORD116*, respectively [17-21].

Igf2r, *Igf2* and *H19* were identified as the first three imprinted genes in mouse in 1991 [22-25]; *SNRPN*, which was the fourth to be identified [26], expresses a bicistronic transcript containing 10 exons. Exons 1-3 code for the protein “SNRPN upstream reading frame” (SNURF), and exons 4-10 code for the polypeptide protein “small nuclear ribonucleoprotein” (SmN) [27-34]. Although the exon 1 promoter drives expression of *SNRPN* ubiquitously, there are multiple untranslated upstream exons called U exons [13]. U1B and U1A function as brain-specific paternal promoters [13, 14], and U5 and U6 as oocyte-specific maternal promoters [35]. In both of brain-specific and oocyte-specific U exon-containing transcripts, U exons skip

exon 1 and splice into exon 2. The U exons also share sequence similarity, suggesting that they arose from genomic duplication [14].

The imprinting center (IC) in this locus is a bipartite IC composed of the shortest regions of deletion overlaps in AS patients (designated AS-SRO) and PWS patients (designated PWS-SRO), and it controls all imprinting of the *SNRPN-UBE3A* domain [13, 30, 36-40]. PWS-SRO, which spans exon 1, is also thought to regulate the expression of a large transcription unit that spans a genomic region of more than 460 kb and contains at least 148 exons, including those of *SNRPN*, *IPW* and *UBE3A-ATS* [41]. This transcript is alternatively spliced and serves as the host for the six intronic C/D box snoRNA species. Two of these species, *SNORD116* and *SNORD115*, are present in tandemly repeated clusters containing 29 and 48 gene copies, respectively [42].

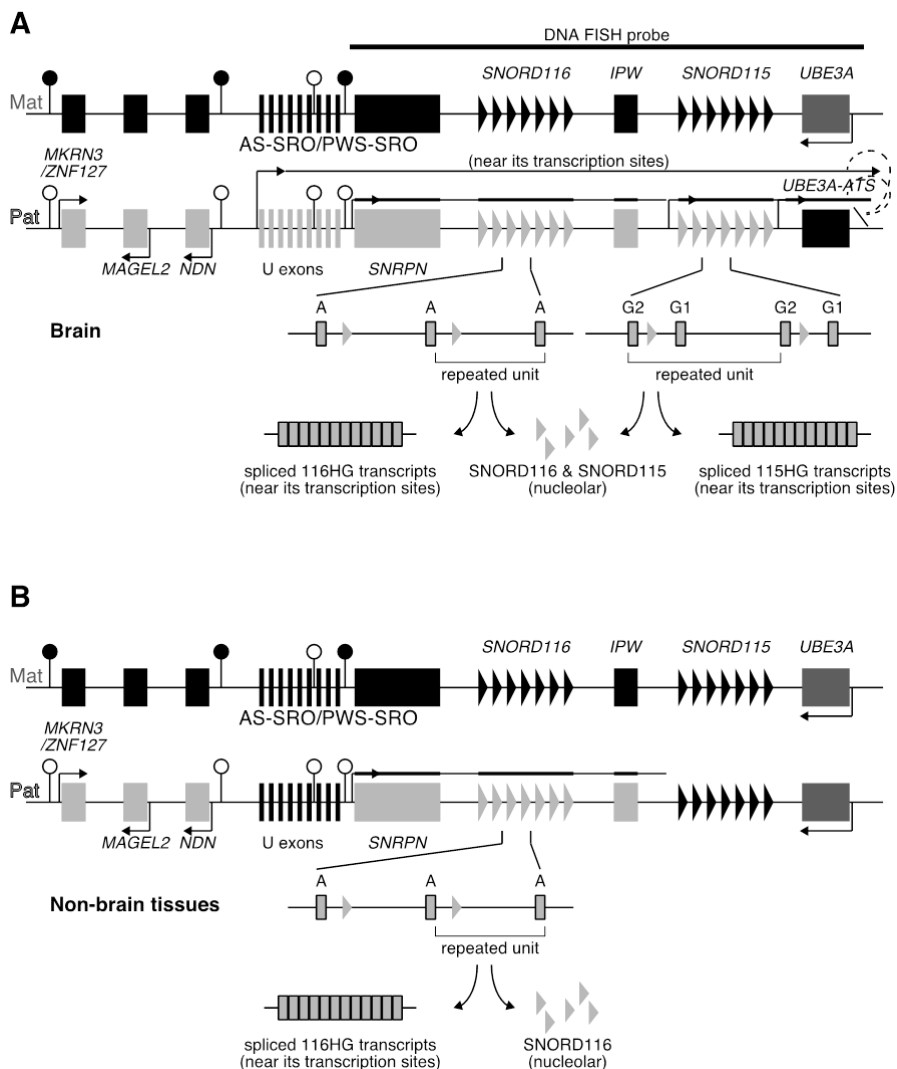


Figure 1. (Continued).

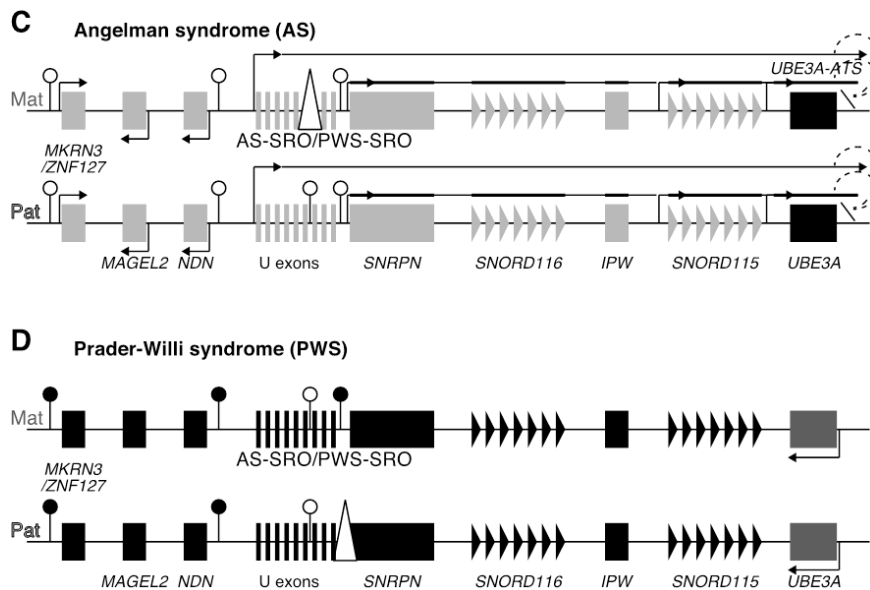


Figure 1. Overview of the *SNRPN-UBE3A* domain. (A, B) Schematic representation of the human *SNRPN-UBE3A* domain in (A) brain and (B) non-brain tissues. Paternally expressed non-coding RNAs, including alternative U exon transcripts, *SNRPN*, snoRNA host gene transcripts (*116HG* and *115HG*), *IPW* and *UBE3A-ATS*, form part of a large transcription unit. The paternal *UBE3A* expression in brain is down-regulated by the neuron-specific paternal *UBE3A-ATS* transcripts driven from the U exons or exon 1, and the shortened *UBE3A-ATS* that possibly starts from a transcription start site (TSS) near *SNORD115*. U exons, *Snord116*, *Snord115* and *Ube3a-ATS* RNAs localize to their transcription sites. *SNRPN*, *116HG* and *IPW* transcripts are found in many tissues. In this and all subsequent figures, maternally and paternally inherited chromosomes are indicated as Mat and Pat, respectively, and figures are not drawn to scale. Genes and U exons are symbolized as rectangles and vertical bars, respectively. The orientation of transcription is shown by arrows. The alleles exhibiting maternal and paternal imprinted expression patterns are depicted in dark gray and light gray, respectively, whereas silenced alleles are depicted in black. Tandemly repeated snoRNA gene clusters contain intron-encoded snoRNAs (triangles) and nuclear-retained host genes (framed rectangles), which accumulate near their transcription sites. Methyated and unmethylated DMRs are indicated as lollipops with black and white circles, respectively. The black bar above the map indicates the location of the DNA FISH probe utilized to show the neuron-specific chromatin decondensation of the active paternal allele. (C) Schematic representation of the *SNRPN-UBE3A* domain of AS patients with maternal transmission of the AS-SRO deletion (white triangle). *UBE3A*, the gene responsible for AS, is not expressed. (D) Schematic representation of the *SNRPN-UBE3A* domain of PWS patients with paternal transmission of the PWS-SRO deletion (white triangle). *SNORD116*, the gene responsible for PWS, is not expressed.

Three of the snoRNA species, *SNORD64/HBII-13* (*SNORD64*), *SNORD107/HBII-436* (*SNORD107*) and *SNORD108/HBII-437* (*SNORD108*), are present as single-copy genes, and the last one, *SNORD109A/HBII-438A* (*SNORD109A*) and *SNORD109B/HBII-438B* (*SNORD109B*) [5, 41, 43, 44], is a double copy gene. *SNORD116* is ubiquitously expressed in substantial amounts, but is expressed at much higher levels in brain [5-7, 42]. *SNORD115* is also expressed primarily in brain [5-7] and modestly in kidney, liver, skeletal muscle and thyroid [42].

UBE3A-ATS is expressed exclusively in brain from the paternal allele [8]. In contrast, *UBE3A* is expressed biallelically in most tissues, but is only expressed from the maternal chromosome in various parts of the brain [11, 12]. Imprinting of *UBE3A* is not associated with

DMRs [45, 46], but the reciprocal expression of *UBE3A-ATS*, which overlaps with at least the 3' half of *UBE3A* and possibly with most of the gene body in humans, is thought to play a role in silencing the paternal *UBE3A* in brain [8, 41, 47].

Bipartite IC of the *SNRPN-UBE3A* Domain

Studies of the first reported human imprinting disorders, AS and PWS, provided evidence on the *cis* elements regulating allele-specific expression [4, 13, 30]. Mapping of micro-deletions of chromosome 15q11-q13 in families of AS and PWS patients identified AS-SRO and PWS-SRO (Figure 1C, D).

AS-SRO is an 880-bp region located 35 kb upstream of *SNRPN* exon 1 [37]. It contains an upstream, oocyte-specific promoter [35] and U exons U5 and U6 of *SNRPN*, which are present in a very small fraction of *SNRPN* splice variants [4, 14, 48]. In contrast, U exon-containing transcripts in oocytes start from these exons but do not include exon 1, which contains the major transcription start sites (TSSs) in brain. U exon-containing transcripts that initiate at U1B or U1A [14], the other common TSSs in brain, are also absent in oocytes, suggesting that *SNRPN* transcription has nearly reciprocal patterns of initiation on the paternal allele in brain and on the maternal allele in oocytes (Figure 2) [35].

PWS-SRO is a 4.1-kb region that includes the paternal allele-specific *SNRPN* promoter and exon 1. Although the maternal and paternal copies of PWS-SRO are, respectively, methylated and unmethylated [31], abnormally methylated paternal PWS-SRO silences the paternal-specific expression of *MKRN3/ZNF127*, *SNRPN* and *IPW* [36, 49]. The unmethylated (active) paternal *SNRPN* promoter is enriched with active histone marks (panacetylated H3, H4 and H3K4me), and the methylated (repressed) maternal *SNRPN* promoter is enriched with repressive histone marks (H3K9me) [50-52].

Developmental Regulation of the Bipartite IC

A series of analyses in mouse elucidated how these *cis* elements mutually interact and regulate one another during development [53]. Transgenic F0 mice that bear an exogenous transgene fragment containing human AS-SRO and PWS-SRO were created [54]. This exogenous fragment transmitted from the male or female F0 founder to their F1 offspring mimicked the imprinting patterns of DNA methylation and gene expression in endogenous paternal or maternal alleles, respectively. This system allowed the establishment and maintenance of differential methylation of AS-SRO and PWS-SRO to be studied in the gamete and especially in the embryo.

AS-SRO is methylated during spermatogenesis but not oogenesis, resulting in its acquisition of the paternally methylated pattern (Figure 2) [54]. This differential methylation is maintained through implantation but is not seen in the adult, consistent with the biallelic DNA methylation pattern in AS-SRO seen in human somatic cells [55, 56]. In contrast, the methylation of PWS-SRO, which has the maternally methylated pattern in somatic cells [31], takes place at the post-implantation stage [54], consistent with the absence of PWS-SRO methylation in human gametes [39].

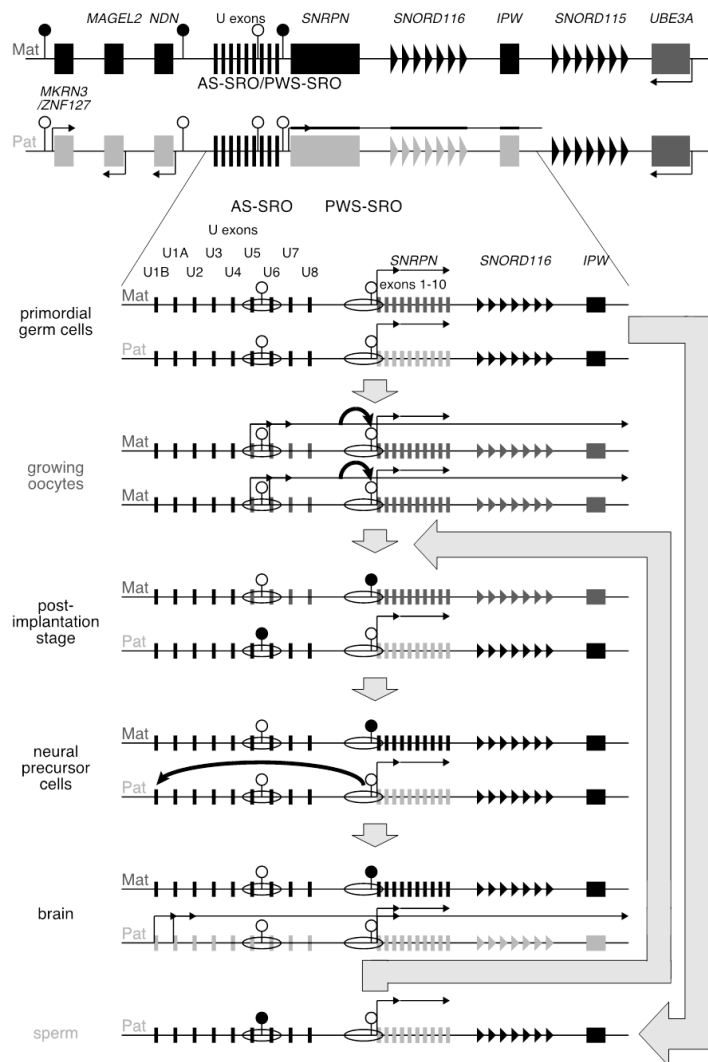


Figure 2. Model for the establishment of brain-specific imprinting at the *SNRPN-UBE3A* domain, derived from human and mouse experiments. Schematic diagrams of the locus near *SNRPN* at various stages of development. Imprinting marks are erased in primordial germ cells, and the U exon transcripts are not expressed. In growing oocytes, U exon transcription is initiated from the oocyte-specific U exons in AS-SRO. Methylated paternal AS-SRO inherited by sperm is maintained through the post-implantation stage but is not seen in adults. Methylation of maternal PWS-SRO/PWS-IC occurs in the post-implantation stage in humans and in growing oocytes in mice. In mice, U exon transcripts themselves may contribute to this methylation and further maternal silencing. Unmethylated paternal PWS-IC is needed for brain-specific paternal expression of U exon-containing transcripts. Biallelic and monoallelic expression of mouse *Snrpn* is observed during gametogenesis and post-fertilization development, respectively. The AS-SRO and PWS-SRO are depicted as ovals overlying U exons 5/6 and *SNRPN* exon 1, respectively. In both of the brain-specific and oocyte-specific U exon-containing transcripts, U exons skip exon 1 and splice into exon 2. Exons of *SNRPN*, snoRNAs of *SNORD116* and *IPW* are symbolized as vertical bars, triangles and rectangles, respectively. The genes and exons exhibiting maternal and paternal imprinted expression patterns are depicted in dark gray and light gray, respectively, whereas silenced alleles are depicted in black. Transcription is shown by arrows. Methylated and unmethylated SROs are indicated as lollipops with black and white circles, respectively. Wide arrows show transitions between developmental stages.

AS-SRO is necessary for the establishment, but not for the maintenance, of PWS-SRO methylation [54]: a transgenic mouse with an exogenous transgene fragment containing human PWS-SRO and AS-SRO flanked by lox sequences was mated with a Cre-expressing mouse to remove AS-SRO; maternally transferred PWS-SRO was unmethylated when AS-SRO was removed prior to gametogenesis and methylated when it was removed after fertilization. A genetic analysis of AS and PWS families produced the same result, and also showed that PWS-SRO is not necessary for the establishment of AS-SRO methylation [56]. These results suggest that the primary, but temporary, imprinting occurs on AS-SRO during gametogenesis (gDMR) and causes allele-specific repression of the maternal PWS-SRO, resulting in the establishment of the secondary, and permanent, imprinting of PWS-SRO during post-fertilization development (sDMR).

The Role of U Exon Transcription in Maternal Silencing of *Snrpn*

Studies of other transgenic mouse lines demonstrated that transcription from the *Snrpn* U exons (U exon transcription) regulates the maternal imprinting at the PWS-IC [57], the region homologous to PWS-SRO in mouse [15, 36, 58, 59]. Although there is no mouse region homologous to the human AS-SRO, an AS-IC was identified as its functional equivalent in a series of knockout and transgenic mice, suggesting that humans and mice share a similar mechanism for establishing genomic imprinting with bipartite elements in this region [57, 60–62].

Two different transgenic mouse lines carrying the same single copy of a modified BAC transgene in different loci exhibited different imprinting patterns [57]. In the *Snrpn* upstream region, there are nine U exons lacking splice acceptor signals and five additional exons containing splice acceptor signals; all 14 exons contain splice donor signals. This BAC transgene contains *Snrpn* and 100 kb of upstream sequence spanning three of the U exons, all of which are deleted in this construct, and three of the additional exons. One transgenic mouse line lacked U exon-containing transcripts in the ovary, resulting in an imprinting defect, biparental *Snrpn* expression and DNA hypomethylation at the PWS-IC. The other line expressed an aberrant transcript containing an additional upstream exon in the ovary and displayed appropriate maternal silencing. These results suggest that U exon transcription through the PWS-IC is essential for AS-IC activity (Figure 2). In human, an AS patient had a point mutation in the splice donor site of U2, located upstream of the oocyte-specific promoter U5, suggesting that the U exon-containing transcript itself and not the process of transcription from the upstream promoter is important for the establishment of maternal silencing [13, 14].

The timing with which maternal silencing is established in the wild-type mouse also supports this idea (Figure 2). Demethylation of the PWS-IC occurs in the primordial germinal cells between 10.5 to 12.5 days postcoitus (dpc), resulting in biallelic expression of *Snrpn* [63–65]. While this biallelic expression pattern of *Snrpn* is maintained in both oocytes and spermatogenic cells, monoallelic expression of *Snrpn* is observed in each blastomere at the 4-cell stage [63, 66]. The U exon-containing transcripts are undetectable in 13.5 dpc primordial germ cells but become evident after birth [57]. Because the PWS-IC is first methylated in growing oocytes between 5 and 25 days postpartum (dpp) in a size-dependent-manner [67, 68], PWS-IC methylation occurs in the presence of U exon-containing transcripts.

A smaller (4.8 kb) deletion of the maternal PWS-IC resulted in a maternal imprinting defect that was accompanied by the expression of paternally expressed genes, such as *Snrpn*, *Snord116* and *Snord115*, from the maternal allele [69]. Brains from mice with a maternal PWS-IC knockout exhibited increased amounts of the maternal U exon-containing transcript that is silenced in the wild-type mouse, suggesting that the PWS-IC is also required for maternal imprinting. Thus, maternal methylation of the PWS-IC, which is mediated by AS-IC in oocytes, has a role not only in maintaining the silencing of ubiquitous *Snrpn* transcription that initiates from the exon 1 promoter, but also in down-regulating maternal U exon transcription that initiates from the oocyte-specific promoter U5, resulting in the ubiquitous maternal expression of *Ube3a* (Figure 2).

Regulation of Paternal Expression of *Ube3a-ATS*

A study of murine embryonic carcinoma cells (ECCs) provided evidence of the correlation between U exon transcription and *Ube3a-ATS* expression on the paternal allele in neuron [16]. Although *Snrpn* was expressed consistently during the time course of ECC neural differentiation, several U exon-containing transcripts and *Ipw/Ube3a-ATS* were up-regulated synchronously. Consistent with this result, detailed expression analyses of U exon-containing transcripts and *Ube3a-ATS* in mouse revealed an overlap in their tissue-specific and developmental expression [70]. Additionally, a knockout mouse with a 35-kb deletion, including a 16-kb upstream region containing the PWS-IC and six exons of the *Snrpn* gene, exhibited loss of paternal expression of the U exon-containing transcript and *Ube3a-ATS* and gain of paternal expression of *Ube3a* [9, 16].

Three additional mouse models with PWS-IC deletions distinguished the elements of the exon 1 promoter and the PWS-IC functional region. The 0.9-kb deletion did not affect the expression of U exon-containing transcripts despite a severe reduction in *Snrpn* transcripts [15, 71]. The 4.8-kb deletion of the paternal PWS-IC caused a partial defect of paternal imprinting of the *Ndn* locus and resulted in decreased expression of the paternal U exon-containing transcript and an increase of paternal *Ube3a* expression in brain, suggesting that the unmethylated paternal PWS-IC activates brain-specific U exon transcription (Figure 2) [69, 71]. The 6-kb deletion, which extended 1 kb further upstream of the 4.8-kb deletion, exhibited results similar to those obtained with the 35-kb PWS-IC deletion [59], including undetectable expression of paternal-only genes such as *Mkx3/Zfp127*, *Magel2* and *Ndn* as well as increased DNA methylation at the *Mkx3/Zfp127* and *Ndn* loci [72]. These results suggest the following two conclusions: (1) elements lying within the 0.9-kb PWS-IC deletion function as an exon 1 promoter to regulate ubiquitous *Snrpn* expression, and (2) elements lying within the 6-kb deletion but outside the 0.9-kb deletion are required for the PWS-IC to establish brain-specific paternal expression of the U exon-containing transcripts.

These two paternal transcripts, U exon-containing transcript and *Ube3a-ATS*, are expressed in neurons (but not glia) derived from primary cortical cultures [73, 74]; they are also linked on cDNAs derived from mouse brain, supporting the regulation of *Ube3a-ATS* expression by U exons [16]. Strand-specific RT-PCR of the region around *Ube3a-ATS* that overlaps the *Ube3a* 3'-UTR identified several transcripts containing some exons of *Ipw* and a part of *Ube3a-ATS* that overlaps with the *Ube3a* 3' end. Additional RT-PCR analysis of the region between *Snrpn* U1 and *Ipw* identified a short transcript that excludes *Snrpn*. These

results suggest that *Ube3a*-ATS expression is driven from brain-specific paternal *Snrpn* U exons and extends over a 1000-kb genomic region (Figure 1A). In expression analyses, *Ube3a*-ATS was reduced by 50% in mice with the 0.9-kb deletion of *Snrpn* exon 1, implying that mouse *Ube3a*-ATS is also transcribed from the *Snrpn* major promoter as it is in humans and that these two forms of the transcripts, *Ube3a*-ATS driven from U exons and exon 1, are expressed at the same level [71].

However, a recent CAGE-seq analysis in human libraries provided evidence of the existence of TSSs between the *SNORD116* and *SNORD115* gene clusters and within the *SNORD115* gene array, suggesting that transcription could also be initiated in the middle of the *SNRPN/UBE3A*-ATS transcription unit [42], producing a shortened *UBE3A*-ATS. Derivation and characterization of human induced pluripotent stem cells (iPSCs) from a PWS patient revealed that the region between *SNRPN* and *SNORD116* plays a role in the silencing of transcripts spanning *SNORD115* and *UBE3A*-ATS in non-neuronal cells and unsilencing them in neurons [75]. iPSCs with a small deletion spanning *SNRPN*, the *SNORD116* cluster and *IPW* expressed paternal *UBE3A*-ATS and silenced paternal *UBE3A* like neurons. The brain-specific U exon-containing transcripts, however, were not expressed in these iPSCs but were normally expressed in neurons differentiated from the same iPSCs, suggesting that the paternal *UBE3A*-ATS expressed in this iPSCs from a PWS patient is a shortened one driven from TSSs near *SNORD115* and neuron-specific paternal *UBE3A* silencing requires only paternal shortened *UBE3A*-ATS expression and no other neuron-specific factors (Figure 1A).

Neuron-Specific Chromatin Decondensation and Localization of Paternal Transcripts to Their Transcription Sites

Combined RNA and DNA fluorescent *in situ* hybridization (FISH) analysis revealed that *Ube3a*-ATS localizes to its transcription site in the nucleus (Figure 1A) [76], suggesting that it down-regulates *Ube3a* expression by forming a complex chromatin structure despite its short half-life of 4 h [71]. DNA FISH analyses revealed that the chromatin structure of the paternal *Snrpn-Ube3a* domain is decondensed and transcriptionally regulated in a neuron-specific manner through the PWS-IC (Figure 1A) [77, 78]. This decondensation occurs within the first 2 weeks after birth, prior to accumulation of *Snord116* in the nucleolus, and is required for increased nucleolar size during neuronal maturation. A combined RNA/DNA FISH analysis also revealed that *Snord116* and *Snord115* transcripts are processed into multiple snoRNAs from introns and into spliced, nuclear-retained host genes (*116HG* and *115HG*) from exons (Figure 1A). These *116HG* and *115HG*, which are repeated-exon-containing mRNA-like transcripts, also accumulate near their transcription sites, juxtaposed to U exon transcripts on the decondensed paternal genome [79, 80].

Combined immunofluorescence of the *Snrpn-Ube3a* domain with the S9.6 antibody, a DNA:RNA-specific antibody, and DNA FISH revealed that the decondensed paternal allele in the *Snrpn-Ube3a* domain overlaps with the DNA:RNA hybrid in adult mouse cerebellum [78]. Although the results of extensive RNase A treatment suggested that the chromatin decondensation might be caused by DNA:DNA rather than DNA:RNA hybridization [77], these results of FISH and immunofluorescence analyses suggest that some or all of the neuron-specific paternal genes, including the U exons, *Snord116*, *Snord115* and *Ube3a*-ATS, would localize to the transcription sites of these genes and form DNA:RNA hybrids.

The *Snord116* region was shown to form DNA:RNA hybrids called R loops *in vitro* and *in vivo* [78-80]. This region is required for topotecan-mediated chromatin decondensation at the *Snrpn-Ube3a* domain and for *Ube3a* silencing in cultured murine neurons, consistent with the human result indicating that topotecan treatment is less effective in repressing the *UBE3A-ATS* in PWS patient-derived iPSCs with a small deletion of the *SNORD116* cluster [78]. Topotecan, a topoisomerase I inhibitor, was found as a novel drug that reduces *Ube3a-ATS* expression by forming supercoils in front of the elongating RNA polymerase and unsilencing paternal *Ube3a* in mouse neurons [81]. Topotecan treatment also stabilizes increased R-loop formation at *Snord116* [78]. Because R loops impair transcription elongation [82-84] and the deletion of *Snord116/SNORD116* gene results in increasing of *Ube3a-ATS/UBE3A-ATS* in mice and humans [75, 78], R-loop formation in the *Snord116* cluster may be involved in the regulation of the paternal *Ube3a-ATS* expression.

Model of Paternal *Ube3a* Silencing by *Ube3a-ATS* Transcripts

Although RT-PCR analyses identified *Ube3a-ATS* transcripts overlapping the middle of the gene body and the 3'-UTR of *Ube3a* [9, 16], an single nucleotide polymorphism (SNP) array analysis and strand-specific microarray revealed that brain-specific *Ube3a-ATS* possibly terminates in the region 40 kb upstream of the *Ube3a* TSS, suggesting that *Ube3a-ATS* is actively transcribed at the *Ube3a* promoter (Figure 1A) [71, 85]. The amount of *Ube3a-ATS* transcripts decreased significantly in the middle of the locus toward the transcription termination site [71]. This is approximately the same region where the biallelic expression of *Ube3a* pre-mRNA is observed [76, 85]. These results support the transcriptional collision hypothesis, in which the two opposing polymerases II transcribing *Ube3a* and *Ube3a-ATS* on the paternal chromosome collide with each other and drop off of the template, thereby aborting transcription of both [71]. Results from the truncation of *Ube3a-ATS* also supported this hypothesis. A transcriptional termination cassette inserted downstream of *Ube3a* on the paternal chromosome reduced *Ube3a-ATS* transcription and unsilenced *Ube3a* expression *in vitro* and *in vivo* [71, 76]. This transcriptional termination cassette also suppressed the behavioral defects of an AS model mouse [76].

However, specific elimination of *Ube3a-ATS* transcripts with antisense oligonucleotides (ASOs) clarified that the transcripts of *Ube3a-ATS*, and not transcription per se, carry out neuron-specific paternal *Ube3a* silencing *in vitro* and *in vivo* [86]. ASOs, which are chemically modified DNAs of ~20 bp in length, penetrate into the nucleus and specifically hybridize with the target RNA. The RNA in the ASO-RNA heteroduplex is cleaved by RNase H and subsequently degraded by exonucleases without inhibiting transcription. Treatment with an ASO, that contained the sequence of the downstream region of the *Ube3a* gene to target *Ube3a-ATS*, reduced paternal expression of *Ube3a-ATS* and unsilenced paternal *Ube3a* in the brain (especially the cortex and hippocampus) and spinal cord. Because *Ube3a-ATS* transcripts span the *Ube3a* promoter and localize to its transcription site [71, 76, 85], *Ube3a-ATS* transcripts may cover *Ube3a* promoter and block the access of RNA polymerase II (Figure 1A).

Overview of the *DLK1-DIO3* Domain

The imprinted *DLK1-DIO3* domain is located in the human 14q32 and mouse 12qF1 regions (Figure 3A). Like the *SNRPN-UBE3A* domain, it contains tandemly repeated C/D box snoRNA gene clusters that are imprinted, in this case *SNORD112/14q(0)* (*SNORD112*), *SNORD113/14q(I)* (*SNORD113*) and *SNORD114/14q(II)* (*SNORD114*), which are present in 1, 9 and 31 gene copies, respectively. These clusters, first identified in mouse as the non-coding RNA gene named *RNA imprinted and accumulated in nucleus (Rian)*, exhibit brain-dominant and brain-specific maternal expression in human and mouse, respectively [6, 87]. Most of these snoRNAs are intron-encoded and processed from a complex transcription unit mapping to *maternally expressed gene 8 (MEG8)* [6]. The imprinted domain *DLK1-DIO3* consists of the paternally expressed protein-coding genes *delta-like 1 (DLK1)*, *retrotransposon-like 1 (RTL1)* and *deiodinase, iodothyronine type III (DIO3)*, as well as the maternally expressed non-coding genes *MEG3*, *MEG8*, *antisense RTL1 (RTL1as)* and one of the largest microRNA (miRNA) clusters in the genome [88, 89]. These maternal genes and their maternally expressed intergenic transcripts are all expressed in the same orientation [6, 90-93], suggesting that they may form a large polycistronic transcription unit like *UBE3A-ATS* [93].

Mouse strains with uniparental duplications of chromosome 12 were used for the identification and characterization of DMRs in this domain [94]. The paternal intergenic DMR (IG-DMR), which is located upstream of *Gtl2/Meg3 (Gtl2)*, the mouse homolog of *MEG3*, has a methylation mark that is transmitted from the sperm as a gDMR [94]. The IG-DMR regulates the methylation status of the *Gtl2* DMR, which is a sDMR spanning the *Gtl2* promoter and its first exon and intron, and regulates monoallelic expression of all genes in this domain [95-97]. In contrast to the regulation of PWS-SRO by AS-SRO in the *SNRPN-UBE3A* domain [54], absence of the maternal IG-DMR results in *Gtl2* DMR hypermethylation and expression of the paternal chromosome, suggesting that the unmethylated maternal IG-DMR functions as an epigenotype switch that changes the default paternal status to the maternal status (Figure 3B, C) [95].

Gtl2, a non-coding poly-adenylated transcript, has multiple alternatively spliced forms of unknown function [98-100]. Short transcript variants (*Gtl2v3* and *4*) have 10 small exons, and long transcript variants (*Gtl2v1* and *2*) contain the largest (10 kb) exon at their 3' ends. An *in situ* hybridization analysis revealed that *Gtl2v1* is expressed in early neurons at 15.5 dpc but not in their progenitors at 12.5 dpc [101]. Because *Gtl2* regulates the occupancy of polycomb repressive complex-2 on chromatin through a direct interaction [102], *Gtl2v1* may be involved in neuron-specific gene expression.

A DNA FISH probe for the mouse *Gtl2* domain that encompasses *Rian* and *microRNA-containing gene (Mirg)*—regions equivalent to human *MEG8* and the large miRNA cluster—exhibited signals with one small compact signal (probably the inactive paternal allele) and one decondensed signal (probably the active maternal allele) in neurons but not glia, similar to *SNORD115* and *SNORD116* in the *SNRPN-UBE3A* domain (Figure 3A) [77]. DNA FISH probes flanking *Gtl2* revealed no chromatin decondensation. Combined RNA/DNA FISH analysis revealed that a spliced, nuclear-retained host gene of *RBII-36*, the snoRNA gene in rat *Dlk1-Dio3* domain, also accumulates near its transcription sites in brain [79].

Overview of the *GRB10* Domain

Growth factor receptor-bound protein 10/MEG1 (*GRB10*) is located on human 7p11.2–p12 and in mouse proximal chromosome 11 regions [103–105] and encodes an adapter protein [106]. *GRB10* is a unique imprinted gene that undergoes a brain-specific imprinting switch during development in human and mouse (Figure 4A, B, C). Human *GRB10* exhibits biallelic expression in almost all tissues and organs, except for the fetal brain, where it is expressed paternally [107–109].

In mouse, although the timing of the imprinting switch differs among strains, *Grb10* has been reported to be expressed maternally and paternally in brains of 11.5 and 14.5 dpc embryos with C57BL/6:CBA mixed genetic background, respectively [110], although it is only expressed maternally in most other fetal and adult tissues [109–111]. As with the U exons in the *Snrpn-Ube3a* domain, a series of downstream alternative promoters, 1C, 1B1 and 1B2, control this brain-specific paternal expression [109, 111–113].

In the region surrounding these brain-specific promoters, there is a gDMR named *Grb10*-DMR, which contains a methylated maternal allele and an unmethylated paternal allele [109, 111, 112]. A mouse-specific repeat sequence found in *Grb10*-DMR functions as a binding site for CCCTC-binding factor (CTCF) in a DNA methylation-sensitive manner [114]. The deletion of *Grb10*-DMR in the paternal allele changes the paternal transcripts in brain from a brain-specific isoform to a major isoform (Figure 4D) and unsilences paternal expression from the major promoter 1A in non-brain tissues, resulting in biallelic expression (Figure 4E). Maternal deletion of the *Grb10*-DMR produced no observable adverse phenotype in any tissue (Figure 4F, G). These results suggest that the paternal *Grb10*-DMR acts as an IC and represses the paternal major promoter 1A. Knockdown of CTCF in mouse embryonic stem cells (mESCs) by using short hairpin RNA also resulted in up-regulation of the major isoform with no change in the repressed brain-specific isoform. Although it is difficult to exclude indirect effects of CTCF knockdown, this result suggests that CTCF represses the paternal major promoter, possibly through its binding to *Grb10*-DMR [115].

Whereas DNA methylation and CTCF account for part of the repression mechanism, histone modification accounts for the activation mechanism (Figure 4A, B, C). The unmethylated paternal *Grb10*-DMR is activated in brain and repressed in non-brain tissues. It is enriched with active histone marks (H3K4me2 and H3K9ac) in brain; however, it has both active (H3K4me2) and repressive (H3K27me3) marks in embryo and kidney [113]. Repressed DNA-methylated maternal *Grb10*-DMR is enriched with repressive histone marks (H3K9me3 and H4K20me3) in both brain and non-brain tissues.

Experiments on *in vitro* neural differentiation of mESCs also revealed that the histone mark on the paternal *Grb10*-DMR changed from H3K27me3/H3K4me2 to H3K9ac/H3K4me2 during differentiation. Additionally, in mESC deficient for EED—a component of the polycomb repressive complex 2 that mediates H3K27 tri-methylation—the brain-specific promoter was derepressed, possibly on the paternal allele normally enriched with H3K27me3. These results suggest that the paternal brain-specific promoter near the *Grb10*-DMR is poised but silenced through the bivalent chromatin domain (H3K4me2/H3K27me3) in undifferentiated cells and most non-brain tissues, whereas replacement of repressive H3K27me3 with active H3K9ac during neural differentiation activates the paternal brain-specific promoter.

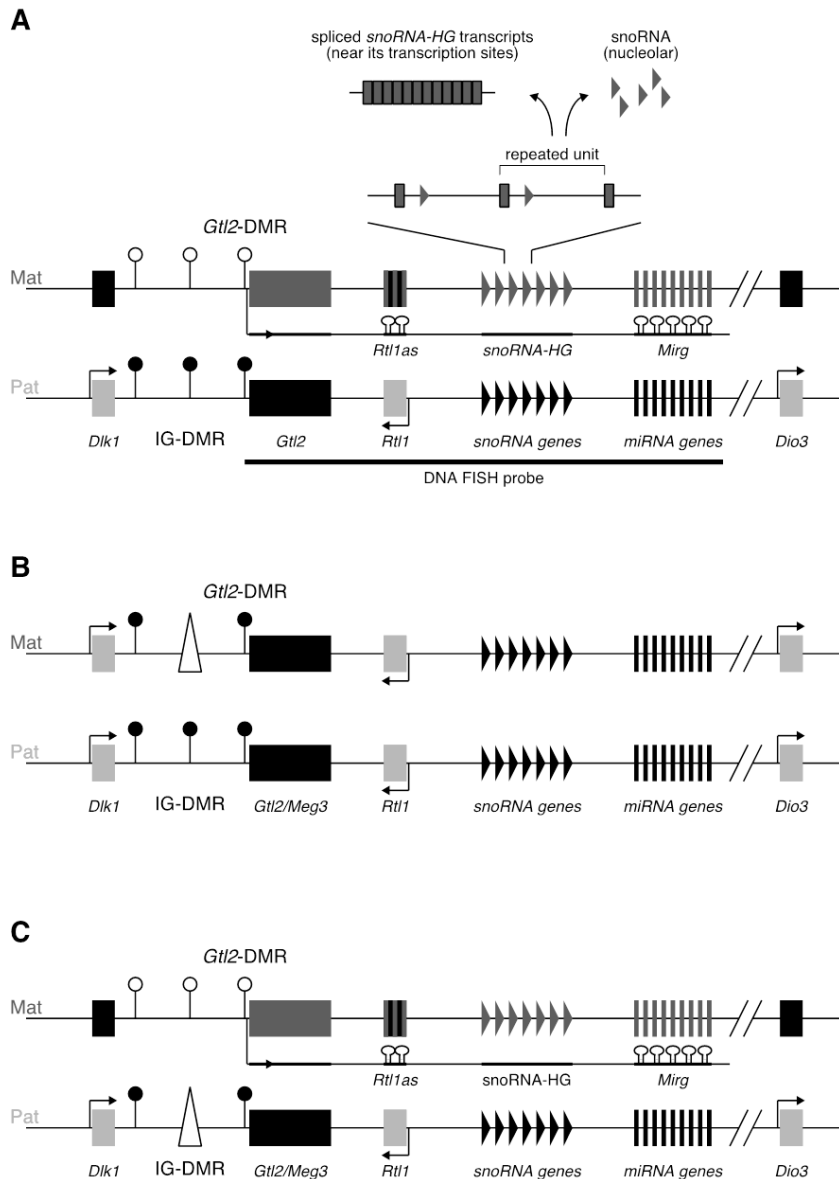


Figure 3. Overview of the *Dlk1-Dio3* domain. (A) Schematic representation of the *Dlk1-Dio3* domain, derived from the mouse and rat experiments. The alleles exhibiting maternal and paternal imprinted expression patterns are depicted in dark gray and light gray, respectively, whereas silenced alleles are depicted in black. Maternally expressed non-coding RNAs (*Gtl2*, *Rtl1as*, *snoRNA-HG* transcripts and *Mirg*) constitute part of a large polycistronic transcription unit. The cluster of tandemly repeated *snoRNA* genes contains intron-encoded *snoRNAs* (triangles) and nuclear-retained *snoRNA-HG* transcripts (framed rectangles), which accumulate near their transcription sites. The *miRNA* and other genes are symbolized as vertical bars and rectangles, respectively. The orientation of transcription is shown by arrows. Methylated and unmethylated DMRs are indicated as lollipops with black and white circles, respectively. The black bar below the map indicates the location of the DNA FISH probe utilized to show the neuron-specific chromatin decondensation (probably the active maternal allele). (B) Maternal transmission of the *IG-DMR* deletion (white triangle) in mouse results in bidirectional loss of imprinting of all genes in the domain. (C) Paternal transmission of the *IG-DMR* deletion in mouse does not affect imprinting.

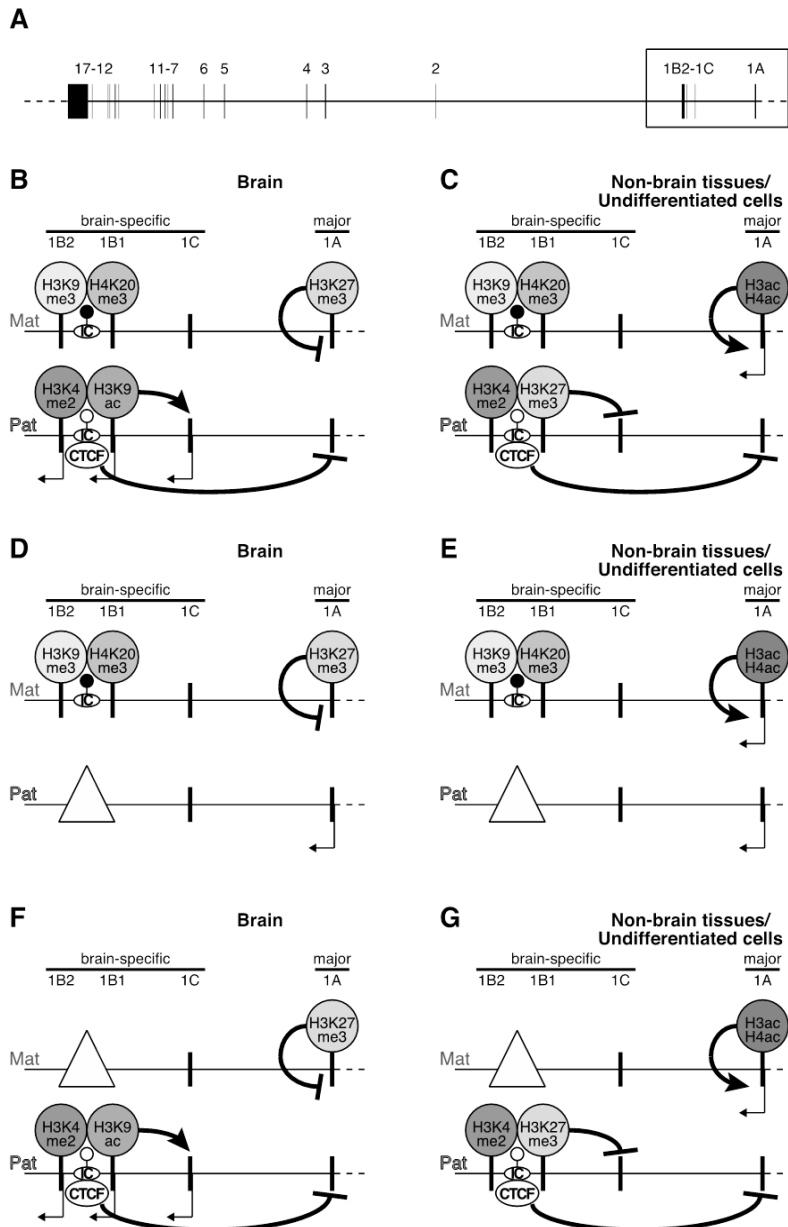


Figure 4. Overview of the *Grb10* locus. (A) Schematic representation of the region spanning mouse *Grb10* exons 1A-17. Exons are designated as vertical bars. The promoter region focused in following panels (B-G) is marked by a rectangle. (B, C) Allele-specific expression patterns and epigenetic features in (B) brain and (C) non-brain tissues and undifferentiated cells. *Grb10*-DMR and CTCF are depicted as ovals labeled IC (imprinting center) and CTCF, respectively. CTCF represses the paternal major promoter, possibly through its binding to the unmethylated paternal *Grb10*-DMR. Replacement of repressive H3K27me3 with active H3K9ac activates the paternal brain-specific promoter near the *Grb10*-DMR. In contrast, replacement of active panacetylated H3 and H4 with repressive H3K27me3 represses the maternal major promoter in neuron. (D, E) Paternal transmission of the *Grb10*-DMR deletion (white triangles) results in derepression of the major promoter in (D) brain and (E) non-brain tissues and undifferentiated cells. (F, G) Maternal transmission of the *Grb10*-DMR deletion does not affect imprinting in (F) brain or (G) non-brain tissues and undifferentiated cells.

Experiments on primary cultured cells revealed similar allele-specific histone modifications in the major promoter. The maternal major promoter is activated in non-brain tissues and repressed in brain. It is enriched with activating chromatin marks (panacetylated H3 and H4) in fibroblasts, but with repressive H3K27me3 in primary neuronal culture [112]. The repressed paternal major promoter is enriched with repressive H3K27me3 in both fibroblasts and primary neuronal cultures.

CONCLUSION

The *Dlk1-Dio3* and *Snrpn-Ube3a* domains are similar, for example, in that they are involved in transcription of the snoRNA clusters or their upstream brain-specific non-coding transcripts (*Gtl2v1* in the *Dlk1-Dio3* domain and U exon transcripts in the *Snrpn-Ube3a* domain) and that accumulation of spliced nuclear-retained host genes occurs at their transcription sites. Although *Gtl2v1* is an alternatively spliced isoform whose promoter is used for transcription in several tissues, these shared features may play a role in establishing the differential chromatin organization of the two parental alleles and brain-specific expression of the downstream imprinted genes (maternal *Mirg* in the *Dlk1-Dio3* domain and paternal *Ube3a-ATS* in the *Snrpn-Ube3a* domain). The treatment with ASOs that target *Snord116* degraded almost all of the *Snord116* and *Snord115* transcripts but approximately half of the *Ube3a-ATS* precursor transcripts [86], supporting the existence of several types of *Ube3a-ATS* transcripts. The *Snord116* transcripts are embedded in the *Ube3a-ATS* transcripts driven from U exons or exon 1, but not in the shortened type of *Ube3a-ATS* transcripts driven from the TSSs near *Snord115*. The similarities between *Dlk1-Dio3* and *Snrpn-Ube3a* domains (existence of several types of transcripts, neuron-specific chromatin decondensation, and localization of transcripts at their transcription sites) may provide a clue to the complicated gene regulation in these domains (Figure 1A, 3A). The similarity between *Grb10* transcription and U-exon transcription is the existence of brain-specific promoters (Figure 1A, 1B, 2, 4B, 4C). Although the epigenetic features of the brain-specific promoters in U exons are not yet clear, the regulation of the promoter switch in *Grb10* may provide some insight.

The role of brain-specific U exon transcripts in the brain-specific expression of *Ube3a-ATS* also remains to be clarified. The regulation of paternal *Ube3a* silencing through *Ube3a-ATS* also requires further study. Further functional experiments to investigate the regulatory regions or functional elements of U exon transcripts, snoRNA clusters and *Ube3a-ATS* will be essential to elucidate the mechanisms that regulate brain-specific imprinting. We consider that longitudinal analyses by *in vitro* neural differentiation [16, 47, 75, 113, 116] will be a useful tool to clarify the developmental changes of the imprinting marks and gene regulation of the brain-specific imprinted genes.

REFERENCES

- [1] Barlow DP, Bartolomei MS. Genomic imprinting in mammals. *Cold Spring Harbor perspectives in biology*. 2014;6(2).

- [2] Chaillet JR, Knoll JH, Horsthemke B, Lalande M. The syntenic relationship between the critical deletion region for the Prader-Willi/Angelman syndromes and proximal mouse chromosome 7. *Genomics*. 1991;11(3):773-6.
- [3] Wagstaff J, Chaillet JR, Lalande M. The GABAA receptor beta 3 subunit gene: characterization of a human cDNA from chromosome 15q11q13 and mapping to a region of conserved synteny on mouse chromosome 7. *Genomics*. 1991;11(4):1071-8.
- [4] Horsthemke B, Wagstaff J. Mechanisms of imprinting of the Prader-Willi/Angelman region. *American journal of medical genetics Part A*. 2008;146a(16):2041-52.
- [5] Cavaille J, Buiting K, Kieffmann M, Lalande M, Brannan CI, Horsthemke B, et al. Identification of brain-specific and imprinted small nucleolar RNA genes exhibiting an unusual genomic organization. *Proceedings of the National Academy of Sciences of the United States of America*. 2000;97(26):14311-6.
- [6] Cavaille J, Seitz H, Paulsen M, Ferguson-Smith AC, Bachellerie JP. Identification of tandemly-repeated C/D snoRNA genes at the imprinted human 14q32 domain reminiscent of those at the Prader-Willi/Angelman syndrome region. *Human molecular genetics*. 2002;11(13):1527-38.
- [7] Castle JC, Armour CD, Lower M, Haynor D, Biery M, Bouzek H, et al. Digital genome-wide ncRNA expression, including SnoRNAs, across 11 human tissues using polyA-neutral amplification. *PloS one*. 2010;5(7):e11779.
- [8] Rougeulle C, Cardoso C, Fontes M, Colleaux L, Lalande M. An imprinted antisense RNA overlaps UBE3A and a second maternally expressed transcript. *Nature genetics*. 1998;19(1):15-6.
- [9] Chamberlain SJ, Brannan CI. The Prader-Willi syndrome imprinting center activates the paternally expressed murine Ube3a antisense transcript but represses paternal Ube3a. *Genomics*. 2001;73(3):316-22.
- [10] Albrecht U, Sutcliffe JS, Cattanaach BM, Beechey CV, Armstrong D, Eichele G, et al. Imprinted expression of the murine Angelman syndrome gene, Ube3a, in hippocampal and Purkinje neurons. *Nature genetics*. 1997;17(1):75-8.
- [11] Rougeulle C, Glatt H, Lalande M. The Angelman syndrome candidate gene, UBE3A/E6-AP, is imprinted in brain. *Nature genetics*. 1997;17(1):14-5.
- [12] Vu TH, Hoffman AR. Imprinting of the Angelman syndrome gene, UBE3A, is restricted to brain. *Nature genetics*. 1997;17(1):12-3.
- [13] Dittrich B, Buiting K, Korn B, Rickard S, Buxton J, Saitoh S, et al. Imprint switching on human chromosome 15 may involve alternative transcripts of the SNRPN gene. *Nature genetics*. 1996;14(2):163-70.
- [14] Farber C, Dittrich B, Buiting K, Horsthemke B. The chromosome 15 imprinting centre (IC) region has undergone multiple duplication events and contains an upstream exon of SNRPN that is deleted in all Angelman syndrome patients with an IC microdeletion. *Human molecular genetics*. 1999;8(2):337-43.
- [15] Bressler J, Tsai TF, Wu MY, Tsai SF, Ramirez MA, Armstrong D, et al. The SNRPN promoter is not required for genomic imprinting of the Prader-Willi/Angelman domain in mice. *Nature genetics*. 2001;28(3):232-40.
- [16] Landers M, Bancescu DL, Le Meur E, Rougeulle C, Glatt-Deeley H, Brannan C, et al. Regulation of the large (approximately 1000 kb) imprinted murine Ube3a antisense transcript by alternative exons upstream of Snurf/Snrpn. *Nucleic acids research*. 2004;32(11):3480-92.

- [17] Kishino T, Lalande M, Wagstaff J. UBE3A/E6-AP mutations cause Angelman syndrome. *Nature genetics*. 1997;15(1):70-3.
- [18] Matsuura T, Sutcliffe JS, Fang P, Galjaard RJ, Jiang YH, Benton CS, et al. De novo truncating mutations in E6-AP ubiquitin-protein ligase gene (UBE3A) in Angelman syndrome. *Nature genetics*. 1997;15(1):74-7.
- [19] Sahoo T, del Gaudio D, German JR, Shinawi M, Peters SU, Person RE, et al. Prader-Willi phenotype caused by paternal deficiency for the HBII-85 C/D box small nucleolar RNA cluster. *Nature genetics*. 2008;40(6):719-21.
- [20] de Smith AJ, Purmann C, Walters RG, Ellis RJ, Holder SE, Van Haelst MM, et al. A deletion of the HBII-85 class of small nucleolar RNAs (snoRNAs) is associated with hyperphagia, obesity and hypogonadism. *Human molecular genetics*. 2009;18(17):3257-65.
- [21] Duker AL, Ballif BC, Bawle EV, Person RE, Mahadevan S, Alliman S, et al. Paternally inherited microdeletion at 15q11.2 confirms a significant role for the SNORD116 C/D box snoRNA cluster in Prader-Willi syndrome. *European journal of human genetics : EJHG*. 2010;18(11):1196-201.
- [22] Barlow DP, Stoger R, Herrmann BG, Saito K, Schweifer N. The mouse insulin-like growth factor type-2 receptor is imprinted and closely linked to the Tme locus. *Nature*. 1991;349(6304):84-7.
- [23] Bartolomei MS, Zemel S, Tilghman SM. Parental imprinting of the mouse H19 gene. *Nature*. 1991;351(6322):153-5.
- [24] DeChiara TM, Robertson EJ, Efstratiadis A. Parental imprinting of the mouse insulin-like growth factor II gene. *Cell*. 1991;64(4):849-59.
- [25] Ferguson-Smith AC, Cattanaach BM, Barton SC, Beechey CV, Surani MA. Embryological and molecular investigations of parental imprinting on mouse chromosome 7. *Nature*. 1991;351(6328):667-70.
- [26] Leff SE, Brannan CI, Reed ML, Ozcelik T, Francke U, Copeland NG, et al. Maternal imprinting of the mouse Snrpn gene and conserved linkage homology with the human Prader-Willi syndrome region. *Nature genetics*. 1992;2(4):259-64.
- [27] Ozcelik T, Leff S, Robinson W, Donlon T, Lalande M, Sanjines E, et al. Small nuclear ribonucleoprotein polypeptide N (SNRPN), an expressed gene in the Prader-Willi syndrome critical region. *Nature genetics*. 1992;2(4):265-9.
- [28] Schmauss C, Brines ML, Lerner MR. The gene encoding the small nuclear ribonucleoprotein-associated protein N is expressed at high levels in neurons. *The Journal of biological chemistry*. 1992;267(12):8521-9.
- [29] Sutcliffe JS, Nakao M, Christian S, Orstavik KH, Tommerup N, Ledbetter DH, et al. Deletions of a differentially methylated CpG island at the SNRPN gene define a putative imprinting control region. *Nature genetics*. 1994;8(1):52-8.
- [30] Buiting K, Saitoh S, Gross S, Dittrich B, Schwartz S, Nicholls RD, et al. Inherited microdeletions in the Angelman and Prader-Willi syndromes define an imprinting centre on human chromosome 15. *Nature genetics*. 1995;9(4):395-400.
- [31] Glenn CC, Saitoh S, Jong MT, Filbrandt MM, Surti U, Driscoll DJ, et al. Gene structure, DNA methylation, and imprinted expression of the human SNRPN gene. *American journal of human genetics*. 1996;58(2):335-46.

- [32] Gabriel JM, Gray TA, Stubbs L, Saitoh S, Ohta T, Nicholls RD. Structure and function correlations at the imprinted mouse *Snrpn* locus. *Mammalian genome : official journal of the International Mammalian Genome Society*. 1998;9(10):788-93.
- [33] Nicholls RD, Saitoh S, Horsthemke B. Imprinting in Prader-Willi and Angelman syndromes. *Trends in genetics: TIG*. 1998;14(5):194-200.
- [34] Gray TA, Saitoh S, Nicholls RD. An imprinted, mammalian bicistronic transcript encodes two independent proteins. *Proceedings of the National Academy of Sciences of the United States of America*. 1999;96(10):5616-21.
- [35] Lewis MW, Brant JO, Kramer JM, Moss JI, Yang TP, Hansen PJ, et al. Angelman syndrome imprinting center encodes a transcriptional promoter. *Proceedings of the National Academy of Sciences of the United States of America*. 2014.
- [36] Ohta T, Gray TA, Rogan PK, Buiting K, Gabriel JM, Saitoh S, et al. Imprinting-mutation mechanisms in Prader-Willi syndrome. *American journal of human genetics*. 1999;64(2):397-413.
- [37] Buiting K, Lich C, Cottrell S, Barnicoat A, Horsthemke B. A 5-kb imprinting center deletion in a family with Angelman syndrome reduces the shortest region of deletion overlap to 880 bp. *Human genetics*. 1999;105(6):665-6.
- [38] Bielinska B, Blaydes SM, Buiting K, Yang T, Krajewska-Walasek M, Horsthemke B, et al. De novo deletions of SNRPN exon 1 in early human and mouse embryos result in a paternal to maternal imprint switch. *Nature genetics*. 2000;25(1):74-8.
- [39] El-Maarri O, Buiting K, Peery EG, Kroisel PM, Balaban B, Wagner K, et al. Maternal methylation imprints on human chromosome 15 are established during or after fertilization. *Nature genetics*. 2001;27(3):341-4.
- [40] Rabinovitz S, Kaufman Y, Ludwig G, Razin A, Shemer R. Mechanisms of activation of the paternally expressed genes by the Prader-Willi imprinting center in the Prader-Willi/Angelman syndromes domains. *Proceedings of the National Academy of Sciences of the United States of America*. 2012;109(19):7403-8.
- [41] Runte M, Huttenhofer A, Gross S, Kiefmann M, Horsthemke B, Buiting K. The IC-SNURF-SNRPN transcript serves as a host for multiple small nucleolar RNA species and as an antisense RNA for UBE3A. *Human molecular genetics*. 2001;10(23):2687-700.
- [42] Galiveti CR, Raabe CA, Konthur Z, Rozhdestvensky TS. Differential regulation of non-protein coding RNAs from Prader-Willi Syndrome locus. *Scientific reports*. 2014;4:6445.
- [43] de los Santos T, Schweizer J, Rees CA, Francke U. Small evolutionarily conserved RNA, resembling C/D box small nucleolar RNA, is transcribed from PWCR1, a novel imprinted gene in the Prader-Willi deletion region, which is highly expressed in brain. *American journal of human genetics*. 2000;67(5):1067-82.
- [44] Meguro M, Mitsuya K, Nomura N, Kohda M, Kashiwagi A, Nishigaki R, et al. Large-scale evaluation of imprinting status in the Prader-Willi syndrome region: an imprinted direct repeat cluster resembling small nucleolar RNA genes. *Human molecular genetics*. 2001;10(4):383-94.
- [45] Lossie AC, Whitney MM, Amidon D, Dong HJ, Chen P, Theriaque D, et al. Distinct phenotypes distinguish the molecular classes of Angelman syndrome. *Journal of medical genetics*. 2001;38(12):834-45.

- [46] Makedonski K, Abuhatzira L, Kaufman Y, Razin A, Shemer R. MeCP2 deficiency in Rett syndrome causes epigenetic aberrations at the PWS/AS imprinting center that affects UBE3A expression. *Human molecular genetics*. 2005;14(8):1049-58.
- [47] Chamberlain SJ, Chen PF, Ng KY, Bourgois-Rocha F, Lemtiri-Chlieh F, Levine ES, et al. Induced pluripotent stem cell models of the genomic imprinting disorders Angelman and Prader-Willi syndromes. *Proceedings of the National Academy of Sciences of the United States of America*. 2010;107(41):17668-73.
- [48] Wawrzik M, Spiess AN, Herrmann R, Buiting K, Horsthemke B. Expression of SNURF-SNRPN upstream transcripts and epigenetic regulatory genes during human spermatogenesis. *European journal of human genetics : EJHG*. 2009;17(11):1463-70.
- [49] Horsthemke B, Buiting K. Imprinting defects on human chromosome 15. *Cytogenetic and genome research*. 2006;113(1-4):292-9.
- [50] Saitoh S, Wada T. Parent-of-origin specific histone acetylation and reactivation of a key imprinted gene locus in Prader-Willi syndrome. *American journal of human genetics*. 2000;66(6):1958-62.
- [51] Fulmer-Smentek SB, Francke U. Association of acetylated histones with paternally expressed genes in the Prader-Willi deletion region. *Human molecular genetics*. 2001;10(6):645-52.
- [52] Xin Z, Allis CD, Wagstaff J. Parent-specific complementary patterns of histone H3 lysine 9 and H3 lysine 4 methylation at the Prader-Willi syndrome imprinting center. *American journal of human genetics*. 2001;69(6):1389-94.
- [53] Kantor B, Shemer R, Razin A. The Prader-Willi/Angelman imprinted domain and its control center. *Cytogenetic and genome research*. 2006;113(1-4):300-5.
- [54] Kantor B, Kaufman Y, Makedonski K, Razin A, Shemer R. Establishing the epigenetic status of the Prader-Willi/Angelman imprinting center in the gametes and embryo. *Human molecular genetics*. 2004;13(22):2767-79.
- [55] Schumacher A, Buiting K, Zeschnigk M, Doerfler W, Horsthemke B. Methylation analysis of the PWS/AS region does not support an enhancer-competition model. *Nature genetics*. 1998;19(4):324-5.
- [56] Perk J, Makedonski K, Lande L, Cedar H, Razin A, Shemer R. The imprinting mechanism of the Prader-Willi/Angelman regional control center. *The EMBO journal*. 2002;21(21):5807-14.
- [57] Smith EY, Futtner CR, Chamberlain SJ, Johnstone KA, Resnick JL. Transcription is required to establish maternal imprinting at the Prader-Willi syndrome and Angelman syndrome locus. *PLoS genetics*. 2011;7(12):e1002422.
- [58] Shemer R, Birger Y, Riggs AD, Razin A. Structure of the imprinted mouse Snrpn gene and establishment of its parental-specific methylation pattern. *Proceedings of the National Academy of Sciences of the United States of America*. 1997;94(19):10267-72.
- [59] Yang T, Adamson TE, Resnick JL, Leff S, Wevrick R, Francke U, et al. A mouse model for Prader-Willi syndrome imprinting-centre mutations. *Nature genetics*. 1998;19(1):25-31.
- [60] Shemer R, Hershko AY, Perk J, Mostoslavsky R, Tsuberi B, Cedar H, et al. The imprinting box of the Prader-Willi/Angelman syndrome domain. *Nature genetics*. 2000;26(4):440-3.

- [61] Wu MY, Chen KS, Bressler J, Hou A, Tsai TF, Beaudet AL. Mouse imprinting defect mutations that model Angelman syndrome. *Genesis* (New York, NY : 2000). 2006;44(1):12-22.
- [62] Johnstone KA, DuBose AJ, Futtner CR, Elmore MD, Brannan CI, Resnick JL. A human imprinting centre demonstrates conserved acquisition but diverged maintenance of imprinting in a mouse model for Angelman syndrome imprinting defects. *Human molecular genetics*. 2006;15(3):393-404.
- [63] Szabo PE, Mann JR. Biallelic expression of imprinted genes in the mouse germ line: implications for erasure, establishment, and mechanisms of genomic imprinting. *Genes & development*. 1995;9(15):1857-68.
- [64] Szabo PE, Hubner K, Scholer H, Mann JR. Allele-specific expression of imprinted genes in mouse migratory primordial germ cells. *Mechanisms of development*. 2002;115(1-2):157-60.
- [65] Hajkova P, Erhardt S, Lane N, Haaf T, El-Maarri O, Reik W, et al. Epigenetic reprogramming in mouse primordial germ cells. *Mechanisms of development*. 2002;117(1-2):15-23.
- [66] Szabo PE, Mann JR. Allele-specific expression and total expression levels of imprinted genes during early mouse development: implications for imprinting mechanisms. *Genes & development*. 1995;9(24):3097-108.
- [67] Lucifero D, Mann MR, Bartolomei MS, Trasler JM. Gene-specific timing and epigenetic memory in oocyte imprinting. *Human molecular genetics*. 2004;13(8):839-49.
- [68] Hiura H, Obata Y, Komiyama J, Shirai M, Kono T. Oocyte growth-dependent progression of maternal imprinting in mice. *Genes to cells : devoted to molecular & cellular mechanisms*. 2006;11(4):353-61.
- [69] Wu MY, Jiang M, Zhai X, Beaudet AL, Wu RC. An unexpected function of the Prader-Willi syndrome imprinting center in maternal imprinting in mice. *PloS one*. 2012;7(4):e34348.
- [70] Le Meur E, Watrin F, Landers M, Sturny R, Lalande M, Muscatelli F. Dynamic developmental regulation of the large non-coding RNA associated with the mouse 7C imprinted chromosomal region. *Developmental biology*. 2005;286(2):587-600.
- [71] Meng L, Person RE, Beaudet AL. Ube3a-ATS is an atypical RNA polymerase II transcript that represses the paternal expression of Ube3a. *Human molecular genetics*. 2012;21(13):3001-12.
- [72] Dubose AJ, Smith EY, Yang TP, Johnstone KA, Resnick JL. A new deletion refines the boundaries of the murine Prader-Willi syndrome imprinting center. *Human molecular genetics*. 2011;20(17):3461-6.
- [73] Yamasaki K, Joh K, Ohta T, Masuzaki H, Ishimaru T, Mukai T, et al. Neurons but not glial cells show reciprocal imprinting of sense and antisense transcripts of Ube3a. *Human molecular genetics*. 2003;12(8):837-47.
- [74] Mapendano CK, Kishino T, Miyazaki K, Kondo S, Yoshiura K, Hishikawa Y, et al. Expression of the Snurf-Snrpn IC transcript in the oocyte and its putative role in the imprinting establishment of the mouse 7C imprinting domain. *Journal of human genetics*. 2006;51(3):236-43.
- [75] Martins-Taylor K, Hsiao JS, Chen PF, Glatt-Deeley H, De Smith AJ, Blakemore AI, et al. Imprinted expression of UBE3A in non-neuronal cells from a Prader-Willi syndrome patient with an atypical deletion. *Human molecular genetics*. 2014;23(9):2364-73.

- [76] Meng L, Person RE, Huang W, Zhu PJ, Costa-Mattioli M, Beaudet AL. Truncation of Ube3a-ATS unsilences paternal Ube3a and ameliorates behavioral defects in the Angelman syndrome mouse model. *PLoS genetics*. 2013;9(12):e1004039.
- [77] Leung KN, Vallero RO, DuBose AJ, Resnick JL, LaSalle JM. Imprinting regulates mammalian snoRNA-encoding chromatin decondensation and neuronal nucleolar size. *Human molecular genetics*. 2009;18(22):4227-38.
- [78] Powell WT, Coulson RL, Gonzales ML, Crary FK, Wong SS, Adams S, et al. R-loop formation at Snord116 mediates topotecan inhibition of Ube3a-antisense and allele-specific chromatin decondensation. *Proceedings of the National Academy of Sciences of the United States of America*. 2013;110(34):13938-43.
- [79] Vitali P, Royo H, Marty V, Bortolin-Cavaille ML, Cavaille J. Long nuclear-retained non-coding RNAs and allele-specific higher-order chromatin organization at imprinted snoRNA gene arrays. *Journal of cell science*. 2010;123(Pt 1):70-83.
- [80] Powell WT, Coulson RL, Crary FK, Wong SS, Ach RA, Tsang P, et al. A Prader-Willi locus lncRNA cloud modulates diurnal genes and energy expenditure. *Human molecular genetics*. 2013;22(21):4318-28.
- [81] Huang HS, Allen JA, Mabb AM, King IF, Miriyala J, Taylor-Blake B, et al. Topoisomerase inhibitors unsilence the dormant allele of Ube3a in neurons. *Nature*. 2012;481(7380):185-9.
- [82] Belotserkovskii BP, Liu R, Tornaletti S, Krasilnikova MM, Mirkin SM, Hanawalt PC. Mechanisms and implications of transcription blockage by guanine-rich DNA sequences. *Proceedings of the National Academy of Sciences of the United States of America*. 2010;107(29):12816-21.
- [83] El Hage A, French SL, Beyer AL, Tollervey D. Loss of Topoisomerase I leads to R-loop-mediated transcriptional blocks during ribosomal RNA synthesis. *Genes & development*. 2010;24(14):1546-58.
- [84] Skourti-Stathaki K, Proudfoot NJ, Gromak N. Human senataxin resolves RNA/DNA hybrids formed at transcriptional pause sites to promote Xrn2-dependent termination. *Molecular cell*. 2011;42(6):794-805.
- [85] Numata K, Kohama C, Abe K, Kiyosawa H. Highly parallel SNP genotyping reveals high-resolution landscape of mono-allelic Ube3a expression associated with locus-wide antisense transcription. *Nucleic acids research*. 2011;39(7):2649-57.
- [86] Meng L, Ward AJ, Chun S, Bennett CF, Beaudet AL, Rigo F. Towards a therapy for Angelman syndrome by targeting a long non-coding RNA. *Nature*. 2015;518(7539):409-12.
- [87] Hatada I, Morita S, Obata Y, Sotomaru Y, Shimoda M, Kono T. Identification of a new imprinted gene, Rian, on mouse chromosome 12 by fluorescent differential display screening. *Journal of biochemistry*. 2001;130(2):187-90.
- [88] Irving MD, Buiting K, Kanber D, Donaghue C, Schulz R, Offiah A, et al. Segmental paternal uniparental disomy (patUPD) of 14q32 with abnormal methylation elicits the characteristic features of complete patUPD14. *American journal of medical genetics Part A*. 2010;152a(8):1942-50.
- [89] Benetatos L, Hatzimichael E, Londin E, Vartholomatos G, Loher P, Rigoutsos I, et al. The microRNAs within the DLK1-DIO3 genomic region: involvement in disease pathogenesis. *Cellular and molecular life sciences : CMLS*. 2013;70(5):795-814.

- [90] Seitz H, Youngson N, Lin SP, Dalbert S, Paulsen M, Bachellerie JP, et al. Imprinted microRNA genes transcribed antisense to a reciprocally imprinted retrotransposon-like gene. *Nature genetics*. 2003;34(3):261-2.
- [91] Seitz H, Royo H, Bortolin ML, Lin SP, Ferguson-Smith AC, Cavaille J. A large imprinted microRNA gene cluster at the mouse Dlk1-Gtl2 domain. *Genome research*. 2004;14(9):1741-8.
- [92] Davis E, Caiment F, Tordoir X, Cavaille J, Ferguson-Smith A, Cockett N, et al. RNAi-mediated allelic trans-interaction at the imprinted Rtl1/Peg11 locus. *Current biology: CB*. 2005;15(8):743-9.
- [93] Tierling S, Dalbert S, Schoppenhorst S, Tsai CE, Oliger S, Ferguson-Smith AC, et al. High-resolution map and imprinting analysis of the Gtl2-Dnchc1 domain on mouse chromosome 12. *Genomics*. 2006;87(2):225-35.
- [94] Takada S, Paulsen M, Tevendale M, Tsai CE, Kelsey G, Cattanach BM, et al. Epigenetic analysis of the Dlk1-Gtl2 imprinted domain on mouse chromosome 12: implications for imprinting control from comparison with Igf2-H19. *Human molecular genetics*. 2002;11(1):77-86.
- [95] Lin SP, Youngson N, Takada S, Seitz H, Reik W, Paulsen M, et al. Asymmetric regulation of imprinting on the maternal and paternal chromosomes at the Dlk1-Gtl2 imprinted cluster on mouse chromosome 12. *Nature genetics*. 2003;35(1):97-102.
- [96] Lin SP, Coan P, da Rocha ST, Seitz H, Cavaille J, Teng PW, et al. Differential regulation of imprinting in the murine embryo and placenta by the Dlk1-Dio3 imprinting control region. *Development (Cambridge, England)*. 2007;134(2):417-26.
- [97] Kagami M, Sekita Y, Nishimura G, Irie M, Kato F, Okada M, et al. Deletions and epimutations affecting the human 14q32.2 imprinted region in individuals with paternal and maternal upd(14)-like phenotypes. *Nature genetics*. 2008;40(2):237-42.
- [98] Schuster-Gossler K, Simon-Chazottes D, Guenet JL, Zachgo J, Gossler A. Gtl2lacZ, an insertional mutation on mouse chromosome 12 with parental origin-dependent phenotype. *Mammalian genome : official journal of the International Mammalian Genome Society*. 1996;7(1):20-4.
- [99] Schuster-Gossler K, Bilinski P, Sado T, Ferguson-Smith A, Gossler A. The mouse Gtl2 gene is differentially expressed during embryonic development, encodes multiple alternatively spliced transcripts, and may act as an RNA. *Developmental dynamics : an official publication of the American Association of Anatomists*. 1998;212(2):214-28.
- [100] Paulsen M, Takada S, Youngson NA, Benchaib M, Charlier C, Segers K, et al. Comparative sequence analysis of the imprinted Dlk1-Gtl2 locus in three mammalian species reveals highly conserved genomic elements and refines comparison with the Igf2-H19 region. *Genome research*. 2001;11(12):2085-94.
- [101] Qu C, Jiang T, Li Y, Wang X, Cao H, Xu H, et al. Gene expression and IG-DMR hypomethylation of maternally expressed gene 3 in developing corticospinal neurons. *Gene expression patterns : GEP*. 2013;13(1-2):51-6.
- [102] Kaneko S, Bonasio R, Saldana-Meyer R, Yoshida T, Son J, Nishino K, et al. Interactions between JARID2 and noncoding RNAs regulate PRC2 recruitment to chromatin. *Molecular cell*. 2014;53(2):290-300.
- [103] Dong LQ, Du H, Porter SG, Kolakowski LF, Jr., Lee AV, Mandarino LJ, et al. Cloning, chromosome localization, expression, and characterization of an Src homology 2 and

- pleckstrin homology domain-containing insulin receptor binding protein hGrb10gamma. *The Journal of biological chemistry*. 1997;272(46):29104-12.
- [104] Jerome CA, Scherer SW, Tsui LC, Gietz RD, Triggs-Raine B. Assignment of growth factor receptor-bound protein 10 (GRB10) to human chromosome 7p11.2-p12. *Genomics*. 1997;40(1):215-6.
- [105] Miyoshi N, Kuroiwa Y, Kohda T, Shitara H, Yonekawa H, Kawabe T, et al. Identification of the Meg1/Grb10 imprinted gene on mouse proximal chromosome 11, a candidate for the Silver-Russell syndrome gene. *Proceedings of the National Academy of Sciences of the United States of America*. 1998;95(3):1102-7.
- [106] Han DC, Shen TL, Guan JL. The Grb7 family proteins: structure, interactions with other signaling molecules and potential cellular functions. *Oncogene*. 2001;20(44):6315-21.
- [107] Blagitko N, Mergenthaler S, Schulz U, Wollmann HA, Craigen W, Eggermann T, et al. Human GRB10 is imprinted and expressed from the paternal and maternal allele in a highly tissue- and isoform-specific fashion. *Human molecular genetics*. 2000;9(11):1587-95.
- [108] Hitchins MP, Monk D, Bell GM, Ali Z, Preece MA, Stanier P, et al. Maternal repression of the human GRB10 gene in the developing central nervous system; evaluation of the role for GRB10 in Silver-Russell syndrome. *European journal of human genetics : EJHG*. 2001;9(2):82-90.
- [109] Hikichi T, Kohda T, Kaneko-Ishino T, Ishino F. Imprinting regulation of the murine Meg1/Grb10 and human GRB10 genes; roles of brain-specific promoters and mouse-specific CTCF-binding sites. *Nucleic acids research*. 2003;31(5):1398-406.
- [110] Garfield AS, Cowley M, Smith FM, Moorwood K, Stewart-Cox JE, Gilroy K, et al. Distinct physiological and behavioural functions for parental alleles of imprinted Grb10. *Nature*. 2011;469(7331):534-8.
- [111] Arnaud P, Monk D, Hitchins M, Gordon E, Dean W, Beechey CV, et al. Conserved methylation imprints in the human and mouse GRB10 genes with divergent allelic expression suggests differential reading of the same mark. *Human molecular genetics*. 2003;12(9):1005-19.
- [112] Yamasaki-Ishizaki Y, Kayashima T, Mapendano CK, Soejima H, Ohta T, Masuzaki H, et al. Role of DNA methylation and histone H3 lysine 27 methylation in tissue-specific imprinting of mouse Grb10. *Molecular and cellular biology*. 2007;27(2):732-42.
- [113] Sanz LA, Chamberlain S, Sabourin JC, Henckel A, Magnuson T, Hugnot JP, et al. A mono-allelic bivalent chromatin domain controls tissue-specific imprinting at Grb10. *The EMBO journal*. 2008;27(19):2523-32.
- [114] Hitchins MP, Bentley L, Monk D, Beechey C, Peters J, Kelsey G, et al. DDC and COBL, flanking the imprinted GRB10 gene on 7p12, are biallelically expressed. *Mammalian genome : official journal of the International Mammalian Genome Society*. 2002;13(12):686-91.
- [115] Plasschaert RN, Bartolomei MS. Tissue-specific regulation and function of Grb10 during growth and neuronal commitment. *Proceedings of the National Academy of Sciences of the United States of America*. 2014.
- [116] Kohama C, Kato H, Numata K, Hirose M, Takemasa T, Ogura A, et al. ES cell differentiation system recapitulates the establishment of imprinted gene expression in a cell-type-specific manner. *Human molecular genetics*. 2012;21(6):1391-401.

Chapter 33

PHARMACOGENOMICS FOCUSING ON PHASE TWO METABOLIZING ENZYMES

Kung-Hao Liang*

Chang Gung Memorial Hospital, Taoyuan City, Taiwan

ABSTRACT

Treatment of human diseases in general has dual goals: to treat effectively, at the cost of minimal adverse side effects. However, person-to-person differences on drug efficacy and adverse reactions have been frequently observed. Pharmacogenomic studies intend to address such differences based on personal genomic variants such as single nucleotide polymorphisms. In the past, success has been achieved on the identification of major histocompatibility complex genomic variants which are associated to immune-mediated adverse drug effects. Additionally, genomic variants on the direct drug targets have been found to be associated to drug efficacy. Besides immunological and drug-target factors, one other critical factor is on the drug metabolism mediated by various xenobiotic metabolizing and detoxification enzymes. They determine how quickly the drugs can be metabolized and then excreted from the body, thereby affecting drug efficacy and toxicity.

Human xenobiotic metabolizing enzymes are categorized by their phases in the metabolizing processes: the modification phase (phase 1), the conjugation phase (phase 2), and further modification and excretion (phase 3). Previously, genes involved in phase 1 have been extensively studied in pharmacogenomics. In comparison, genes in phase 2 received less attention, despite the fact that they are no less important. These genes include Uridine Diphospho-Glucuronosyltransferases and others. This Chapter presents the genomic variants on phase 2 genes which can account for the drug efficacy and adverse reactions.

* Corresponding Author's Email: kunghao@gmail.com (Tel: + 886-3-3281200 ext 8129).

INTRODUCTION

The human genome encodes the blueprint of the human body. It is the shared human heritage yet every person has a slightly different version of the genome. The variations between versions often occur in hotspots, such as the single nucleotide polymorphisms or short tandem repeats, which could cause personal difference in health and disease. All of us suffer from diseases sometime in our life, either mild or severe. Contemporary treatments of diseases often follow medical guidelines; and the drugs are prescribed by their optimum doses established by an extensive series of preclinical lab tests, large-scale clinical trials and post-market studies. Under such standalized treatments, different levels of responses and toxicity are often observed in different people. Importantly, life-threatening adverse reactions may occur to certain patients for unknown reasons. Patients' age, gender and ethnicity cannot explain the observed variations completely. The remaining unexplained variations could be ascribed to the personal genetic background.

Pharmacogenomics studies aim to elucidate the unexplained variations by interrogating the genotype-efficacy/safety relationships. The candidate gene approach interrogates a focused panel of genes which are hypothesized to underlie the investigated variation. Alternatively, the genome-wide approach interrogates the entire genome, using multiplexing platforms such as SNP microarrays or high-throughput sequencing techniques. Expected results are genes where different genotypes associated to different outcomes of the same treatment (efficacious vs. non-efficacious, or with adverse effects vs. non-adverse effects).

Till 2014, ascertained genes often belong to one of three typical categories. The first class is involved in the targeting mechanisms of drugs. For example, personal variations of the promoter sequence of the IL28B gene, also known as Interferon lambda, was found to account for the difference on the efficacy of interferon-based treatments of chronic hepatitis C infection (Ge et al., 2009). IL28B was considered the downstream genes in the interferon pathways.

The second class is the immunological genes. Variants on major histocompatibility complex genes can be used to identify high-risk subjects of life-threatening allergic reactions such as the Steven-Johnson's syndrome, if they were given Carbamazepine which is an anticonvulsant and a mood-stabilizing drug (Chen et al., 2011). A prospective study with 4877 subjects in Taiwan showed that conducting the genetic test before the use of Carbamazepine can achieve zero incidence of Steven-Johnson's syndrome, as opposed to a historical incidence of 0.23% (Chen et al., 2011).

The third class genes encode xenobiotic metabolizing enzymes. Some drugs can be excreted in its original form without being chemically modified by the body. Others undergo a series of modifications mediated by enzymes in the liver, kidney and gastrointestinal tract. Such modification include three phases, the modification phase (phase 1), the conjugation phase (phase 2), and further modification and excretion (phase 3) (Omiecinski et al., 2011). Not surprisingly, genomic variations affect the enzymatic activities, thereby affecting the effective dose concentrations of the drugs. Phase I enzymes often involve the oxydation, reduction and hydrolysis of the substances to produce polar, hydrophilic metabolites. Typical phase 1 enzymes, such as CYP2C9 or CYP4A6 in the cytochrome P450 gene family, are the metabolizers of warfarin. The genetic variants can be utilized to adjust the administration dosage of warferin (The international Warfarin pharmacogenetics consortium, 2009).

Phase 2 enzymes can increase the polarity and water solubility of substrates for their ease of excretion by bile or urine (Omiecinski et al., 2011). Substrates are modified by the covalent attachment of small, endogeneous, hydrophilic molecules including glucuronic acid, sulfate, or glycine. Different modifications are catalyzed by different groups of enzymes.

UGT GENE SUPER FAMILY ENCODES CRITICAL PHASE TWO METABOLIZING ENZYMES

The Uridine-Diphospho Glucuronosyltransferases (UGT) gene super family encodes enzymes whose major roles are to catalyze glucuronidations of substrates, be them xenobiotic drugs, endogeneous hormones or bilirubin. The molecular weights of these enzymes are around 60 kilodaltons. An UGT protein has a transmembrane domain near the C-terminus, a substrate-binding domain near the N-terminus, and an UDPGA domain in the middle for catalyzing glucuronide conjugations. The transmembrane domain anchors the protein into the golgi apparatus of the cells. The human UGT super family comprises two families, UGT1 and UGT2. The UGT1 family has 9 genes, all of them located in the forward strand of chromosome 2q37.1. They have distinct first exons but share exons 2-5. Table 1 summarizes the currently known UGT1 variants. Some variants can encode changes of amino acids in exon 1 which may confer variations on substrate specificity and binding efficacy. Other variants may play regulatory roles by affecting the transcription and translation of genes. Note that variants near the shared exons 2-5 may affect all UGT1 genes.

Table 1. Genomic variants of the UGT 1 genes sorted by their locations in human chromosome 2. Variants may encode amino acid changes which therefore affect the enzymatic functions or alternatively, they may encode the difference of regulatory effects

Gene	Chr 2 Location	dbSNP ID	Amino acid variation	Note
UGT1A8	234 526 666	rs1126788		
	234 526 668	rs1126790		
	234 526 680	rs1126792		
	234 526 681	rs1126793		
	234 526 705	rs1126798		
	234 527 001	rs1126804		
	234 527 183	rs17863762		
UGT1A10	234 545 345	rs56935833	I59M	
	234 545 765	rs45523834		
	234 545 773	rs58704432	T202I	
UGT1A9	234 580 463	rs3832043		
	234 580 589	rs72551329	C3Y	
	234 580 678	rs72551330	M33T	

Table 1. (Continued)

Gene	Chr 2 Location	dbSNP ID	Amino acid variation	Note
UGT1A7	234 581 306	rs66915469	Y242X	
	234 581 346	rs58597806	N256D	
	234 590 926	rs61261057	G115S	
UGT1A6	234 602 191	rs2070959		
	234 602 277	rs17863783		
	234 621 922	rs17874942		
UGT1A4	234 622 061	rs3755320		
	234 627 536	rs6755571		
	234 637 789	rs28898617		
UGT1A3	234 637 853	rs6706232		
	234 637 905	rs45625338		
	234 665 659	rs4124874		
UGT1A1	234 665 782	rs10929302		
	234 667 582	rs3755319		
	234 668 570	rs887829		
	234 668 881	rs8175347		Short tandem repeat (TA)5/6/7/8
	234 669 144	rs4148323		
	234 669 155	rs72551340	Y74X	
	234 669 457	rs72551341	Q175L	
	234 669 462	rs72551342	R177C	
	234 669 558	rs72551343	R209W	
	234 669 631	rs72551344	L233R	
	234 669 759	rs72551345	R276G	
	234 675 738	rs62625011	E308G	
UGT1 Shared Exons	234 675 807	rs72551348	Q331R	
	234 676 519	rs72551349	R341X	
	234 676 567	rs72551350	Q357X	
	234 676 568	rs72551351	Q357R	
	234 676 880	rs55750087	R367G	
	234 676 883	rs72551352	T368A	
	234 676 905	rs72551353	S375F	
	234 676 924	rs72551354	S381R	
	234 676 982	rs72551355	P401A	
	234 680 912	rs72551357	K437X	
	234 681 090	rs72551361	L496X	
	234 681 416	rs10929303		
	234 681 544	rs1042640		
	234 681 645	rs8330		

The UGT2 family can be further subdivided to UGT2A and UGT2B subfamilies based on the disparity of amino acid sequences. The UGT2B genes include 7 members, where UGT2B10, 2B7 and 2B28 reside in the forward strand, while UGT2B17, 2B15, 2B11 and 2B4 reside in the reverse strand of the chromosome 4q13.2 regions. Table 2 summarizes the currently known UGT2 variants.

Table 2. Genomic variants of the UGT2 genes sorted by their locations in human chromosome 4. Variants may encode amino acid changes which therefore affect the enzymatic functions or alternatively, they may encode the difference of regulatory effects. Genes are encoded either in the forward or the reverse strand of the chromosome

Gene	Chromosomal orientation	Chr 4 location	dbSNP ID	Amino acid variation*
UGT2B17	reverse	69 415 555	rs7436962	
		69 417 570	rs28374627	
		69 420 232	rs4860305	
UGT2B15	reverse	69 512 637	rs4148271	
		69 512 655	rs3100	
		69 512 847	rs4148269	
		69 536 084	rs1902023	
UGT2B7	forward	69 961 912	rs7662029	
		69 962 078	rs7668258	
		69 962 449	rs12233719	A71S
		69 964 271	rs28365062	T245T
		69 964 337	rs7438284	P267P
		69 964 338	rs7439366	Y268H
		69 972 952	rs4348159	Y354Y
UGT2B11	reverse	70 078 295	rs3890590	P289L
		70 079 975	rs7697037	C156R
UGT2B28	forward	70 160 259	rs41292949	N441S
		70 160 299	rs41292951	V454V
		70 160 309	rs6828191	H458D
UGT2B4	reverse	70 345 904	rs1131878	
		70 346 057	rs1051752	
		70 346 127	rs1966151	
		70 346 564	rs13142440	R459R
		70 346 565	rs13119049	D458E
		70 351 050	rs72552707	L396F
		70 355 211	rs1845555	T316T

* Numbers represent the amino acid locations of the variants. Since variants are sorted in the ascending order by the chromosomal orientation, the amino acid locations appear in a descending order in those genes reside in the reverse strand.

Drugs metabolized by UGT enzymes are only partially known (Guillemette et al., 2014; Stingl et al., 2014). Commonly prescribed drugs and endogenous molecules confirmed to be metabolized by UGT enzymes are summarized in Figure 1 for a snapshot of the current

knowledge. Genes are organized by a hierarchical clustering method based on the pairwise correlation of their substrate profiles. UGT2B28, 2B11 and 1A5 have no clearly confirmed substrates currently. UGT1A4, 2B15, 2B7, 1A9 and 1A1 has been discovered to be able to metabolize multiple drugs. UGT1A7 and 1A10 have high correlation in their substrate profiles. In fact, they can only metabolize SN-38 but not the other molecules in the current list. UGT1A1 and 1A3, as well as UGT1A6 and 2B4 are also pairs of enzymes with similar substrate profiles (Figure 1).

SN-38 is the active-form metabolite of Irinotecan, which is an approved chemotherapeutic agent for treating colon cancers. SN-38 can be further metabolized by multiple enzymes including UGT1A1, 1A7, 1A9 and 1A10 to become capable of being excreted from the human body (Figure 1). Since SN-38 has a narrow therapeutic index, poor phase 2 metabolisms by the UGT genes may result in the accumulation of SN-38, causing adverse effects including diarrhea and neutropenia (Iyer et al., 2002). Thus, personal variants of the four UGT genes may account for the adverse effects. The most commonly studied variant is rs8175347, a short tandem repeat with 5-8 repetitive "TA" nucleotides in the UGT1A1 promoter region. The higher number of "TA" has been shown to be associated to poor phase 2 metabolism of SN-38 by UGT1A1 (Iyer et al., 2002).

Sorafenib is a multikinase inhibitor approved for the treatment of advanced hepatocellular carcinoma and renal cell carcinoma. It was reported to be a potent inhibitor of UGT1A1, while at the same time metabolized by UGT1A9. We can therefore infer that combined usage of sorafenib and irinotecan may cause adverse effects due to the suppression of UGT1A1 by sorafenib, which suppressed the metabolism of SN-38.

Other examples of one-to-many drug-UGT relationships include mycophenolic acids, which are immunosuppressants metabolized by UGT1A8, 1A9 and 2B7 (Picard et al., 2005), as well as olanzapine, which is an antipsychotic metabolized by UGT1A4 and 2B10 (Figure 1). Tamoxifen, an antagonist of estrogen receptors for treating estrogen-positive breast cancers, is metabolized by UGT1A4 and 2B15 (Figure 1). Resistance to the treatment, as well as adverse effects on deep-vein thrombosis and hot flashes, has been observed in some patients.

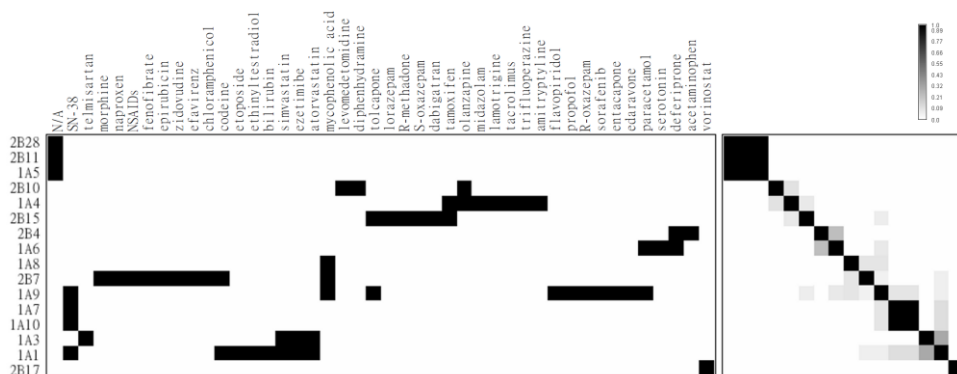


Figure 1. UGT enzymes arranged by the correlations of their substrate profiles. The right panel shows the pairwise correlations of substrate profiles by grey-scale levels. The left panel is an enzyme-substrate relationship matrix where a dark cell represent a well-established relationship summarized in a review paper (Guillemette et al., 2014). UGT enzymes with higher correlations of the substrate profiles are presented in nearby rows by a hierarchical clustering method of the Generalized Association Plot software (Chen, 2002).

FUTURE PERSPECTIVE

The above analysis shows that genomic variations on the UGT genes can certainly affect drug metabolism, causing differences on drug efficacy and safety. However, the relationship may be complex, as a drug can be metabolized by multiple UGT genes (Figure 1). Furthermore, these genes have many common variations, each of which may exert certain degree of effect on the drugs. Fortunately, multiplexing technique is already commercially available for assessing a collection of variants in a single assay. Genotypes of most of the variants in Tables 1 and 2 can be easily detected by the Affymetrix DMETTM Plus microarrays. With such information available, treatment can be tailored personally for maximal efficacy and minimal toxicity.

New drugs are constantly being developed to fulfill two major goals: to treat effectively and to elicit minimal adverse effects, both has been primary concerns of healthcare authorities. Regarding safety, how efficient the drugs can be metabolized, and whether the medication will elicit unwanted or unexpected adverse effects are two major issues. Although pharmacokinetic studies are routinely performed in pre-clinical studies to identify their metabolizing enzymes, the genotypes are rarely used currently for the design of clinical trials in selecting adequate patient groups. This may be one major reason for the high attrition rate of drug developments at this moment. We will witness soon that the genomic variants being effectively incorporated for the successful development of drugs.

REFERENCES

- Chen, C. H. (2002), Generalized Association Plots: Information Visualization via Iteratively Generated Correlation Matrices. *Statistica Sinica* 12, 7-29.
- Chen P., Lin J. J., Lu C. S. et al., Carbamazepine-induced toxic effects and HLA-B*1502 screening in Taiwan. *N Engl J Med*. 2011 Mar 24; 364(12): 1126-33.
- Ge et al. (2009), Genetic variation in IL28B predicts hepatitis C treatment-induced viral clearance. *Nature* 461, 399-401.
- Guillemette C., Lévesque É., Rouleau M. Pharmacogenomics of human uridine diphosphoglucuronosyltransferases and clinical implications. *Clin Pharmacol Ther*. 2014 Sep; 96(3):324-39.
- International Warfarin Pharmacogenetics Consortium, Klein T. E., Altman R. B., Eriksson N., Gage B. F., Kimmel S. E., Lee M. T., Limdi N. A., Page D., Roden D. M., Wagner M. J., Caldwell M. D., Johnson J. A. Estimation of the warfarin dose with clinical and pharmacogenetic data. *N Engl J Med*. 2009 Feb 19; 360(8):753-64.
- Iyer L., Das S., Janisch L., Wen M., Ramírez J., Karrison T., Fleming G. F., Vokes E. E., Schilsky R. L., Ratain M. J. UGT1A1*28 polymorphism as a determinant of irinotecan disposition and toxicity. *Pharmacogenomics J*. 2002; 2(1):43-7.
- Omicinski C. J., Vanden Heuvel J. P., Perdew G. H., Peters J. M. Xenobiotic metabolism, disposition, and regulation by receptors: from biochemical phenomenon to predictors of major toxicities. *Toxicol Sci*. 2011 Mar; 120 Suppl 1:S49-75.

- Picard N., Ratanasavanh D., Prémaud A., Le Meur Y., Marquet P. Identification of the UDP-glucuronosyltransferase isoforms involved in mycophenolic acid phase II metabolism. *Drug Metab Dispos.* 2005 Jan; 33(1):139-46.
- Stingl J. C., Bartels H., Viviani R., Lehmann M. L., Brockmöller J. Relevance of UDP-glucuronosyltransferase polymorphisms for drug dosing: A quantitative systematic review. *Pharmacol Ther.* 2014 Jan; 141(1):92-116.

Chapter 34

SEXUAL DIMORPHISM IN INSECT LONGEVITY: INSIGHTS FROM EXPERIMENTAL EVOLUTION

Jelica Lazarević^{1,*}, Biljana Stojković² and Nikola Tucić¹

¹Institute for Biological Research, University of Belgrade, Belgrade, Serbia

²Faculty of Biology, University of Belgrade, Belgrade, Serbia

ABSTRACT

In species with separate sexes, gender differences in longevity are widespread and the extent and direction of these differences varies tremendously among taxa. To understand sexual dimorphism in longevity and explain how different forms of selection shape longevity and other fitness-related traits within and among species, it is important to obtain information on the genetic architecture (the number of genes and degree of inter- and intra-genic interactions) and various mechanistic causes which underlie mortality variation between the sexes. Here we review recent empirical studies on gender differences in longevity in insect species, from both mechanistic and evolutionary perspective. Whenever it was possible, we focus on data obtained from the laboratory evolution experiments because the study of evolution under controlled conditions may provide valuable evidence not only for the effects of natural and sexual selection in shaping sex-specific longevities and mortality rates but also it offers novel insights into the mechanistic basis of these differences.

INTRODUCTION

Over the last 30 years our understanding of the proximate (mechanistic) causes involved in age-specific decline in survival and reproductive performance (senescence or ageing) has been improved dramatically. However, many fundamental questions concerning the relative importance of evolutionary mechanisms (i.e., ultimate causes: Mayr 1961) that directly or indirectly influence the rate of senescence are not entirely answered. For example, despite the

* Corresponding Author's Email: jellaz@ibiss.bg.ac.rs (Institute for Biological Research, University of Belgrade, Despot Stefan Blvd. 142, 11000 Belgrade, Serbia).

fact that we now have a plethora of information with respect to longevity differences in genetic architecture and physiology between females and males (see below), ultimate causes why females and males age at different rates are still poorly understood.

Gender differences in longevity are widespread in insect species although the extent and direction of the difference depend on species phylogenetic position, environmental conditions and mating status (Table 1). Moreover, even when females and males have the same longevity, they may differ in demographic parameters, i.e., the shape of age-specific survival curves.

Antagonistic relationship between somatic maintenance and reproduction has been revealed by many longevity-extending genetic and environmental manipulations, as well as in artificial selection and laboratory evolution experiments (Hughes and Raynolds 2005). The cost of reproduction in many insects is frequently manifested as increased mortality rate of mated relative to unmated females. Mating is energetically costly to females not only because of high amount of resources allocated to egg production but also due to impairment of immune functions (Rolff and Siva-Jothy 2002; Fedorka and Zuk 2007), oxidative damage (Archer et al. 2012a) and direct physical harassment by males (Morrow and Arnquist 2003). In general, mated individuals live shorter than virgins, although many exceptions from this ‘rule’ (Lazarević 1994; Carey et al. 2002; De Loof 2011) imply that mating costs and benefits for female longevity may vary among different populations (Shuker et al. 2006) and species (Rönn et al. 2006). For example, multiple mating is beneficial for *Callosobruchus maculatus* (Rönn et al. 2006) and *Gryllus lineaticeps* (Wagner et al. 2001), detrimental for *C. chinensis* (Rönn et al. 2006), *Drosophila buzzati* (Scannapieco et al. 2007, 2009), *Acanthoscelides obtectus* (Šešlija et al. 2008), *Gryllus bimaculatus* (Bateman et al. 2006), and does not affect female longevity in *Telostylinus angusticollis* (Adler et al. 2013) and *Gryllodes sigillatus* (Ivy and Sakaluk 2005).

Table 1. Sexual dimorphism in insect longevity and demographic parameters depending on mating status

Species ¹	Mating ²	Food ³	Line, population	Females vs. Males				Reference
				Best fit model ⁴	a	b	s (and/or Mean M) longevity	
<i>Gryllodes sigillatus</i> (O)	SM-I	+	8 inbred lines	G	<	=	<	Archer et al. 2012b
<i>Teleogryllus commodus</i> (O)	V	+	Up (selection for ↑ male longevity)	G	=	=	>	Hunt et al. 2006
	V	+	Down (↓ male longevity)	G	=	<	>	
	MM	+	Seminatural cond.	GM	<	>	>	Zajitschek et al. 2009a
	MM♀, V♂	+		G	=	>	<	Zajitschek et al. 2009b
	MM♀, V♂	+/-		G	<	>	<	
<i>Ceratitis capitata</i> (D)	V-Gr	+		L	>	<	=	Pujol-Lereis et al. 2012
<i>Drosophila melanogaster</i> (D)	V-Gr	+	Oregon R	G	<	=	<	Vaiserman et al. 2013
	V-Gr	+	Canton S	G	<	=	=	
	V-Gr	+	Um	G	<	>	<	

Species ¹	Mating ²	Food ³	Line, population	Females vs. Males					Reference
				Best fit model ⁴	a	b	s (and/or M)	Mean longevity	
	MM	+	FTF x Crete (Interpopulation crosses)	LM	<	>	> (<)	<	Rand et al. 2006
	MM	+	Crete x FTF (Interpopulation crosses)	LM	=	=	< (=)	<	
	V-I	+	Canton S	G	<	=		=	Iliadi et al. 2009
	SM-I	+		G	>	=		<	
	V-Gr	+		G	=	=		=	
	SM-Gr	+		G	=	=		<	
	MM	+		G	=	=		=	
	V-I	+	Oregon R	G	<	=		>	
	SM-I	+		G	=	=		=	
	V-Gr	+		G	<	=		>	
	SM-Gr	+		G	=	=		=	
	MM	+		G	=	=		=	
	V-Gr	+	Wt, peach orchard	GM	=	=	=	=	Promislow and Haselkorn 2002
<i>Drosophila affinis</i>	V-Gr	+	Wt, peach orchard	GM	<	=	>	<	
<i>Drosophila hydei</i>	V-Gr	+	<i>Drosophila</i> stock centre	GM	<	>	>	<	
<i>Drosophila simulans</i>	V-Gr	+	Wt, peach orchard	GM	<	>	>	=	
<i>Drosophila virilis</i>	V-Gr	+	<i>Drosophila</i> stock centre	GM	<	<	=	>	
<i>Drosophila koepferae</i>	V-Gr	+	Lab pop. (7gen.)	G	=	<		>	Scannapieco et al. 2007
<i>Drosophila buzzati</i>	V-Gr	+	Lab pop. (7gen.)	G	>	<		>	
	MM	+	control	G	=(<)>			<	Scannapieco et al. 2009
	MM	+	L (selection for ↑ lifespan) 30gen.	G	>	=		<	
	MM	+	Lab pop. (7gen.)	G	>	<		<	
	MM	+	Low latitude (natural population)	G	>	<		<	
	MM	+	High-latitude (natural population)	G	>	<		<	
<i>Drosophila mojavensis</i>	V-Gr		Agria mainland Mexico (natural population)	G♀, GM♂	>	<		=	Jaureguy and Etges 2007
	V-Gr	Lab food	„	L♀, LM♂	>	<	<	>	
<i>Acanthoscelides obtectus</i> (C)	V-I	-	E (early reproduction)	G	<	>		<	Stojković and Savković 2011
	V-I	-	L (late reproduction)	G	=	=		>	
	SM-I	-	E (early reproduction)	G	>	<		=	
	SM-I	-	L (late reproduction)	G	>	<		>	
	V-Gr		E (early reproduction)	G	=	<		>	Lazarević et al. 2013
<i>Callosobruchus maculatus</i> (C)	V-Gr		L (late reproduction)	G	<	<		>	
<i>Sator limbatus</i> (C)	V-I	-	Vigna radiata pop	L	>	<	<	>	Fox et al. 2003
<i>Pararge aegeria</i> (Lep)	V-I	-		L	<	=	=	=	
	V-Gr	+	Madeira pop.	L	<	>	=	=	Gotthard et al. 2000
	V-Gr	+	Sweden pop.	L	>	<	<	>	

¹O-Orthoptera, D-Diptera, C-Coleoptera, Lep-Lepidoptera²Mating: I-individually kept; Gr-kept in group V-virgin; SM- single mated, MM- multiple mating³Adult food: + (*ad libitum*), - (without food), +/- dietary restriction⁴L-logistic, G – Gompertz, GM - Gompertz-Makeham, LM – Logistic-Makeham

Importance of genetic background in shaping insect mortality rates and sexual dimorphism in longevity was confirmed in many studies in which life-span of unmated (virgin) individuals were compared among different natural populations or laboratory lines (Yilmaz et al. 2008; Iliadi et al. 2009) and species of *Drosophila* (Promislow and Haselkorn 2002). *D. hydei* females live about 20% shorter than males, while life-span of *D. virilis* females is almost twice as long as male's (Promislow and Haselkorn 2002). *Canton S* and *Oregon R* lines of *D. melanogaster* differed in mean longevity of females, but not males, when they were kept individually, while in a single sex group of 20 individuals *Oregon R* females live longer and males live shorter comparing to *Canton S* line (Iliadi et al. 2009). Same authors showed that sexual dimorphism for longevity of virgin individuals was expressed only in *Oregon R* line, where females outlived males due to lower initial mortality rate. However, the results with *Oregon R* and *Canton S* lines may vary depending on experimental design (Cui et al. 2001; Moskalev et al. 2009; Vaiserman et al. 2013). As expected, QTL and quantitative genetic analysis revealed that genes affecting longevity are sex-specific and environment specific (Nuzhdin et al. 1997; Šešlija and Tucić 2008; Hallsson and Björklund 2011) implying that evolution of longevity may proceed by different mechanisms in females and males (Tower and Arbeitmann 2009) and may take different courses in diverse environmental circumstances.

Generally, life-span duration is influenced by female fitness-related traits more strongly compared with males, which is an expected consequence of positive correlation between female longevity and realized fecundity. Interestingly, this association between fecundity and life-span is even more important under stressful conditions. There is a plethora of evidence that females and males differently respond to the variations in environments. Substantial sex-specific plasticity (i.e., significant „sex $\times$ environment“ interaction) in adult longevity has been detected in environments with varying nutritional value of diet (Adler et al. 2013; Zajitschek et al. 2009b), temperature (Norry and Loeschcke 2002), oxidative stress (Lazarević et al. 2013), etc. (reviewed in Section 3.3. of this chapter). As will be elaborated in proceeding sections of this chapter, different evolutionary mechanisms can be involved in shaping sexual dimorphism in longevity and patterns of ageing, as well as in disparity of environmental sensitivity of these characteristics among genders.

First, natural selection acting either directly on life-span duration in the two sexes or indirectly through environmental effects on other fitness-related traits (e.g., developmental time, body size, fecundity, etc.) can mould longevity patterns in specific habitats (Section 1). For example, in a lepidopterous species *Pararge aegeria*, temperate (Atlantic island of Madeira) and north populations (southern Sweden) exhibit phenological differences in adult eclosion. Since eclosion is more synchronized in north than temperate populations, northern females are available to males during a shorter period of time which weakens selective pressure for long life in males and promote higher sexual dimorphism in longevity (Gotthard et al. 2000).

Another evolutionary force which shape insect longevity and mortality rate is sexual selection (Section 2). Sexual selection occurs whenever sexes differ in investment to reproduction. Theory predicts that males adopt “live fast – die young” strategy while females follow “low-risk, low-wear-and-tear” strategy. Male-biased mortality is usually related to risky behaviour, such as male sexual competition in polygamous mating systems, and higher allocation of resources to reproduction and secondary sexual traits which improve sexual presentation. In monogamous species, or species where mating success increases with age, selection may favour lower mortality in males than females (Bonduriansky et al. 2008).

Effects of mating on longevity depend on mating system of a species. Comparisons of virgin and mated individuals between polyandrous and monoandrous species of fireflies from Lampyridae family revealed that in polyandry longevity was not affected by copulation while in monoandry single mated females doubled their lifespan relative to virgin counterparts (Fu et al. 2012). Analyses of fireflies from genus *Photinus* showed that, in comparison with males from monoandrous species, polyandrous males allocate more resources towards reproductive tissues, which maximize their ability to produce nuptial gifts and increase reproductive success under high levels of male-male competition (Demary and Lewis 2007). Among lepidopterous species, investment in early reproduction in females is negatively correlated with longevity and degree of polyandry possibly because polyandrous females obtain more nutrients through nuptial gifts than monoandrous ones (Jervis et al. 2005). Intra-species comparisons in *Acanthoscelides obtectus* pointed that monoandrous conditions, inadvertently created during laboratory evolution for early reproduction and short-life, decrease the magnitude of sexual dimorphism in longevity (Stojković and Savković 2011).

As stressed by Bonduriansky et al. (2008), females and males do not differ in the level but rather in the nature and scheduling of reproductive effort. In natural conditions, *Teleogryllus commodus* females live longer, show lower baseline mortality and higher mortality rate than males. At the same time, long life is related to early reproductive effort in females and late reproductive effort in males (Zajitschek et al. 2009a). Under laboratory conditions, similar age-dependent pattern of reproduction has been recorded, but males outlived females due to lack of extrinsic mortality causes (Maklakov et al. 2009b). This implies that environmental conditions determine optimal relationship between longevity and reproduction.

1. THE ROLE OF NATURAL SELECTION IN SHAPING INSECT LONGEVITY

Fundamental to the standard theory of ageing is the idea that senescence is a result of decreasing intensity of natural selection after the onset of reproduction (Medawar, 1952; Williams 1957; Hamilton 1966). This means that any delay in the start of reproduction should lead to a postponement of senescence. This prediction has received quite extensive support, primarily in the experimental evolution studies that have been conducted on three insect species, *Drosophila melanogaster* (e.g., Rose and Charelsworth 1980; Rose 1984; Partridge and Flower 1992), *Acanthoscelides obtectus* (Tucić et al. 1996, 2004) and *Musca domestica* (Reed and Bryant 2000). By manipulating the ages at which selection is strong (i.e., using early vs. late reproduction selection regimes), these studies have shown that natural selection is indeed able to shape the evolution of ageing patterns. Experimental evolution studies have also suggested that an important genetic mechanism underlying postponed senescence could be antagonistic pleiotropy, which assumes the existence of alleles that improve early-life performance but have deleterious effects at later ages. Because the intensity of natural selection declines with advancing age, these alleles will increase in frequency and cause the senescence due to their negative effects on late-age performance (Williams 1957). The observations of a trade-off between early and late life performance, as in the above mentioned studies, are empirical support of the antagonistic pleiotropy hypothesis. A second prediction of this hypothesis is the existence of negative genetic correlation between early and late-life fitness-

related traits in populations with frequent segregating variation. These expectations also received experimental confirmation (e.g., Tucić et al. 1988), although in a less consistent way (Rose et al. 2005). It is noteworthy that an “early-late” trade-off in performance is also integral part of the “disposable soma” hypothesis (Kirkwood 1977), which proposes that senescence is caused by accumulation of damage due to processes that result in progressive decrease of the amount of resources allocated toward somatic maintenance and repair. This hypothesis assumes a physiological trade-off between reproduction and somatic maintenance and predicts that investment in reproduction (or somatic maintenance at earlier ages) should result in decrease of performance at later ages.

However, senescence might arise because of the accumulation of mutation with deleterious effects confined to later ages (Medawar 1952). Mutation accumulation hypothesis assumes that, because the strength of natural selection declines with age, genetic drift and recurrent deleterious mutations produce decline in individual performance at advancing ages. One of the effects of these processes could be upsurge of inbreeding depression for fitness-related traits with age. This is empirically testable prediction because crosses between populations, if they are subjected to age-specific mutation accumulation, should produce higher hybrid vigour (as the consequences of increased inbreeding load within populations) on fitness-related traits at later ages relative to the early ages (Charlesworth and Hughes 1996). Several studies of *Drosophila* have documented age-specific hybrid vigour for longevity and other fitness-related traits (Charlesworth and Hughes 1996; Hughes et al. 2002; Gong et al. 2006), but few others have shown that this relationship between age and inbreeding load varies among populations or are sex-specific (Lesser et al. 2006; Reynolds et al. 2007; Vaiserman et al. 2013). Furthermore, studies on two seed beetle species (*Callosobruchus chinensis*; Tanaka, 1990; and *C. maculatus*, Fox et al. 2006; Fox and Stillwell 2009) have produced results that were inconsistent with mutation accumulation hypothesis.

Experimental evolution studies, where replicate populations are exposed to different age-specific selection regimes, are especially suitable tool for testing mutation accumulation hypothesis. Here, the critical test of the hypothesis is the exploration of hybrid vigour in late-life fitness component in crosses between replicate populations subject to early reproduction. In study on short-lived early-reproducing *D. melanogaster* experimental populations, Rose et al. (2002) did not find differences between hybrid and nonhybrid populations in late-mortality rates implying that mutation accumulation could not be considered as a relevant explanation of genetic mechanisms involved in late mortality. Also, in subsequent work of Borash et al. (2007) on the same set of fruit flies populations no detectable heterotic effect was seen for female and male longevity and their hybrids. The only fitness-related trait whose age-specific changes were consistent with mutation accumulation hypothesis was male mating ability. Interestingly, in the same kind of experiment on the seed beetle *Acanthoscelides obtectus*, in contrast to females, mating ability of males exhibited inbreeding depression (Đorđević et al., in preparation), indicating that underlying genetic mechanisms for reproductive effort differed between sexes in populations selected for early reproduction and short life. It is noteworthy that, in these nonhybrid populations of the seed beetle, virgin males outlived virgin females, while in the long-lived nonhybrid populations females lived longer than males (sex-reversal with respect to longevity in the short-lived populations was also observed in Stojković and Savković 2011). It seems that in short lived populations evolutionary course reversed sex-specific mortality rates being that in the ancestral experimental population (i.e., base population), from which short- and long-lived populations have been derived, the longevity of

females were substantially higher than in males (Šešlija et al. 2008). Although this finding of sex-specific responses to inbreeding is quite unexpected, it is consistent with the result obtained by Bilde et al. (2009) in another seed beetle, *Callosobruchus maculatus*.

As pointed out by Bilde et al. (2009), one possible explanation as to why inbreeding, which increases homozygosity, affects males less than females presumes that the heterogametic sex (males in *A. obtectus* and *C. maculatus*) does not suffer from increased expression of X-linked recessive deleterious mutations. Another potential explanation for the sex-specific sensitivity to inbreeding is that males evolve mechanisms to counteract the negative effect of mitochondrial mutations on longevity, possibly by nuclear genes that influence expression of the mitochondrial genes (Yee et al. 2013). This hypothesis is based on the fact that mitochondrial genomes commonly function sub-optimally when expressed in males, as a consequences of maternal inheritance of mitochondrial genes and and their selection within females. Therefore, it is possible that the mismatch between mitochondrial and nuclear genes, which affect longevity, may be more pronounced in the inbred females than males. Consistent with this idea is the finding that in *D. simulans* females homozygous for some mutations in one nuclear-encoded mitochondrial protein pay a greater survival cost than do males (Melvin and Ballard 2011). The role of mitochondria in ageing process and sex-specific mortality patterns will be more elaborated in Section 3.3.1.

2. SEXUAL CONFLICT AND EVOLUTION OF SEXUAL DIMORPHISM IN LONGEVITY

The opposing sex-specific responses to inbreeding observed in the two seed beetle species indicate that an optimal longevity has evolved not only as a results of trade-offs between reproductive performance early and late in life, but also that this process could be influenced by changes in patterns of sexual selection. Bilde et al. (2009) and Maklakov and Lummaa (2013) argue that it is likely that sex-specific selection, which assumes that the optimal reproductive strategies of females and males differ, has also played an important role in the evolution of sexual dimorphism in longevity and mortality rates. Sexually divergent selection may produce two distinct forms of conflict with very different evolutionary outcomes. Intra-locus sexual conflict, which occurs whenever reproductive behaviours that are optimal for one sex are deleterious for the other sex, prevents traits which are shared by the sexes to evolve to their sex-specific phenotypic optima, because the shared traits have a common genetic basis. In contrast, inter-locus conflict involves different genes in each sex and it can cause open-ended cycles of adaptation and counter-adaptation between the sexes (Parker 1979). Since intra-locus conflict could lead a subset of genes coding for the shared traits to sex limited expression (see e.g., Lande 1980; Bonduriansky 2007 and references therein), it is expected that traits under this form of sexual selection possess the sex-specific genetic architecture. This prediction is supported in both seed beetle species, *C. maculatus* (Fox et al. 2004) and *A. obtectus* (Šešlija and Tucić 2008). In both studies it has been shown that genes coding for extended longevity were dominant in females and recessive in males, which raises the possibility that inbreeding could decrease female longevity and, at the same time, increase life span in males. Thus, it seems that the genetic architecture of longevity in these species evolves in a way that reduces genetic constraints on the independent evolution of the sexes. In one experiment on *D.*

melanogaster, Wayne et al. (2007) found that variation in gene expression was mostly the result of non-additive interactions between genes in females, while in males additive interactions of alleles prevailed. Having in mind that additive genetic variation respond to selection more readily, it is possible that genes affecting longevity may be more exposed to selection in males than in females, thereby reducing inbreeding load for male-specific deleterious mutations in populations.

Comparison of laboratory strains of *D. melanogaster* revealed that large part of transcriptional variation can be attributed to sexual dimorphism especially for genes expressed in germ-line tissues (Jin et al. 2001; Wayne et al. 2007; Baker et al. 2007; Wilson et al. 2013). These studies, in which sex-specific differences in the mode of transcriptome inheritance were identified, provide an additional insight into the gender differences in genetic architecture of the fitness-related traits. Gene expression pattern changes in response to mating and sexual selection and depending on mating system can have different evolutionary outcomes in females and males (Mank et al. 2013). For example, in *Drosophila* males, higher changes in transcription level during ageing was revealed for genes affecting protein synthesis, whereas in females starch and sucrose metabolism was more affected (Wilson et al. 2013). Magwire et al. (2010) analyzed effects of longevity-extending mutations on transcriptional pattern and found that each mutation had at least one deleterious pleiotropic effect and that these effects differed between the sexes.

The key feature of intra-sexual conflict in the evolution of longevity and ageing is that selection is working in the opposite directions in females and males, i.e., when female fitness is associated with increased longevity, male longevities should co-evolve, because of the intersexual genetic correlation for longevity, resulting in decreased male fitness. The predicted sex-specific relationship between fitness and longevity has been obtained in the seed beetle *C. maculatus*, where divergent sex-specific selection on longevity has been conducted (Berg and Maklakov 2012). In this experimental evolution study, selection for long-life produced replicate populations with low male fitness and high female fitness. More results implying that different reproductive strategies may lead females and males towards different optima in longevity has also been shown in the decorated cricket, *Gyllodes sigillatus* (Archer et al., 2012b). The sexual dimorphism in this species arises because females and males alter their reproductive effort with age; the reproductive effort increases with age in males but not in females. These strategies have led to an interesting evolutionary outcome - males live longer and age more slowly than females. Consistent with the view that intra-locus sexual conflict has the potential to play an important role in the evolution of sex differences in longevity and ageing are also findings obtained by Lewis et al. (2011) on indian meal moth, *Plodia interpunctella*. By using quantitative genetic approach these authors have shown that the pattern of multivariate selection acting on longevity and two other sexually dimorphic fitness-related traits was working in the opposing directions in the two sexes.

Several recent studies have indicated that some sexual adaptations in males, such as genital armour or increased toxicity of seminal fluid that polygynous males evolved to maximize their sperm competitive ability (possibly by destroying or disabling the sperm of previous mates), should lead to evolution of accelerated senescence in females. For example, in the laboratory evolution experiment on *C. maculatus*, Maklakov et al. (2007) have shown decreased rates of mortality and elevated longevity in monogamous virgin females relative to virgin females evolved under enforced polyandry. Similarly, female *D. melanogaster* had longer life-span after mating with males from lines selected for monogamy relative to females mated with

polyandrous males (Holland and Rice 1999). Thus, removal of sperm competition may have promoted evolution of more benign mates for females, which liberated them from costly counter-adaptations. Also, under such conditions, males did not trade-off resources between traits associated with male-to-male competition and longevity. Monogamous and polyandrous mating regimes have also been applied in other insect species: *D. pseudoobscura* (Bacigalupe et al. 2007), the dung flies, *Sepsis synipsea* and *Scathophaga stercoraria* (Hosken et al. 2001; Martin and Hosken 2003), the dung beetle, *Onthophagus taurus* (Simmons and Garcia-Gonzalez 2008). These studies support the idea that sexual selection can lead to significant gender differences in longevity and mortality rates.

Despite the potential importance of combined effect of sexual selection and natural selection for the evolution of longevity and mortality rates, there are few empirical studies that have attempted to tease apart these two evolutionary mechanisms. Maklakov and Fricke (2009), Maklakov et al. (2009a) and Maklakov et al. (2010) pioneered the approach of exploring consequences of simultaneous manipulations of mating systems (monogamy vs. polyandry) and age-specific selection (early vs. late reproduction) on the evolution of longevity in *C. maculatus*. The results showed that natural selection affects the evolution of longevity much more than did sexual selection.

The general effects of the conflict between members of the same sex over opportunity to mate on the evolution of longevity and mortality rates can mediate, to some extent, the species-specific adaptations. For example, because mated females with repressed oviposition live longer than virgin females, it seems that seed beetle (*A. obtectus*) females under starvation benefit from nutrients and/or water received with ejaculates (Tucić et al. 1996; Maklakov et al. 2005; Šešljija et al. 2008). Since *A. obtectus* is capital breeder (adults do not need water or food to reproduce successfully), this can be treated as an important female adaptation. Similar effects of male ejaculates on female fitness have also been observed in other bruchid insects (Savalli and Fox 1999; Edvardsson 2007).

To date, the role of sexual conflict in shaping longevity has been tested in very few species and focusing mostly on the consequences of conflict from the female's perspective. From the male's point of view, Hall et al. (2009) have revealed that nuptial feeding could be an important mechanism of antagonistic interactions between females and males and that in some cases, when resources are scarce, male reproductive investment in offspring, together with the associated costs of mating, can exceed the costs experienced by females. Under these circumstances, as pointed out by Hall et al. (2009), "the sex roles can reverse with females actually competing for access to males as potential food sources" (p. 873). In other words, changes in nature of sexual conflict may shape longevity and mortality rates of males in the same way as elevated sexual conflict do in females.

3. MECHANISTIC CAUSES OF SEX-SPECIFIC AGEING

Many experiments have shown that various genetic and environmental interventions exhibit sex-specific effects on longevity. Because of these gender differences Burger and Promislow (2004) emphasized the importance of studying mechanisms of aging separately in females and males (Table 2).

Table 2. Relative effect (% of change relative to control) of different gene manipulations in *Drosophila* females and males. With the exception of Cox7A where effects of gene mutation were examined for *D. simulans* all other data are for *D. melanogaster*

Gene	Function	Manipulation	Specificity	Trait	conditions	Females	Males	
Cox7A homozygotes	Cytochrome c oxidase subunit	Hypomorphic mutation	Ubiquitous in females, testis specific in males	50% survival,	virgin 23°C, 50% humidity	-18	No effect	Melvin and Ballard 2011
Methuselah homozygotes	Cellular signalling	Hypomorphic mutation		Mean longevity vs. <i>w¹¹¹⁸</i> control	virgin dark, 29°C, 75-90% humidity	+18	+15	Mockett and Sohal 2006
				Mean longevity vs. <i>w¹¹¹⁸</i> control	virgin 25°C, 50% humidity	+5	+22	
Hsf4 homozygotes	Heatshock response			Mean longevity vs. <i>Canton-S</i> control	Virgin 25°C	-37	-33.5	Moskalev et al. 2009
Hsf2/CyO Canton-S × Hsf2/CyO Hsf ⁰		Heat-sensitive		Mean longevity vs. Virgin, 25°C rescued mutant	Virgin, 25°C, high lipid food	-9 +33 +28 -26	+10 +12 No effect -29	Sørensen et al. 2007
UAS-D-GADD45 × GAL4-1407	Stress response, DNA repair, apoptosis	conditional overexpression	Nervous system, adults	Mean longevity vs. UAS-D-GADD45parent	Males and non-virgin females	+25.5	+73**	Plyusnina et al. 2011
				Mean longevity vs. GAL4-1407 parent		+58	+71	
p53, homozygotes	Apoptosis, mitochondrial metabolism	null mutant	-	Mean longevity	Virgin, 25°C	+13	+12	Waskar et al. 2009
p53		conditional overexpression	whole tissue	Mean longevity	29°C	-4 to -6	+10 to +18	Shen and Tower 2010
p53		conditional overexpression	Nervous system			+11, +12	-9, -13***	
GLaz -/-, Nlaz+/+	Stress response, fat storage	null mutant	Nervous system	50% survival vs. GLaz +/+, Nlaz+/+	Virgin, 25°C, 60% humidity	No effect	-15.5	Ruiz et al. 2011
GLaz +/-, Nlaz-/-	Stress response, systemic inhibition of IIS	null mutant	Nervous system,,			-30.8	-22.5	

Gene	Function	Manipulation	Specificity	Trait	conditions	Females	Males	
Human SOD	Antioxidative defence	overexpression	motoneurons	Mean longevity vs. UGA UAS-HS control strains	Virgin, 24°C	+43	+28	Spencer et al. 2003
<i>InRp5545/InRE19</i>	Insulin-like receptor	hypomorphic		Mean longevity	Mated, 25°C	+85	No effect	Tatar et al. 2001
Chico homozygous mutant	Substrate for InR	hypomorphic		Life expectancy	Mated, 25°C, 40% humidity	+57.6	+5.5	Tu et al. 2002
Chico heterozygous mutant						+36	+50	
dFOXO homozygous	Transcription factor in IIS/IGF-1 pathway	null mutant	-	Mean longevity vs. Canton S	Virgin, 25°C	-51	-68	Moskalev et al. 2011
<i>Lnk^{Del29}</i> homozygotes	Insulin/IGF-1 signaling	loss-of-function		50% survival vs. <i>w¹¹¹⁸</i> control	Mated, 25°C	+11.5	+13	Slack et al. 2010
<i>Lnk^{Del29}</i> heterozygotes						-3	No effect	
Sirt2 homozygous	protein deacetylase	null	-	Mean longevity vs. Canton S	Virgin, 25°C	+30	-27	Moskalev et al. 2011
<i>Sir2^{EP2300}</i>	protein deacetylase	overexpression	ubiquitous	50% survival	Virgin, 25°C	+57	+32	Rogina and Helfand 2004
		overexpression	Nervous system			+52	+20	
		overexpression	Fat body	50% survival	Mated, 25°C	+12.3	+13	Hoffmann et al. 2013
Enigma heterozygotes	β oxidation of fatty acids	null mutant		50% survival vs. <i>w¹¹¹⁸</i> control	25°C	+19.5	No effect	Mourikis et al. 2006
Gr63a homozygotes	Receptor in olfactory neurons	null mutant		Mean longevity vs. <i>w¹¹¹⁸</i> control	Virgin, 25°C, 60%humidity	+30	No effect	Poon et al. 2010
Tudor(1) homozygotes	Sexual differentiation	null mutant		Mean longevity vs. <i>Oregon R</i> control	Germ-line ablated, 25°C	-21,-2	+20, +20****	Shen et al. 2009
<i>Snazarus X</i> linked,	Endocytosis, signal transduction	transgene	fat body	Mean longevity vs. <i>w¹¹¹⁸</i> control	Virgin, 22-23°C	+62	+85	Stenesen 2011

average effect of 3 replications, *two independent transgenes

Proximate causes of sex-specific variation in longevity, which are the fuel for natural and sexual selection in shaping ageing patterns across sexes, include, besides genetic variability, hormonally regulated differences in behaviour, metabolic and immune functions, as well as tolerance to oxidative and other stresses between females and males (May 2007). In addition, it is important to keep in mind that differential allocation of resources between longevity and reproduction also depend on developmental stage and environmental conditions (Boggs 2009).

3.1. Neuroendocrine Regulation of Ageing

Neuroendocrine system regulate interaction between nutrition, metabolism, reproduction and ageing and adjust organism behaviour and physiology according to developmental and ecological context. This system of regulation presumes complex network of large number of interdependent elements whose empirical investigation requires thorough biochemical analyses in each population/ species/environment arrangement. Among various signalling pathways and gene functions involved in ageing-pattern regulation, here we elaborate several mechanisms that were extensively explored in insects.

Insulin signalling pathways link longevity, oxidative stress resistance and immunity. In *Drosophila*, down-regulation of insulin/IGF-1 signalling in mutants or individuals with ablated medial neurosecretory neurons (MNSN) increases longevity due to changes in the level of secondary hormones - juvenile hormone and ecdysone (Tatar et al. 2003). These hormones affect many aspects of reproduction such as oogenesis and spermatogenesis, synthesis of vitellogenin and accessory gland proteins, courtship behaviour, expression of secondary sexual traits etc. (Gruntenko and Rauschenbach 2008; Ishimoto et al. 2009; Moczek 2009; Toivonen and Partridge 2009). As expected, mutations in insulin/IGF-1 signalling pathway affects females more strongly than males since their investment in reproduction is higher (Clancy et al. 2001; Tatar et al. 2001; Tu et al. 2002). Ablation of MNSN increased median lifespan by 10.5% in males, 18.5% in virgin and 33.5% in mated females (Broughton et al. 2005). Beside reduced fecundity, mated females with ablated MNSN, also show increased lipid and carbohydrate storage, increased resistance to starvation and oxidative stress and reduced tolerance to heat and cold. The observed effects of insuline/IGF-1 signalling pathway on ageing are associated with its influence on levels of expression of some genes. For example, down-regulation of this signalling pathway prevent phosphorylation of dFOXO transcription factor and enables its translocation to nucleus leading to increased transcription of genes such as those related to oxidative stress resistance and response to nutrient variation. On the other hand, overexpression of dFOXO increases lifespan (Hwangbo et al. 2004), while null mutants for *dFOXO* gene exhibit significantly shorter life-span in both females and males (Moskalev et al. 2011). The importance of dFOXO activity for determining ageing pattern has been described in many studies. This transcription factor is involved in hormetic response to some stresses (Jünger et al. 2003; Moskalev et al. 2011), and also, by its influence on transcription of other genes, such as suppression of *Lnk* gene, dFOXO is involved in regulation of insulin/IGF-1 signalling pathway (Slack et al. 2010). Homozygous mutants for *Lnk* gene have increased lifespan in females and males, but females are more resistant to starvation and oxidative stress, and accumulate more lipids and trehalose than males (Slack et al. 2010).

dFOXO transcription factor, together with p53, control the expression of stress signalling gene *D-GADD45*. Its ubiquitous overexpression is lethal (Peretz et al. 2007). However, in

nervous system it extends lifespan without reduction in fecundity and locomotor activity (Plyusnina et al. 2011). Moreover, fecundity was higher which might explain differences in the mutation effects on females and males (**Table 2**). Another example of gene expressed in nervous tissue, which differently affects female and male longevity, is *GLaz* gene involved in regulation of stress resistance and fat storage. Null mutants showed reduced longevity only in males and this sex-specific modulation of longevity is related to changes in protein homeostasis (Ruiz et al. 2011). Mutant flies also showed reduced fecundity, deficient courtship and lower mating success. On the other hand, mutation in the gene for olfactory receptor increases both lifespan and reproduction in females, as well as fat storage and oxidative stress resistance (Poon et al. 2010).

As mentioned previously, insuline-like pathways control juvenile hormone and ecdysone - two signalling molecules that were found to be significant for sex-specific development of ageing patterns. Juvenile hormone is a pro-ageing hormone known to suppress stress resistance and immunity (Flatt et al. 2005) and to mediate trade-off between reproduction and immune functions. Addition of methopren (juvenile hormone analog) in larval food increases early fecundity and decreases lifespan in *D. melanogaster* (Flatt and Kawecki 2007). Mutations in ecdysone receptor (EcR) are lethal in homozygotes while heterozygous mutants show increased longevity by 45% both in females and males (Simon et al. 2003). However, mutant females exhibit higher resistance to heat stress, while males become more resistant to oxidative stress and dry starvation. In one experiment where *Drosophila* populations were selected for postponed reproduction, it was found that young females have decreased levels of 20-OH-ecdysone, while in males activity of this hormone was unaffected by selection regime (Harshman 1999). Similarly, higher investment in reproduction in short winged cricket females was related to increase in 20-OH-ecdysone, whereas short- and long-winged males did not differ in the level of 20-OH-ecdysone (Zera et al. 2007).

Together with juvenile hormone and ecdysone, another product of neurosecretory neurons that is involved in regulation of reproduction and mating behaviour is dopamine neurotransmitter (Gruntenko and Rauschenbach 2008). In addition, it seems that dopamine system also has a great impact on shaping longevity. Namely, it has been found in natural *Drosophila* populations that polymorphism in Dopa decarboxylase, enzyme of the last step of dopamine synthesis, account for a large fraction of genetic variation in longevity (De Luca et al. 2003). As shown, the level of dopamine is negatively correlated with lifespan (Vermeulen et al. 2006a). However, in response to selection for short life-span increased dopamine level was recorded in females, while in males the same dopamine response was obtained in populations selected for long life-span.

3.2. Metabolic Differences between Sexes and Longevity Consequences

Reallocation of limited resources among different functions underlies longevity-reproduction trade-off. However, in many species this physiological trade-off may not be obvious or complete being that different functions may require qualitatively different compounds for their energy supply. For example, although short-winged morph of *Gryllus* allocates more resources to ovaries comparing to long-winged morph, there is negative correlation between tryglyceride and phospholipide accumulation since tryglycerides are main flight fuel while phospholipids are main component of oocytes (Zera 2005).

Furthermore, strategy of metabolite accumulation during preadult development and their usage during ageing may differ between sexes. In *Ceratitis capitata* females and males differ in the pattern of change in lipid content with age (Pujol-Lereis et al. 2012). In *D. melanogaster* lipid content increases with ageing while opposite trend was recorded in males (Vermeulen et al. 2006b). On the other hand, in species which do not feed at adult stages, all metabolites decrease during ageing (Lazarević et al. 2012).

QTL analysis on *D. melanogaster* documented that lipid storage is under control of sex-specific genetic mechanisms (De Luca et al. 2005). Comparisons among different *Drosophila* species demonstrated significant among-species differences in sexual dimorphism of metabolite pool content (Matzkin et al. 2009). As shown on *D. melanogaster*, males have, on average, more triacylglycerols than females (De Luca et al. 2005). Selection for postponed reproduction in the same species increased carbohydrate and lipid content in females, while in males only lipid content was increased (Djawdan 1996). In accordance with metabolite pool changes, both sexes in long-lived flies showed increased starvation resistance. However, only long-lived females were more resistant to desiccation comparing to control population (Djawdan 1998). In *Acanthoscelides obtectus*, on the other hand, selection for postponed reproduction and long life resulted in increased carbohydrate and lipid content in males, whereas in females only glycogen content was elevated (Lazarević et al. 2012). At the same time, selection for early reproduction and short life-span resulted in females with lower lipid and higher protein contents. Similar trade-off pattern was found for short-winged early-reproduction female crickets. Results of biochemical analyses on this species suggest that these changes are accompanied with alterations in activity of enzymes of intermediary metabolism (Zera 2005).

3.2.1. Nutritional Basis of Ageing

Studies on the nutritional basis of ageing revolve around positive effects of dietary restriction on life-span (Partridge et al. 2005; Piper and Partridge 2007) and the role of specific macronutrients and their ratio in shaping patterns of ageing and reproduction (Lee et al. 2008; Maklakov et al. 2008; Maklakov et al. 2009b; Simpson and Raubenheimer 2009). Generally, most studies have documented that two sexes differ in their sensitivity to diet quality and quantity.

Due to their higher investment in reproduction, it is expected that females are more responsive to nutritionally poor food than males (Nakagawa et al. 2012). Dietary restriction (40% dilution of sucrose-yeast medium) in Dahomey wild type virgin females of *Drosophila melanogaster* led to 60% increase in longevity while males lived only 30% longer (Magwere et al. 2004). However, when mated individuals of Canton S *Drosophila* line were fed on corn meal-sucrose-yeast medium, no sex differences were recorded (Min and Tatar 2006). Nutrient poor diet in *Tellegrylus commodus* did not affect male longevity while in females longevity was increased for about 60% (Zajitschek et al. 2009b). On the other hand, in *Telostylinus angusticollis*, both sexes elevated life-span for 65% on nutrient poor diet (Adler et al. 2013). Since life-extending effects could be achieved in female *D. melanogaster* without ovaries (Mair et al. 2004) it was suggested that reallocation of resources to reproduction is not the main cause of observed effects.

As stressed by Piper et al. (2011), exploring the effects of dietary balance is more informative than dietary dilution. Insects encountering imbalanced food can rebalance nutrient intake and utilization by behavioural and physiological mechanisms (Bede et al. 2007; Behmer

2009; Clissold et al. 2010). Higher consumption of nutritionally poor food has been described as compensatory response in many insects. However, in experiment of Min and Tatar (2006) males did not show compensatory response and females even increased intake of nutrient rich food. As for most other physiological processes, regulatory capacity, nutrient requirements and age-specific changes in nutrient needs may differ between the sexes. For example, behavioural adjustment to sex-specific dietary requirements could be achieved by sex-specific dietary preference. Female and male larvae of the gypsy moth in advanced instars differ in preference for lipids and proteins (Stockhoff 1993). When they were offered two complementary cubes of artificial diet, females chose to eat more proteins and males consume more lipids. This finding is in line with high energy demands in flying adult males, on one hand, and high investment of proteins in egg production in females, on the other hand. In cockroaches, males prefer carbohydrate biased food which increases their weight gain and lipid accumulation. In addition, this specific diet increases attractiveness and pheromone production in males implying that levels and courses of sexual selection may be largely influenced by dietary manipulations (South and al. 2011). Similarly, in *Telleogrylus commodus*, females prefer food with higher protein to carbohydrate ratio than males, which mirrors sex-specific difference in nutrient requirements for reproduction (Maklakov et al. 2008). However, in contrast to males, preferred food does not maximize reproduction in females implying that there might be intra-locus sexual conflict over nutrient regulation. Moreover, life-span and reproductive effort are maximized at different protein to carbohydrate ratios in females and males. In addition to the observed sex differences, Maklakov et al. (2008) also found that nutrient requirements for maximizing longevity differ from optimal nutrition for maximizing reproduction. In males, nutrient requirements for the two fitness components lay in the similar region of two-dimensional nutrient space, but in females, reproduction is maximized in much higher protein to carbohydrate ratio than longevity indicating that nutrient intake should make compromise between these two traits. Discrepancy in nutrient requirements between longevity and reproduction in females has also been recorded in *Drosophila melanogaster*, *Bactrocera tryoni*, *Anastrepha ludens* (reviewed in Simpson and Raubenheimer 2009) and explain shorter female life-span compared with males in these species. In difference to virgin males in the study of Maklakov et al. (2008), mated *T. commodus* males have different nutrient landscapes for longevity and reproductive effort (Archer 2012). It seems that, in presence of females, males have higher demands for proteins because of increased calling efforts, increased sperm production and production of spermatophore. When given a choice, mated males eat even more proteins than females. Geometric framework approach was also applied for studying effects of nutrient ratio on immune functions and it was shown that different functions (e.g., phenoloxidase and lysozyme activity) are maximized at different ratios of proteins and carbohydrates (Cotter et al. 2011).

Among genes that were found important for dietary restriction response (Banerjee et al. 2012) and fat metabolism (Reis et al. 2010) in insects, detail analysis was performed on protein deacetylase, sirtuin Sir2. Its overexpression in nervous system increased life-span more in females than males (Rogina and Helfand 2004). Overexpression of Sir2 in fat body provokes similar increase in longevity in both sexes, although sex-specific changes in transcriptional pattern of some genes were recorded (Hoffmann et al. 2013). The role of Sir2 in regulating longevity can be understood in the light of its effect on p53 transcriptional factor which is downstream target of Sir 2 (Bauer et al. 2009).

3.2.2. Stress Resistance and Life-Span

There is a significant overlap in gene expression patterns between stress response and ageing which point to involvement of stress resistance mechanisms in somatic maintenance (Vermeulen and Loeschke 2007). Besides, selection for extended longevity is often related to increased stress resistance (Wit et al. 2013).

Exposure to a stressor activates various protective responses in cells and organism which may act as anti-ageing agents (Demirović and Rattan 2013). This hypothesis is confirmed by various beneficial effects of mild stresses, i.e., “hormesis”, that were found in many animal species (see Rattan 2012). Hormetic effects, as other explained mechanisms involved in determination of life-span, were also shown to be sex-specific and dependent on dose and genetic background (Sarup and Loeschke 2011). Generally, hormesis has been documented more frequently in males than females (Minois 2000). Male longevity was increased for hypergravity and cold (Le Bourg et al. 2009), X-radiation (Vaiserman et al. 2003), and heat (Sørensen et al. 2007). On the other hand, female *D. melanogaster* needed higher doses of morphin chloride for life-extending effect (Dubiley et al. 2011).

In *Drosophila melanogaster*, females are more resistant to desiccation (Gibbs et al. 1997; Chippindale et al. 1998), starvation (Harshman et al. 1999; Le Bourg 2007a), oxidative stress (Le Bourg 2007b; Minois 2001; Moskalev et al. 2009), high temperature (Le Bourg 2007a), and cold stress (Norry and Loeschke 2002). On the other hand, they are more sensitive to the presence of galactose in food (Cui et al. 2004), hypergravity (Le Bourg et al. 2009), radiation (Vaiserman et al. 2003), fungus infection (Le Bourg 2009) and combined effects of the stresses (Le Bourg 2012). In *Drosophila buzzatti* selected for increased longevity, increased heat stress resistance was found only in females (Scannapieco et al. 2009).

Investigations on the mechanistic basis of ageing focus mostly on the relationship between longevity and oxidative stress resistance. Even under optimal conditions, metabolic functions are related to production of reactive oxygen species (ROS) in mitochondria which may damage macromolecules and accelerate ageing if they are not balanced with ROS scavenging and damage repair systems (Isaksson et al. 2011). Despite many experiments which have succeeded to confirm basic postulates of free radical / oxidative stress theory of ageing (e.g., Arking et al. 2000, 2002; Wang et al. 2004; Orr et al. 2005) there are also many of them which failed to find positive correlation between oxidative stress resistance and longevity (Bayne et al. 2005; Loder 2006), between ROS production and longevity (Miwa et al. 2004; Sanz et al. 2010) or between activity of antioxidative enzymes and oxidative stress resistance (Seto et al. 1990; Mockett et al. 2001). It appeared that antioxidative defence is more relevant for survival under stressful conditions than for natural variation in ageing and longevity (Le Bourg 2001; Haenold et al. 2005). It must be kept in mind that reactive oxygen species, besides their detrimental effects, also have crucial regulatory role in gene expression, cell signalling, differentiation and apoptosis (Sohal and Orr 2012). They are involved in immune defence and mediate trade-off between immune function and somatic growth as well as between immunity, sexual ornamentation and sperm quality (Dowling and Sommons 2009). Thus, it is not surprising that many studies have revealed duplicitous role of oxidative stress determined by dose dependence of positive or negative effects of many antioxidant and prooxidant chemicals (Bahadorani et al. 2008; Bakkali et al. 2008; Ernst et al. 2013; Lazarević et al. 2013).

Depending on stress intensity (i.e., dose) and species' life-history strategies, oxidative stress can mediate trade-off between longevity and reproduction either as a consequence of differential allocation of antioxidants between soma and reproductive tissues or due to higher

energy demands of increased reproductive effort (Monaghan et al. 2009). It is well known that oocytes accumulate antioxidant proteins such as superoxide dismutase, catalase, glutathione S transferase, yolk proteins (Kurama et al. 1995; Collins et al. 2004; Freitas et al. 2007; Diaz-Albiter et al. 2011) which can be used for defence against oxidative challenge (Seehus et al. 2006; Schneider et al. 2011). In an experiment on *A. obtectus*, Lazarević et al. (2013) have found hormetic effect of low concentration of Paraquat (PQ) pro-oxidant agent on base line mortality of females selected for early reproduction and short life-span. Authors hypothesized that early-life hormesis was a consequence of higher egg load and, accordingly, increased content of yolk proteins known to serve as ROS scavenger in PQ challenged insects (Seehus et al. 2006) and be involved in regulation of iron homeostasis and antioxidative defence (Nichol et al. 2002; Pham and Winzerling 2010; Geiser and Winzerling 2012).

In experiment on *Gryllodes sigillatus*, Archer et al. (2012a, 2012b) determined calling effort in males and egg production in females as measures of investment to reproduction, similarly to other experiments with crickets (Hunt et al. 2006; Maklakov et al. 2008; Zajitschek et al. 2009a, 2009b), in order to explore the relationship between oxidative stress and longevity and reproduction. First, they found divergent reproductive strategies between sexes: females maximize reproductive output early in life, while males continually increase calling during ageing. Then, genetic correlations between life history traits and levels of oxidative damage (protein carbonyls) and levels of antioxidative defence (total antioxidative capacity) were determined. As expected, shorter lived sex (females) had higher level of damaged proteins. In both sexes, protein damage was positively correlated with baseline mortality and negatively correlated with life-span and early reproduction. In both sexes, increased damage is followed by increase in antioxidative capacity (positive genetic correlations). However, in contrast to males, late-life protein damage which affected life-span, ageing rate and antioxidative capacity, were not correlated with reproductive effort in females. These results, together with positive genetic correlations across sexes for the level of oxidative damage and antioxidative protection, suggest that sexes differ in managing their redox states and that intra-locus sexual conflict constrained their evolution towards optimal values.

Despite the known importance of mechanisms which repair or remove oxidatively damaged macromolecules, data on sexual dimorphism in these systems are scarce (Østergaard Hansen et al. 2012). Effects of selection for increased longevity and effects of genetic manipulations on the expression of genes for heat shock proteins were mostly studied in males (Kurapati et al. 2000; Tower 2010). Overexpression of Hsp proteins showed some fitness cost (Silberman and Tatar 2000) and effects depended on the tissue and developmental stage (reviewed in Hughes and Reynolds 2005). During ageing Hsp22-GFP and Hsp70-GFP reporters are up-regulated slightly more in males than females and more in muscle and nervous tissues (Tower et al. 2013). Mutations in *Hsp* genes have detrimental effects more to female than male longevity and diminishes female ability to survive under oxidative stress comparing to control flies (Moskalev et al. 2009). Under control conditions, mutations in genes for transcription factors for heat-shock proteins affects more female than male longevity and some of the mutations showed hormetic effect more often in females than males (Moskalev et al. 2009).

In *D. melanogaster*, activity of superoxide dismutase affects female longevity more than male's and positive effects of SOD overexpression was detected in more genetic backgrounds (Spencer et al. 2003). In *Acanthoscelides obtectus* females (longer lived sex) have higher level of catalase activity (Šešljija et al. 1999). In non-virgin *D. simulans*, females (shorter lived sex)

have higher production of hydrogen peroxide, lower activity of catalase and mitochondrial SOD, higher proton leak, ADP: O ratio and mtDNA density (Ballard et al. 2007). Since mating is energetically more costly to females, it may be expected that they are more capable to increase metabolic activity in response to reproduction.

3.2.3. The Roles of Mitochondria

Since mitochondria are major generator and target of intracellular ROS, they are suggested to play central role in the ageing process. Age-related decline in mitochondrial function, as well as life-extending effects of mitochondria-targeted genetic and environmental manipulations, demonstrate important role of mitochondria in longevity determination (reviewed in Morrow and Tanguay 2008 and Bratić and Trifunović 2010).

It is well known that functioning of the cell depends on communication between nucleus and mitochondria which, according to direct and retrograde signals, enables adjustment of metabolic reactions with energy requirements. Impairment of mitochondrial function, due to exposure to stress or lack of co-adaptation between nuclear and mitochondrial genomes, decreases ATP production, increases free radical leak and activate regulatory mechanisms which may lead either towards increased expression of respiratory complexes or towards apoptosis (Lane 2011). Outcome will depend on the apoptotic threshold level. Low threshold is related to good mito-nuclear match, low fertility and delayed ageing. As was shown in *Drosophila*, production of reactive oxygen species is lower in long-lived than short-lived lines (Driver and Tawadros 2000) and in females than males (Vina et al. 2003). Additionally, mitochondria-targeted antioxidants show sex-specific effects on longevity of *D. melanogaster* (Magwere et al. 2006; Krementsova et al. 2012)

It has been suggested that maternal inheritance of mitochondria contributes to sexual dimorphism in ageing and longevity by promoting higher co-adaptation of mitochondrial and nuclear genome in female sex (Zeh and Zeh 2005; Rand 2005; Tower 2006; Wolf and Gemmell 2012). On the other hand, it attenuates the strength of selection on male mitochondrial genome leading to increased mutation load and greater mitochondrial genetic variance for male than female longevity and ageing rate (Gemmell et al. 2004; Camus et al. 2012). In *D. melanogaster*, these mutations affect nuclear gene expression only in males modifying nearly 10% of transcripts mostly in reproductive tissues (Innocenti et al. 2011). As a consequences of asymmetric inheritance of mitochondrial genome, natural and sexually antagonistic selection can maintain mutations highly deleterious to males if they are beneficial to females, or if nuclear genome harbours modifiers which diminish their negative effects. The consequence of the disruption of mito-nuclear interactions may also be revealed as a change in gene expression pattern which lead to increased longevity (Rand et al. 2006).

Several genetic elements, which were mentioned earlier as important elements in regulating life-span, have also been found to have roles in controlling mitochondrial metabolism. Among other functions, transcription factor p53 is involved in regulation of apoptosis and exhibits antagonistic pleiotropic effects on longevity depending on age and sex (Shen and Tower 2010). Interestingly, sexual dimorphism of p53 effects is under the control of FOXO, while the level of the effects depends on Sir2 gene. Additionally, the mechanisms by which p53 exerts longevity effects might overlap with regulatory pathways which determine feeding behaviour and nutrient utilization.

ACKNOWLEDGMENTS

This work was supported by Ministry of Education, Science and Technological Development, grant No. 173007.

REFERENCES

- Adler MI, Cassidy EJ, Fricke C, Bonduriansky R (2013) The lifespan-reproduction trade-off under dietary restriction is sex-specific and context-dependent. *Exp Gerontol* 48: 539–548.
- Archer RC (2012) *Food, sex and death: costs of reproduction and the mechanistic basis of ageing*. PhD Dissertation, University of Exeter.
- Archer CR, Sakaluk SK, Selman C, Royle NJ, Hunt J (2012a) Oxidative stress and the evolution of sex differences in life span and ageing in the decorated cricket, *Gryllodes sigillatus*. *Evolution* 67: 620–634.
- Archer CR, Zajitschek F, Sakaluk SK, Royle NJ, Hunt J (2010b) Sexual selection affects the evolution of lifespan and ageing in the decorated cricket *Gryllodes sigillatus*. *Evolution* 66: 3088–3100.
- Arking R, Burde V, Graves K, Hari R, Feldman E, Zeevi A, Soliman S, Saraiya A, Buck S, Vettraino J, Sathrasala K, Wehr N, Levine RL (2000) Forward and reverse selection for longevity in *Drosophila* is characterized by alteration of antioxidant gene expression and oxidative damage patterns. *Exp Gerontol* 35: 167–185.
- Arking R, Buck S, Novoseltev VN, Hwangbo D-S, Lane M (2002) Genomic plasticity, energy allocations, and the extended longevity phenotypes of *Drosophila*. *Ageing Res Rev* 1: 209–228.
- Bacigalupe LD, Crudington HS, Hunter F, Moore AJ, Snook RR (2007) Sexual conflict does not drive reproductive isolation in experimental populations of *Drosophila pseudoobscura*. *J Evol Biol* 20:1763-1771.
- Bahadorani S, Bahadorani P, Phillips JP, Hilliker AJ (2008) The effects of vitamin supplementation on *Drosophila* life span under normoxia and under oxidative stress. *J Gerontol: Biol Sci* 63: 35–42.
- Bakkali F, Averbeck S, Averbeck D, Idaomar M (2008) Biological effects of essential oils – A review. *Food Chem Toxicol* 46: 446–475.
- Baker DA, Meadows LA, Wang J, Dow JAT, Russell S (2007) Variable sexually dimorphic gene expression in laboratory strains of *Drosophila melanogaster*. *BMC Genomics* 8: 454.
- Ballard JWO, Melvin RG, Miller JT, Katewa SD (2007) Sex differences in survival and mitochondrial bioenergetics during aging in *Drosophila*. *Aging Cell* 6: 699–708.
- Banerjee KK, Ayyub C, Ali SZ, Mandot V, Prasad NG, Kolthur-Seetharam U (2012) dSir2 in the adult fat body, but not in muscles, regulates life span in a diet-dependent manner. *Cell Rep* 2: 1485-1491.
- Bateman PW, Ferguson JWH, Yetman CA (2006) Courtship and copulation, but not ejaculates, reduce longevity of female field crickets (*Gryllus bimaculatus*). *J Zool* 268: 314–346.
- Bauer JH, Morris SN, Chang C, Flatt T, Wood JG, Helfand SL (2009) dSir2 and Dmp53 interact to mediate aspects of CR-dependent lifespan extension in *D. melanogaster*. *Aging* 1:38-48.

- Bayne ACV, Mockett RJ, Orr WC, Sohal RS (2005) Enhanced catabolism of mitochondrial superoxide/hydrogen peroxide and aging in transgenic *Drosophila*. *Biochem J* 391: 277–284.
- Bede JC, McNeil JN, Tobe SJ (2007) The role of neuropeptides in caterpillar nutritional ecology. *Peptides* 28: 185–196.
- Behmer ST (2009) Insect herbivore nutrient regulation. *Ann Rev Entomol* 54: 165–87.
- Berg CB, Maklakov AA (2012) Sexes suffer from suboptimal lifespan because of genetic conflict in a seed beetle. *Proc R Soc Lond B* 279: 4296–4302.
- Bilde T, Maklakov AA, Meisner K, la Guardia L, Friberg U (2009) Sex differences in the genetic architecture of lifespan in a seed beetle: Extreme inbreeding extends male lifespan. *BMC Evol Biol* 9: 33.
- Boggs CL (2009) Understanding insect life histories and senescence through a resource allocation lens. *Funct Ecol* 23: 27–37.
- Bonduriansky R (2007) The evolution of condition dependent sexual dimorphism. *Am Nat* 167: 9–19.
- Bonduriansky R, Maklakov A, Zajitschek F, Brooks R (2008) Sexual selection, sexual conflict and the evolution of ageing and life span. *Funct Ecol* 22: 443–453.
- Borash DJ, Rose MR, Mueller LD (2007) Mutation accumulation affects male virility in *Drosophila* selected for later reproduction. *Physiol Biochem Zool* 80: 461–472.
- Bratić I, Trifunović A (2013) Mitochondrial energy metabolism and ageing. *Biochim Biophys Acta* 1797: 961–967.
- Broughton SJ, Piper MDW, Ikeya T, Bass TM, Jacobson J, Driege Y, Martinez P, Hafen E, Withers DJ, Leivers SJ, Partridge L (2005) Longer lifespan, altered metabolism, and stress resistance in *Drosophila* from ablation of cells making insulin-like ligands. *Proc Natl Acad Sci USA* 102: 3105–3110.
- Burger JM, Promislow DE (2004) Sex-specific effects of interventions that extend fly life span. *Sci Aging Knowledge Environ*, pe30.
- Camus MF, Clancy DJ, Dowling DK (2012) Mitochondria, maternal inheritance, and male aging. *Curr Biol* 22: 1717–1721.
- Carey JR, Liedo P, Harshman L, Zhang Y, M H-G, Partridge L, Wang J-L (2002) A mortality cost of virginity at older ages in female Mediterranean fruit flies. *Exp Gerontol* 37: 507–512.
- Charlesworth B, Hughes KA (1996) Age-specific inbreeding depression and components of genetic variance in relation to the evolution of senescence. *Proc Natl Acad Sci USA* 93: 6140–6145.
- Chippindale AK, Gibbs AG, Sheik M, Yee KJ, Djawdan M, Bradley TJ, Rose MR (1998) Resource acquisition and the evolution of stress resistance in *Drosophila melanogaster*. *Evolution* 52: 1342–1352.
- Clancy DJ, Gems D, Harshman LG, Oldham S, Stocker H, Hafen E, Leivers SJ, Partridge L (2001) Extension of life-span by loss of CHICO, a *Drosophila* insulin receptor substrate protein. *Science* 292: 104–106.
- Clissold FJ, Tedder BJ, Conigrave AD, Simpson SJ (2010) The gastrointestinal tract as a nutrient-balancing organ. *Proc R Soc* 277B: 1751–1759.
- Collins AM, Williams V, Evans JD (2004) Sperm storage and antioxidative enzyme expression in the honey bee, *Apis mellifera*. *Insect Molec Biol* 13: 141–146.

- Cotter SC, Simpson SJ, Raubenheimer D, Wilson K (2011) Macronutrient balance mediates trade-offs between immune function and life history traits. *Funct Ecol* 25: 186–198.
- Cui X, Wang L, Zuo P, Han Z, Fang Z, Li W, Liu J (2004) D-Galactose-caused life shortening in *Drosophila melanogaster* and *Musca domestica* is associated with oxidative stress. *Biogerontology* 5: 317–325.
- De Loof A (2011) Longevity and aging in insects: Is reproduction costly; cheap; beneficial or irrelevant? A critical evaluation of the “trade-off” concept. *J Insect Physiol* 57: 1–11.
- Demary KC, Lewis SM (2007) Male reproductive allocation in fireflies (*Photinus* spp.). *Invertebr Biol* 126: 74–80.
- Demirović D, Rattan SIS (2013) Establishing cellular stress response profiles as biomarkers of homeodynamics, health and hormesis. *Exp Gerontol* 48: 94–98.
- De Luca M, Roshina NV, Geiger-Thornsberry GL, Lyman RF, Pasyukova EG, Mackay TFC (2003) Dopa decarboxylase (Ddc) affects variation in *Drosophila* longevity. *Nat Genet* 34: 429–33.
- De Luca M, Yi N, Allison DB, Leips J, Ruden DM (2005) Mapping quantitative trait loci affecting variation in *Drosophila* triacylglycerol storage. *Obesity Res* 13: 1596–1605.
- Diaz-Albiter HD, Mitford R, Fernando A, Genta FA, Sant’Anna MRV, Dillon RJ (2011) Reactive oxygen species scavenging by catalase is important for female *Lutzomyia longipalpis* fecundity and mortality. *PLoS ONE* 6(3): e17486.
- Djardjan M, Chippindale AK, Rose MR, Bradley TJ (1998) Metabolic reserves and evolved stress resistance in *Drosophila melanogaster*. *Physiol Zool* 71: 584–594.
- Djardjan M, Sugiyama TT, Schlaeger LK, Bradley TJ, Rose MR (1996) Metabolic aspects of the trade-off between fecundity and longevity in *Drosophila melanogaster*. *Physiol Zool* 69: 1176–1195.
- Dowling DK, Simmons LW (2009) Reactive oxygen species as universal constraints in life-history evolution. *Proc R Soc* 276B: 1737–1745.
- Driver C, Tawadros N (2000) Cytoplasmic genomes that confer additional longevity in *Drosophila melanogaster*. *Biogerontology* 1: 255–260.
- Dubiley TA, Rushkevich YE, Koshel NM, Voitenko VP, Vaiserman AM (2011) Life span extension in *Drosophila melanogaster* induced by morphine. *Biogerontology* 12: 179–184.
- Edvardsson M (2007) Female *Callosobruchus maculatus* mate when they are thirsty: resource rich ejaculates as mating effort in a beetle. *Anim Beh* 74: 183–188.
- Ernst IMA, Pallauf K, Bendall JK, Paulsen L, Nikolai S, Huebbe P, Roeder T, Rimbach G (2013) Vitamin E supplementation and lifespan in model organisms. *Ageing Res Rev* 12: 365–375.
- Fedorka KM and Zuk M (2005) Sexual conflict and female immune suppression in the cricket, *Allonemobius socius*. *J Evol Biol* 18: 1515–1522.
- Flatt T, Kawecki TJ (2007) Juvenile hormone as a regulator of the trade-off between reproduction and life span in *Drosophila melanogaster*. *Evolution* 61: 1980–1991.
- Flatt T, Tu M-P, Tatar M (2005) Hormonal pleiotropy and the juvenile hormone regulation of *Drosophila* development and life history. *BioEssays* 27: 999–1010.
- Fox CW, Czesak ME, Wallin WG (2004) Complex genetic architecture of population differences in adult lifespan of a beetle: nonadditive inheritance, gender differences, body size and a large maternal effect. *J Evol Biol* 17: 1007–1017.
- Fox CW, Dublin L, Pollitt SJ (2003) Gender differences in lifespan and mortality rates in two seed beetle species. *Funct Ecol* 17: 619–626.

- Fox CW, Scheibly KL, Wallin WG, Hitchcock Lj, Stillwell RC, Smith BP (2006) The genetic architecture of life span and mortality rates: gender and species differences in inbreeding load of two seed-feeding beetles. *Genetics* 174: 763-773.
- Fox CW, Stillwell RC (2009) Environmental effects on sex differences in the genetic load for adult lifespan in a seed-feeding beetle. *Heredity* 103: 62-72.
- Freitas DRJ, Rosa RM, Moraes J, Campos E, Logullo C, Da Silva Vaz Jr. I, Masuda A (2007) Relationship between glutathione S-transferase, catalase, oxygen consumption, lipid peroxidation and oxidative stress in eggs and larvae of *Boophilus microplus* (Acarina: Ixodidae). *Comp Biochem Physiol* 146A: 688-694.
- Fu X, South A, Lewis SM (2012) Sexual dimorphism, mating systems, and nuptial gifts in two Asian fireflies (Coleoptera: Lampyridae). *J Insect Physiol* 58: 1485-1492.
- Geiser DL, Winzerling JJ (2012) Insect transferrins: multifunctional proteins. *Biochim Biophys Acta* 1820: 437-451.
- Gibbs AG, Chippindale AK, Rose MR (1997) Physiological mechanisms of evolved desiccation resistance in *Drosophila melanogaster*. *J Exp Biol* 200: 1821-1832.
- Gong Y, Thompson JN, Woodruff RC (2006) Effect of deleterious mutations on life span in *Drosophila melanogaster*. *J Gerontol Series A Biol Sci Med Sci* 61: 1246-1252.
- Gotthard K, Nylin S, Wiklund C (2000) Mating opportunity and the evolution of sex-specific mortality rates in a butterfly. *Oecologia* 122: 36-43.
- Gruntenko NE, Rauschenbach IYu (2008) Interplay of JH, 20E and biogenic amines under normal and stress conditions and its effect on reproduction. *J of Insect Physiol* 54: 902-908.
- Haenold R, Wassef DM, Heinemann SH and Hoshi T (2005) Oxidative damage, aging and anti-aging strategies. *Age* 27: 183-199.
- Hall MD, Bussiere LF, Brooks R (2009) Diet-dependent female evolutionary influence male lifespan in a nuptial feeding insect. *J Evol Biol* 22:873-881.
- Hallsson LR, Björklund M (2012) Sex-specific genetic variances in life-history and morphological traits of the seed beetle *Callosobruchus maculatus*. *Ecol Evol* 2: 128-138.
- Hamilton WD (1966) The moulding of senescence by natural selection. *J Theor Biol* 12: 12-45.
- Harshman LG (1999) Investigation of the endocrine system in extended longevity lines of *Drosophila melanogaster*. *Exp Gerontol* 34: 997-1006.
- Harshman LG, Moore KM, Sty MA, Magwire MM (1999) Stress resistance and longevity in selected lines of *Drosophila melanogaster*. *Neurobiol Aging* 20: 521-529.
- Hoffmann J, Romey R, Fink C, Yong L, Roeder T (2013) Overexpression of *Sir2* in the adult fat body is sufficient to extend lifespan of male and female *Drosophila*. *Aging* 5: 315-327.
- Holland B, Rice WR (1999) Experimental removal of sexual selection reverses intersexual antagonistic coevolution and removes a reproductive load. *Proc Natl Acad Sci, USA* 96: 5083-5088.
- Hosken DJ, Garner TWJ, Ward PI (2001) Sexual conflict selects for male and female reproductive characters. *Curr Biol* 11: 489-493.
- Hughes KA, Alipaz JA, Drnevich JM, Reynolds RM (2002) A test of evolutionary theories of aging. *Proc Natl Acad Sci, USA* 99: 14286-14291.
- Hughes KA, Reynolds RM (2005) Evolutionary and mechanistic theories of aging. *Ann Rev Entomol* 50: 421-445.

- Hunt J, Jennions MD, Spyrou N, Brooks R (2006) Artificial selection on male longevity influences age-dependent reproductive effort in the black field cricket *Teleogryllus commodus*. *Am Nat* 168: E72–E86.
- Hwangbo DS, Gershman B, Tu MP, Palmer M, Tatar M (2004) *Drosophila* dFOXO controls lifespan and regulates insulin signalling in brain and fat body. *Nature* 429: 562–566.
- Iliadi KG, Natalia N, Iliadi NN, Boulianne GL (2009) Regulation of *Drosophila* life-span: Effect of genetic background, sex, mating and social status. *Exp Gerontol* 44: 546–553.
- Innocenti P, Morrow EH, Dowling DK (2011) Experimental evidence supports a sex-specific selective sieve in mitochondrial genome evolution. *Science* 332: 845–848.
- Isaksson C, Sheldon BC, Uller T (2011) The challenges of integrating oxidative stress into life-history biology. *BioScience* 61: 194–202.
- Ishimoto H, Sakaia T, Kitamoto T (2009) Ecdysone signaling regulates the formation of long-term courtship memory in adult *Drosophila melanogaster*. *Proc Natl Acad Sci* 106: 6381–6386.
- Ivy TM, Sakaluk SK (2005) Polyandry promotes enhanced offspring survival in decorated crickets. *Evolution* 59: 152–159.
- Jauregui LM, Etges WJ (2007) Assessing patterns of senescence in *Drosophila mojavensis* reared on different host cacti. *Evol Ecol Res* 9: 1–17.
- Jervis MA, Boggs CL, Ferns PN (2005) Egg maturation strategy and its associated trade-offs: a synthesis focusing on Lepidoptera. *Ecol Entomol* 30: 359–375.
- Jin W, Riley RM, Wolfinger RD, White KP, Passador-Gurgel G, Gibson G (2001) The contributions of sex, genotype and age to transcriptional variance in *Drosophila melanogaster*. *Nat Genet* 29: 389–395.
- Jünger M, Rintelen F, Stocker H, Wasserman J, Végh M, Radimerski T, Greenberg M, Hafen E (2003) The *Drosophila* Forkhead transcription factor FOXO mediates the reduction in cell number associated with reduced insulin signaling. *J Biol* 2(3):20.
- Kirkwood TBL (1977) Evolution of ageing. *Nature* 270: 301–304.
- Krementsova AV, Roshina NV, Tsybul'ko EA, Rybina OY, Symonenko AV, Pasyukova EG (2012) Reproducible effects of the mitochondria-targeted plastoquinone derivative SkQ1 on *Drosophila melanogaster* lifespan under different experimental scenarios. *Biogerontology* 13: 595–607.
- Kurama T, Kurata S, Natori S (1995) Molecular characterization of an insect transferrin and its selective incorporation into eggs during oogenesis. *Eur J Biochem* 228: 229–235.
- Kurapati R, Passananti HB, Rose MR, Tower J (2000) Increased *hsp22* RNA levels in *Drosophila* lines genetically selected for increased longevity. *J Gerontol: Biol Sci* 55A: B552–B559.
- Lande R (1980) Sexual dimorphism, sexual selection and adaptation in polygenic characters. *Evolution* 38:292–305.
- Lane N (2011) Mitonuclear match: Optimizing fitness and fertility over generations drives ageing within generations. *BioEssays* 33: 860–869.
- Lazarević J (1994) *The effect of host plant on digestive enzymes and variation of fitness components in the gypsy moth Lymantria dispar* L. MS Thesis, University of Belgrade.
- Lazarević J, Đorđević M, Stojković B, Tucić N (2013) Resistance to prooxidant agent paraquat in the short- and long-lived lines of the seed beetle (*Acanthoscelides obtectus*). *Biogerontology* (2013) 14:141–152.

- Lazarević J, Tucić N, Šešlija Jovanović D, Večeřa J, Kodrík D (2012) The effects of selection for early and late reproduction on metabolite pools in *Acanthoscelides obtectus* Say. *Insect Sci* 19: 303–314.
- Le Bourg É (2001) Oxidative stress, aging and longevity in *Drosophila melanogaster*. *FEBS Letters* 498: 183–186.
- Le Bourg É (2007a) Hormetic effects of repeated exposures to cold at young age on longevity, aging and resistance to heat or cold shocks in *Drosophila melanogaster*. *Biogerontology* 8: 431–444.
- Le Bourg É (2007b) Hormetic effects on longevity of hydrogen peroxide in *Drosophila melanogaster* flies living on a poorly nutritious medium. *Biogerontology* 8: 327–344.
- Le Bourg É (2012) Combined effects of two mild stresses (cold and hypergravity) on longevity, behavioral aging, and resistance to severe stresses in *Drosophila melanogaster*. *Biogerontology* 13: 313–328.
- Le Bourg É, Massou I, Vanessa Gobert V (2009) Cold stress increases resistance to fungal infection throughout life in *Drosophila melanogaster*. *Biogerontology* 10: 613–625.
- Lee KP, Simpson SJ, Clissold FJ, Brooks R, Ballard JWO, Taylor PW, Soran N, Raubenheimer D (2008) Lifespan and reproduction in *Drosophila*: New insights from nutritional geometry. *Proc Natl Acad Sci* 105: 2498–2503.
- Lesser KJ, Paiusi IC, Leips J (2006) Naturally occurring genetic variation in the age-specific immune response of *Drosophila melanogaster*. *Aging Cell* 5: 293–295.
- Lewis Z, Wedell N, Hunt J (2011) Evidence for strong intralocus sexual conflict in the Indian meal moth, *Plodia interpunctella*. *Evolution* 65: 2085–2097.
- Loder D (2006) *Effect of hsp70 overexpression on life span and stress resistance in Drosophila melanogaster*. Doctoral dissertation, Queen's University, Kingston, ON.
- Magwere T, Chapman T, Partridge L (2004) Sex differences in the effect of dietary restriction on lifespan and mortality rates in female and male *Drosophila melanogaster*. *J Gerontol A: Biol Sci Med Sci* 59: B3–B9.
- Magwere T, West M, Riyahi K, Murphy MP, Smith RAJ and Partridge L (2006) The effects of exogenous antioxidants on lifespan and oxidative stress resistance in *Drosophila melanogaster*. *Mech Ageing Dev* 127: 356–370.
- Magwire MM, Yamamoto A, Carbone MA, Roshina NV, Symonenko AV, Pasyukova EG, Morozova TV, Mackay TFC (2010) Quantitative and molecular genetic analyses of mutations increasing *Drosophila* life span. *PLoS Genet* 6(7): e1001037.
- Mair W, Sgrò CM, Johnson AP, Chapman T, Partridge L (2004) Lifespan extension by dietary restriction in female *Drosophila melanogaster* is not caused by a reduction in vitellogenesis or ovarian activity. *Exp Gerontol* 39: 1011–1019.
- Maklakov AA, Bonduriansky R, Brooks RC (2009a) Sex differences, sexual selection, and ageing: An experimental evolution approach. *Evolution* 63: 2491–2503.
- Maklakov AA, Hall MD, Simpson SJ, Dessmann J, Clissold FJ, Zajitschek F, Lailvaux SP, Raubenheimer D, Bonduriansky R, Brooks RC (2009b) Sex differences in nutrient-dependent reproductive ageing. *Aging Cell* 8:324–330.
- Maklakov AA, Cayetano L, Brooks RC, Bonduriansky R (2010) The role of life-history selection and sexual selection in the adaptive evolution of mating behavior in a beetle. *Evolution* 64: 1273–1282.
- Maklakov AA, Fricke C (2009) Sexual selection did not contribute to the evolution of male lifespan under curtailed age at reproduction in a seed beetle. *Ecol Entomol* 34: 638–643.

- Maklakov AA, Fricke C, Arnqvist G (2007) Sexual selection affects lifespan and ageing in the seed beetle. *Ageing Cell* 6: 739–744.
- Maklakov AA, Kremer N, Arnqvist G (2005) Adaptive male effects on female ageing in seed beetles. *Proc R Soc* 272: 2485–2489.
- Maklakov AA, Lummaa V (2013) Evolution of sex differences in lifespan and aging: cause and constraints. *BioEssays* Doi 10.1002/bies.201300021.
- Maklakov AA, Simpson SJ, Zajitschek F, Hall MD, Dessmann J, Clissold F, Raubenheimer D, Bonduriansky R, Brooks RC (2008) Sex-specific fitness effects of nutrient intake on reproduction and lifespan. *Curr Biol* 18: 1062–1066.
- Mank JE, Wedell N, Hosken DJ (2013) Polyandry and sex-specific gene expression. *Philos Trans R Soc Lond B Biol Sci* 368: 20120047.
- Martin OY, Hosken DJ (2003) Costs and benefits of evolving under experimentally enforced polyandry or monogamy. *Evolution* 57: 2765–2772.
- Matzkin LM, Mutsaka K, Johnson S, Markow TA (2009) Metabolic pools differ among ecologically diverse *Drosophila* species. *J Insect Physiol* 55: 1145–1150.
- May RC (2007) Gender, immunity and the regulation of longevity. *BioEssays* 29: 795–802.
- Mayr E (1961) Cause and effect in biology. *Science* 134: 1501–1506.
- Medawar PB (1952) *An unsolved problem of biology*. HK Lewis, London.
- Melvin RG, Ballard WO (2011) Females with a mutation in a nuclear-encoded mitochondrial protein pay a higher cost of survival than do males in *Drosophila*. *J Gerontol A Biol Sci Med Sci* 66A: 765–770.
- Min K-J, Tatar M (2006) Restriction of amino acids extends lifespan in *Drosophila melanogaster*. *Mech Ageing Dev* 127: 643–646.
- Minois N (2000) Longevity and aging: beneficial effects of exposure to mild stress. *Biogerontology* 1: 15–29.
- Minois N (2001) Resistance to stress as a function of age in transgenic *Drosophila melanogaster* overexpressing Hsp70. *J Insect Physiol* 47: 1007–1012.
- Miwa S, Riyahi K, Partridge L, Brand MD (2004) Lack of correlation between mitochondrial reactive oxygen species production and life span in *Drosophila*. *Ann NY Acad Sci* 1019: 388–391.
- Mockett RJ, Orr WC, Rahmandar JJ, Sohal BH and Sohal RS (2001) Antioxidant status and stress resistance in long- and short-lived lines of *Drosophila melanogaster*. *Exp Gerontol* 36: 441–463.
- Mockett RJ, Sohal RS (2006) Temperature-dependent trade-offs between longevity and fertility in the *Drosophila* mutant, *methuselah*. *Exp Gerontol* 41: 566–573.
- Moczek AP (2009) Phenotypic plasticity and the origin of diversity: a case study on horned beetles, pp 81–134. In *Phenotypic Plasticity of Insects* (ed. by DW Whitman and T.N. Ananthakrishnan), Science Publishers, Enfield, NH, USA.
- Monaghan P, Metcalfe NB, Torres R (2009) Oxidative stress as a mediator of life history trade-offs: mechanisms, measurements and interpretation. *Ecol Letters* 12: 75–92.
- Morrow EH, Arnqvist G (2003) Costly traumatic insemination and a female counter-adaptation in bed bugs. *Proc R Soc Lond B* 270B: 2377–2381.
- Morrow G, Tanguay RM (2008) Mitochondria and ageing in *Drosophila*. *Biotechnol. J* 3: 728–739.

- Moskalev A, Shaposhnikov M, Turysheva E (2009) Life span alteration after irradiation in *Drosophila melanogaster* strains with mutations of Hsf and Hsps. *Biogerontology* 10: 3–11.
- Moskalev AA, Plyusnina EN, Shaposhnikov MV (2011) Radiation hormesis and radioadaptive response in *Drosophila melanogaster* flies with different genetic backgrounds: the role of cellular stress-resistance mechanisms. *Biogerontology* 12:253–263.
- Mourikis P, Hurlbut GD, Artavanis-Tsakonas S (2006) Enigma, a mitochondrial protein affecting lifespan and oxidative stress response in *Drosophila*. *Proc Natl Acad Sci USA* 103: 1307–1312.
- Nakagawa S, Lagisz M, Hector KL, Spencer HG (2012) Comparative and metaanalytic insights into life extension via dietary restriction. *Aging Cell* 11: 401–409.
- Nichol H, Law JH, Winzerling JJ (2002) Iron metabolism in insects. *Annu Rev Entomol* 47: 535–559.
- Norry FM, Loeschcke VR (2002) Longevity and resistance to cold stress in cold-stress selected lines and their controls in *Drosophila melanogaster*. *J Evol Biol* 15: 775–783.
- Nuzhdin SV, Pasyukova EG, Dilda CL, Zeng Z-B, Mackay TFC (1997) Sex-specific quantitative trait loci affecting longevity in *Drosophila melanogaster*. *Proc Natl Acad Sci USA* 94: 9734–9739.
- Orr WC, Radyuk SN, Prabhudesai L, Toroser D, Benes JJ, Luchak JM, Mockett RJ, Rebrin I, Hubbard JG, Sohal RS (2005) Overexpression of glutamate-cysteine ligase extends life span in *Drosophila melanogaster*. *J Biol Chem* 280: 37331–37338.
- Østergaard Hansen T, Sarup P, Loeschcke V, Rattan SIS (2012) Age-related and sex-specific differences in proteasome activity in individual *Drosophila* flies from wild type, longevity-selected and stress resistant strains. *Biogerontology* 13: 429–438.
- Parker GA (1979) Sexual selection and sexual conflict, pp. 123–166. In: *Sexual Selection and Reproductive Competition in Insects*, (ed. by MS Blum and NA Blum). Academic Press: New York.
- Partridge L, Fowler K (1992) Direct and correlated responses to selection on age at reproduction in *Drosophila melanogaster*. *Evolution* 46: 76–91.
- Partridge L, Piper MD, Mair W (2005) Dietary restriction in *Drosophila*. *Mech Ageing Dev* 126: 938–950.
- Peretz G, Bakhrat A, Abdu U (2007) Expression of the *Drosophila melanogaster* GADD45 homolog (CG11086) affects egg asymmetric development that is mediated by the c-Jun N-terminal kinase pathway. *Genetics* 177:1691–1702.
- Pham DQD, Winzerling JJ (2010) Insect ferritins: typical or atypical? *Biochim Biophys Acta* 1800: 824–833.
- Piper MDW, Partridge L (2007) Dietary restriction in drosophila: delayed aging or experimental artefact? *PLoS Genetics* 3: e57.
- Piper MDW, Partridge L, Raubenheimer D, Simpson SJ (2011) Dietary restriction and aging: A unifying perspective. *Cell Metabol* 14:154–160.
- Plyusnina EN, Shaposhnikov MV, Moskalev AA (2011) Increase of *Drosophila melanogaster* lifespan due to D-GADD45 overexpression in the nervous system. *Biogerontology* 12:211–226.
- Poon PC, Kuo T-H, Linford NJ, Roman G, Pletcher SD (2010) Carbon Dioxide Sensing Modulates Lifespan and Physiology in *Drosophila*. *PLoS Biol* 8(4): e1000356.

- Promislow DEL and Haselkorn TS (2002) Age-specific metabolic rates and mortality rates in the genus *Drosophila*. *Aging Cell* 1: 66–74.
- Pujol-Lereis LM, Rabossi A, Quesada-Allué LA (2012) Lipid profiles as indicators of functional senescence in the medfly. *Exp Gerontol* 47: 465–472.
- Rand DM (2005) Mitochondrial genetics of aging: intergenomic conflict resolution. *Sci. Aging Knowledge Environ* re5.
- Rand DM, Fry A, Sheldahl L (2006) Nuclear–mitochondrial epistasis and *Drosophila* aging: introgression of *Drosophila simulans* mtDNA modifies longevity in *D. melanogaster* nuclear backgrounds. *Genetics* 172: 329–341.
- Rattan SIS (2012) Biogerontology: from here to where? The Lord Cohen Medal Lecture—2011. *Biogerontology* 13: 83–91.
- Reed DH, Braynt EH (2000) The evolution of senescence under curtailed life span in laboratory populations of *Musca domestica* (the housefly). *Heredity* 85: 115–121.
- Reis T, Van Gilst MR, Hariharan IK (2010) A buoyancy-based screen of *Drosophila* larvae for fat-storage mutants reveals a role for Sir2 in coupling fat storage to nutrient availability. *PLoS Genet* 6: e1001206.
- Reynolds RM, Temiyasathit S, Reedy MM, Ruedi EA, Drnevich JM, Leips J, Hughes KA (2007) Age specificity of inbreeding load in *Drosophila melanogaster* and implications for the evolution of late-life mortality plateaus. *Genetics* 177: 587–595.
- Rogina B, Helfand SL (2004) Sir2 mediates longevity in the fly through a pathway related to calorie restriction. *Proc Natl Acad Sci* 101: 15998–16003.
- Rolff J, Siva-Jothy MT (2002) Copulation corrupts immunity: a mechanism for a cost of mating in insects. *Proc Natl Acad Sci USA* 99: 9916–9918.
- Rönn J, Katvala M, Arnqvist G (2006) The costs of mating and egg production in *Callosobruchus* seed beetles. *Animal Behav* 72: 335–342.
- Rose MR (1984) Laboratory evolution of postponed senescence in *Drosophila melanogaster*. *Evolution* 38: 1004–1010.
- Rose MR, Charlesworth B (1980) A test of evolutionary theories of senescence. *Nature* 287: 141–142.
- Rose MR, Drapeau MD, Yazdi PG, Shah KH, Moise DB, Thakar RR, Rauser CL, Mueller LD (2002) Evolution of late-life mortality in *Drosophila melanogaster*. *Evolution* 56: 1982–1991.
- Rose MR, Passananti HB, Chippindale AK, Phelan JP, Matos M, Teotonio H, Mueller LD (2005) The effects of evolution are local: evidence from experimental evolution in *Drosophila*. *Integr Comp Biol* 45: 486–491.
- Ruiz M, Sanchez D, Canal I, Acebes A, Ganfornina MD (2011) Sex-dependent modulation of longevity by two *Drosophila* homologues of human Apolipoprotein D, GLaz and NLaz. *Exp Gerontol* 46: 579–589.
- Sanz A, Fernández-Ayala DJM, Stefanatos RKA, Jacobs HT (2010) Mitochondrial ROS production correlates with, but does not directly regulate lifespan in *Drosophila*. *Aging* 2: 200–223.
- Sarup P, Loeschcke V (2011) Life extension and the position of the hormetic zone depends on sex and genetic background in *Drosophila melanogaster*. *Biogerontology* 12: 109–117.
- Savalli VM, Fox CW (1999) The effect of male size, age, and mating behaviour on sexual selection in the beetles *Callosobruchus maculatus*. *Ecol Ethol Evol* 11: 49–60.

- Scannapieco AC, Sambucetti P, Norry FM (2009) Direct and correlated responses to selection for longevity in *Drosophila buzzatii*. *Biol J Linn Soc* 97: 738–748.48.
- Scannapieco AC, Sørensen JG, Loeschcke V, Norry FM (2007) Heat-induced hormesis in longevity of two sibling *Drosophila* species. *Biogerontology* 8:315–325.
- Schneider SA, Schrader C, Wagner AE, Boesch-Saadatmandi C, Liebig J, Rimbach G, Roeder T (2011) Stress resistance and longevity are not directly linked to levels of enzymatic antioxidants in the Ponerine Ant *Harpegnathos saltator*. *PLoS ONE* 6(1): e14601. doi:10.1371/journal.pone.0014601.
- Seehuus S-C, Norberg K, Gimsa U, Krekling T, Amdam GV (2006) Reproductive protein protects functionally sterile honey bee workers from oxidative stress. *Proc Natl Acad Sci USA* 103: 962–967.
- Seto NOL, Hayashi S, and Tener GM (1990) Overexpression of Cu-Zn superoxide dismutase in *Drosophila* does not affect life-span. *Proc Natl Acad Sci USA* 8: 4270–4274.
- Shen J, Ford D, Landis GN, Tower J (2009) Identifying sexual differentiation genes that affect *Drosophila* life span. *BMC Geriatrics* 9: 56.
- Shen J, Tower J (2010) *Drosophila* foxo acts in males to cause sexual-dimorphism in tissue-specific p53 life span effects. *Exp Gerontol* 45: 97–105.
- Shuker DM, Ballantyne GA, Wedell N (2006) Variation in the cost to females of the sexual conflict over mating in the seed bug, *Lygaeus equestris*. *Animal Behav* 72: 313–321.
- Silbermann R, Tatar M (2000) Reproductive costs of heat shock protein in transgenic *Drosophila melanogaster*. *Evolution* 54: 2038–2045.
- Simmons LW, García-González F (2008) Evolutionary reduction in testes size and competitive fertilization success in response to the experimental removal of sexual selection in dung beetles. *Evolution* 62: 2580–2591.
- Simon AF, Shih C, Mack A, Benzer S (2003) Steroid control of longevity in *Drosophila melanogaster*. *Science* 299: 1407–1410.
- Simpson SJ, Raubenheimer D (2009) Macronutrient balance and lifespan. *Aging* 1: 875–880.
- Slack C, Werz C, Wieser D, Alic N, Foley A, Stocker H, Withers DJ, Thornton JM, Hafen E, Partridge L (2010) Regulation of lifespan, metabolism, and stress Responses by the *Drosophila* SH2B protein, Lnk. *PLoS Genet* 6(3): e1000881.
- Sohal RS, Orr WC (2012) The Redox Stress Hypothesis of Aging. *Free Radic Biol Med* 52: 539–555.
- Sørensen JG, Kristensen TN, Kristensen KV, Loeschcke V (2007) Sex specific effects of heat induced hormesis in Hsf-deficient *Drosophila melanogaster*. *Exp Gerontol* 42: 1123–1129.
- South SH, House CM, Moore AJ, Simpson SJ, Hunt J (2011) Male cockroaches prefer a high carbohydrate diet that makes them more attractive to females: implications for the study of condition dependence. *Evolution* 65: 1594–1606.
- Spencer CC, Howell CE, Wright AR, Promislow DEL (2003) Testing an ‘aging gene’ in long-lived *Drosophila* strains: Increased longevity depends on sex and genetic background. *Aging Cell* 2: 123–130.
- Stenesen DS (2011) *In pursuit of a molecular fountain of youth: the identification and characterization of lifespan regulators in Drosophila*. PhD Dissertation, University of Texas.
- Stockhoff BA (1993) Ontogenetic change in dietary selection for protein and lipid by gypsy moth larvae. *J Insect Physiol* 39: 677–686.

- Stojković B, Savković U (2011) Gender differences in longevity in early and late reproduced lines of the seed beetle. *Arch Biol Sci*, Belgrade 63: 129-136.
- Šešlija D, Blagojević D, Spasić M and Tucić N (1999). Activity of superoxide dismutase and catalase in the bean weevil (*Acanthoscelides obtectus*) selected for postponed senescence. *Exp Gerontol* 34: 185-195.
- Šešlija D, Marečko I, Tucić N (2008) Sexual selection and senescence, do seed beetle males (*Acanthoscelides obtectus*, Bruchidae, Coleoptera) shape longevity of their mates? *J Zool Syst Evol Res* 46: 323–330.
- Šešlija D, Tucić N (2008) The genetic architecture of extended life span in the seed beetle *Acanthoscelide obtectus* (Coleoptera: Bruchidae). *Eur J Entomol* 105: 553-560.
- Tanaka Y (1990) Age specificity of inbreeding depression during a life cycle of *Callosobruchus chinensis* (Coleoptera: Bruchidae). *Res Pop Ecol* 32: 329-335.
- Tatar M, Bartke A, Antebi A (2003) The endocrine regulation of aging by insulin-like signals. *Science* 299: 1346-1351.
- Tatar M, Kopelman A, Epstein D, Tu MP, Yin CM, Garofalo RS (2001) A mutant *Drosophila* insulin homolog that extends life-span and impairs neuroendocrine function. *Science* 292: 107–110.
- Toivonen JM, Partridge L (2009) Endocrine regulation of ageing and reproduction in *Drosophila*. *Molec Cellular Endocrinol* 299: 39-50.
- Tower J (2006) Sex-specific regulation of aging and apoptosis. *Mech Ageing Dev* 127: 705–718.
- Tower J, Arbeitman M (2009) The genetics of gender and life span. *J Biol* 8: 38 (doi:10.1186/jbiol141).
- Tower J (2010) Heat shock proteins and *Drosophila* aging. *Exp Gerontol* 46: 355-362.
- Tower J, Landis G, Gao R, Luan A, Lee J, Sun Y (2013) Variegated expression of *Hsp22* transgenic reporters indicates cell-specific patterns of aging in *Drosophila* oenocytes. *J Gerontol A Biol Sci Med Sci* doi:10.1093/gerona/glt078.
- Tu M-P, Epstein D, Tatar M (2002) The demography of slow aging in male and female *Drosophila* mutant for the insulin-receptor substrate homologue *chico*. *Aging Cell* 1: 75–80.
- Tucić N, Cvetković D, Milanović D (1988) The genetic variation and covariation among fitness components in *Drosophila melanogaster* females and males. *Heredity* 60:55-60.
- Tucić N, Gliksman I, Šešlija D, Milanović D, Mikuljanac S, Stojković O (1996) Laboratory evolution of longevity in the bean weevil (*Acanthoscelides obtectus*). *J Evol Biol* 9: 485-503.
- Tucić N, Šešlija D, Stanković V (2004) The short-term and long-term effects of parental age in the bean weevil (*Acanthoscelides obtectus*). *Evol Ecol* 18: 187–201.
- Vaiserman AM, Koshel NM, Litoshenko AY, Mozhukina TG, Voitenko VP (2003) Effects of X-irradiation in early ontogenesis on the longevity and amount of the S1 nuclease-sensitive DNA sites in adult *Drosophila melanogaster*. *Biogerontology* 4: 9–14.
- Vaiserman AM, Zabuga OG, Kolyada AK, Pisaruk AV, Kozeretska IA (2013) Reciprocal cross differences in *Drosophila melanogaster* longevity: an evidence for non-genomic effects in heterosis phenomenon? *Biogerontology* DOI 10.1007/s10522-013-9419-6.
- Vermeulen CJ, Cremers TIFH, Westerink BHC, Van De Zande L, Bijlsma R (2006a) Changes in dopamine levels and locomotor activity in response to selection on virgin life span in *Drosophila melanogaster*. *Mech Ageing Dev* 127: 610-617.

- Vermeulen CJ, Van De Zande L, Bijlsma R (2006b) Developmental and age-specific effects of selection on divergent virgin life span on fat content and starvation resistance in *Drosophila melanogaster*. *J Insect Physiol* 52: 910-919.
- Vermeulen CJ, Loeschcke V (2007) Longevity and the stress response in *Drosophila*. *Exp Gerontol* 42: 153-159.
- Vina J, Sastre J, Pallardo F, Borras C (2003) Mitochondrial theory of aging: importance to explain why females live longer than males. *Antioxid Redox Signal* 5: 549-556.
- Wagner WE, Kelley RJ, Tucke KR, Harper CJ (2001) Females receive a life-span benefit from male ejaculates in a field cricket. *Evolution* 55: 994-1001.
- Wang H-D, Kazemi-Esfarjani P, Benzer S (2004) Multiple-stress analysis for isolation of *Drosophila* longevity genes. *Proc Natl Acad Sci USA* 101: 12610-12615.
- Waskar M, Landis GN, Shen J, Curtis C, Tozer K, Abdueva D, Skvortsov D, Tavaré S, Tower J (2009) *Drosophila melanogaster* p53 has developmental stage-specific and sex-specific effects on adult life span indicative of sexual antagonistic pleiotropy. *Aging* 1: 903-936.
- Wayne ML, Telonis-Scott M, Bono LM, Harshman L, Kopp A, Nuzhdin SV, McIntyre LM (2007) Simpler mode of inheritance of transcriptional variation in male *Drosophila melanogaster*. *Proc Natl Acad Sci, USA* 104: 18577-18582.
- Williams GC (1957) Pleiotropy, natural selection and the evolution of senescence. *Evolution* 11: 398-411.
- Wilson RH, Lai CQ, Lyman RF, Mackay TFC (2013) Genomic response to selection for postponed senescence in *Drosophila*. *Mech Ageing Dev* 134: 79-88.
- Wit J, Sarup P, Lupsa N, Malte H, Frydenberg J, Loeschcke V (2013) Longevity for free? Increased reproduction with limited trade-offs in *Drosophila melanogaster* selected for increased life span. *Exp Gerontol* 48: 349-357.
- Wolf JN, Gemmell NJ (2012) Mitochondria, maternal inheritance, and asymmetric fitness: why males die younger? *BioEssays* 35: 93-99.
- Yee WKW, Sutton KL, Dowling DK (2013) In vivo male fertility is affected by naturally occurring mitochondrial haplotypes. *Curr Biol* 23: R55-56.
- Yılmaz M, Özsoy ED, Bozcuk AN (2008) Maternal age effects on longevity in *Drosophila melanogaster* populations of different origin. *Biogerontology* 9:163-168.
- Zajitschek F, Bonduriansky R, Zajitschek SRK, Brooks RC (2009a) Sexual dimorphism in life history: age, survival, and reproduction in male and female field crickets *Teleogryllus commodus* under seminatural conditions. *Am Nat* 173: 792-802.
- Zajitschek F, Hunt J, Jennions MD, Hal MD, Brooks RC (2009b) Effects of juvenile and adult diet on ageing and reproductive effort of male and female black field crickets, *Teleogryllus commodus*. *Funct Ecol* 23: 602-611.
- Zeh JA, Zeh DW (2005) Maternal inheritance, sexual conflict and the maladapted male. *TRENDS Gen* 21: 281-286.
- Zera AJ (2005) Intermediary metabolism and life history trade-offs: lipid metabolism in lines of the Wing-polymorphic cricket, *Gryllus firmus*, selected for flight capability vs. early age reproduction. *Integr Comp Biol* 45: 511-524.
- Zera AJ, Zhao Z, Kaliseck K (2007) Hormones in the field: Evolutionary endocrinology of juvenile hormone and ecdysteroids in field populations of the Wing-Dimorphic Cricket *Gryllus firmus*. *Physiol Biochem Zool* 80: 592-606.

Chapter 35

**EVIDENCE OF NATURAL AND SEXUAL SELECTION
SHAPING THE SIZE OF NUPTIAL GIFTS
AMONG A SINGLE BUSH-CRICKET GENUS
(*POECILIMON*; TETTIGONIIDAE):
AN ANALYSIS OF SPERM TRANSFER PATTERNS**

***J. McCartney^{1,2,*}, M. A. Potter¹, A. W. Robertson¹,
K-G. Heller² and D. T. Gwynne³***

¹Ecology Group, Institute of Agriculture and Environment, Massey University,
Palmerston North, New Zealand

²Friedrich Alexander Universität, Institute of Biology, Erlangen – Nürnberg, Germany

³Biology Department, University of Toronto in Mississauga, Mississauga, Canada

ABSTRACT

During mating, male bush-crickets transfer a complex spermatophore to the female. The spermatophore is comprised of a large nuptial gift which the female consumes while the sperm from the ejaculate-containing ampulla are transferred into her. Two main functions of the nuptial gift have been proposed: the ejaculate protection hypothesis and the parental investment hypothesis. The former, founded on sexual selection theory, predicts that the time to consume the gift is no longer than necessary to allow for full ejaculate transfer. The latter maintains that gift nutrients increase the fitness or quantity of offspring and hence the gift is likely to be larger than is necessary for complete sperm transfer. With an aim to better understanding the primary function of nuptial gifts, we examined sperm transfer data from field populations of five *Poecilimon* bush-cricket taxa with varying spermatophore sizes. In the species with the largest spermatophore, the gift was four times larger than necessary to allow for complete sperm transfer and is thus likely to function as paternal investment. Species with medium and small gifts were respectively sufficient and insufficient to allow complete sperm transfer and are likely to represent, to various degrees, ejaculate protection. We also found that species that produce larger

* Corresponding Author's Email: J.McCartney@massey.ac.nz.

spermatophores transfer greater proportions of available sperm than species producing smaller spermatophores, and thus achieve higher paternal assurance.

Keywords: ejaculate protection, mating effort, paternal investment, spermatophore size, sperm transfer, sperm competition

INTRODUCTION

Nuptial feeding has been observed in several insect taxa (Thornhill and Alcock, 1983; Vahed, 1998). Male bush-crickets (Tettigoniidae) transfer a substantial, and often costly, spermatophore to the females for consumption during mating (Wedell, 1994a, 1994b; Vahed, 2007a). The spermatophore consists of an ampulla that contains the sperm, and a spermatophylax that, in most species, is a large gelatinous mass. The female first eats the spermatophylax and then eats the smaller ampulla along with any remaining sperm and seminal fluid (Bowen *et al.*, 1984). There is debate over the selective pressures that maintain nuptial gift size in bush-crickets (for reviews see Thornhill and Alcock, 1983; Simmons and Parker, 1989; Vahed, 1998; Gwynne, 2001; Vahed, 2007; Gwynne, 2008; McCartney, *et al.* 2008, 2010, 2012). Despite recent discussions concerning the effect of sexual conflict on nuptial gift size (e.g., Vahed, 2007b; Gwynne, 2008; Lehman, 2012) two hypotheses remain central to understanding the role of gift size: the ejaculate protection hypothesis and the parental investment hypothesis.

The ejaculate protection hypothesis argues that the nuptial gift is sexually selected; it increases fertilisation success by diverting the female away from the sperm ampulla while maximum insemination is achieved (Gerhard, 1913; Boldyrev, 1915; Gwynne, 1984; Sakaluk and Eggert, 1996; Vahed and Gilbert, 1996; Simmons, 2001). The parental investment hypothesis proposes that the function of the nuptial gift is derived from its nutritive value and that these nutrients are passed into the donating males' offspring; the gift is thus under natural selection to increase the quality and/or the quantity of the male's offspring (Trivers, 1972; Thornhill, 1976; Gwynne, 1986, 1988a, 1988b, 1990; Reinhold, 1999).

The ejaculate protection and paternal investment hypotheses are not mutually exclusive (Quinn and Sakaluk, 1986) and present research focuses on the relative importance of the two hypotheses in different taxa. It is likely that the spermatophylax evolved through sexual selection for ejaculate protection in bush-crickets (Gwynne, 1986, 1990, 1997, 2001), but there is evidence that both functions can be involved in the maintenance of spermatophylax size in various tettigoniid species (for reviews see Vahed, 1998; Gwynne, 2001; McCartney *et al.* 2008).

Nuptial gifts that function to protect the ejaculate are predicted to be smaller, less nutritious, and of a size that co-varies with either sperm number and/or ampulla size and should be no larger than necessary to allow for complete insemination (Reinhold and Heller, 1993; Wedell, 1993a, 1994a, 1994b; Heller and Reinhold, 1994; Vahed and Gilbert, 1996). Nuptial gifts that are influenced by paternal investment are likely to be large, nutritious (Wedell, 1994a, 1994b), and take longer to consume than it takes to transfer a full complement of sperm (Wedell, 1994b). While it can be relatively simple to test the prediction of the ejaculate protection hypothesis, at least three further criteria underpin paternal investment in nuptial-gift-bearing species and are needed to distinguish it from the ejaculate protection hypothesis: 1) the

degree of last-male mating advantage; 2) the time that it takes for the nutrients of the spermatophylax to directly affect the donating males' offspring; and 3) the relationship between female mating interval and egg laying interval (see Vahed, 1998 and references cited therein).

The ejaculate protection hypothesis is supported by comparative studies across taxa showing positive correlations between spermatophylax size and ampulla mass or sperm number (Wedell, 1993a; Vahed and Gilbert, 1996; McCartney *et al.*, 2008, 2012), as well as studies within species showing that the size of the nuptial gift or the consumption time of the gift is roughly similar to the time that it takes for the majority of sperm to transfer into the female (e.g., Wedell and Arak, 1989; Wedell, 1991; Reinhold and Heller, 1993; Heller and Reinhold, 1994; Vahed, 1994; Simmons, 1995a). Evidence of paternal investment has also been observed in some species (Gwynne *et al.*, 1984; Gwynne, 1986, 1988a, 1988b; Simmons, 1990; Wedell, 1994a, 1994b; Simmons *et al.*, 1999; Reinhold, 1999), yet almost all insect species studied thus far, including those with properties of paternal investment, have nuptial gifts (or nuptial gift consumption times) that approximate the size necessary for complete sperm transfer (Heller and Reinhold, 1994; Simmons, 1995a; Simmons and Gwynne, 1991; Vahed, 1994), and are therefore likely to be maintained primarily through sexual selection via the ejaculate protection hypothesis (Vahed, 1998). Diverse examples of this rule can be found in Mecoptera as prey and salivary masses, Diptera as nuptial prey and regurgitated food, Coleoptera and Zoraptera as cephalic gland secretions, and other Orthoptera, as hind-wing and glandular secretion feeding (Vahed, 1998 and references cited therein).

Possibly the only exception is *Requena verticalis*, initially reported to have a spermatophylax twice as large as necessary to allow for complete sperm transfer of the ampulla (Gwynne *et al.*, 1984; Gwynne, 1986, 1988b). However, further research on this species (Simmons, 1995a, 1995b; Simmons *et al.*, 1999) and different interpretations of what constitutes 'complete' sperm transfer (Vahed, 1994, 1998; Simmons, 1995a) suggest that complete sperm transfer may not be achieved until close to, or even after gift consumption (Vahed, 1998). Additionally, males have a substantial first-male paternity advantage (Gwynne, 1988b; Simmons and Achmann, 2000; Simmons *et al.*, 2007) and variable spermatophylax sizes, perhaps as a result of variability in female availability, re-mating interval (Simmons, 1995b), and sexual status (Simmons *et al.*, 1993). At times, therefore, gift size approximates the size necessary for complete sperm transfer.

In order to better understand the relationship between nuptial gift size and sperm transfer pattern and the selective pressure that most influences its variation, there is perhaps no better model than the bush-cricket genus *Poecilimon* (Tettigoniidae) (McCartney *et al.* 2008, 2012). This genus contains species with a large diversity in mating behaviours. Comparisons among species within genera can be particularly useful as characters shared by congeners are often held constant and thus control to a large degree for similarities that may be caused by relatedness (Harvey, 1991; Harvey and Pagel, 1991). With around 140 described *Poecilimon* species (Eades and Otte, 2008), the variation in nuptial gift size is unmatched among Orthoptera and approaches the magnitude of family-wide variation (McCartney *et al.*, 2008), with spermatophore size varying from 6.1% (*Poecilimon laevissimus*) to 37% (*P. thessalicus*) of the relative body mass of the male (McCartney *et al.*, 2008). This clearly represents large variation in male reproductive investment.

Few bush-cricket studies have investigated nuptial gift function from a sperm transfer perspective and, of these most have used laboratory-reared individuals despite concerns about the validity of this approach (see McCartney *et al.*, 2008 for discussion). Even fewer studies

still, have considered sperm transfer patterns within field populations (eg. Heller & von Helversen, 1991; Reinhold, 1994; Vahed and Gilbert, 1996). Furthermore, interpretation of data has been complicated by the diversity of taxa involved; variations in sperm transfer may ultimately be linked to taxon differences and not nuptial gift size *per se* (for discussion see Gwynne, 1995; Vahed and Gilbert, 1996; McCartney, 2010).

Our aim here was to better understand the premise that nuptial gift size relates to function. First, in order to assess the match between nuptial gift consumption time and optimum sperm transfer time across closely related species with marked variation in nuptial size, we combined published sperm transfer and nuptial gift consumption time data from two field-observed *Poecilimon* taxa that produce medium and large gifts (Reinhold and Heller, 1993, Heller & Reinhold, 1994), with sperm transfer and gift consumption data from three novel field-observed *Poecilimon* species; two with small gifts and one with very large gifts. A close match between gift consumption and sperm transfer would be consistent with the sperm protection hypothesis, whereas if complete sperm transfer occurs long before spermatophylax gift consumption is completed, we have grounds to infer a paternal investment function. Secondly, we controlled for body mass and relatedness, and compared spermatophore size between species to the proportion of sperm that has transferred into the female by the time she has consumed the spermatophore. A significant relationship would indicate that males of *Poecilimon* taxa that produce larger spermatophores have increased confidence of sperm transfer, and thus paternal assurance, compared to taxa producing smaller spermatophores.

MATERIALS AND METHODS

Species and Sites

Poecilimon is a genus of bush-crickets (Phaneropterinae, tribe Barbistini) (Orthoptera: Ensifera: Tettigoniidae), with about 65 European species that are mostly situated in the east Mediterranean (Heller, 2004). Three species, *Poecilimon laevisissimus* (Fischer, 1853), *P. erimanthos* Willemse and Heller, 1992, and *P. thessalicus* Brunner von Wattenwyl, 1891, were chosen to represent the genus in this study, as a previous study found that these species had some of the largest differences in relative spermatophore size and sperm number within the genus (McCartney *et al.*, 2008). The spermatophore sizes of *P. laevisissimus* and *P. thessalicus* represent the upper and lower limits, with *P. erimanthos* producing a small to medium-sized spermatophore of 7.2% relative mass (McCartney *et al.*, 2008). Sperm number from single matings range between 90,000 and 140,000 - 210,000 for *P. laevisissimus* and *P. erimanthos* respectively, and up to about 14,500,000 in *P. thessalicus* (McCartney *et al.*, 2008). Data for two further species, *P. v. minor* and *P. v. veluchianus* were obtained from the literature because these species represent medium to large-size spermatophores and sperm numbers respectively (Reinhold and Heller, 1993, Heller & Reinhold, 1994). All species examined here are nocturnal except *P. erimanthos* which is diurnal and mates during the day.

Any important differences between the methods used on the novel species presented here, *P. laevisissimus*, *P. erimanthos* and *P. thessalicus*, and previously published species, *P. v. veluchianus* and *P. v. minor*, are outlined below. However, see Reinhold and Heller (1993) and Heller and Reinhold (1994) for detailed methods on *P. v. veluchianus* and *P. v. minor*.

Fieldwork on all novel species was carried out during the summers of 1990, 1997 and 1998 on the Peloponnese Peninsula and mainland Greece. *Poecilimon erimanthos* and *P. laevisissimus* were observed at Erimanthos Valley (east of the village of Kumani, N. Elia, 37°46'N, 21°47'E.), and *P. thessalicus* at a site inland from Katerini (north-west of the village of Elatochori, 40°19'N, 22°15'E). Both sites were semi-pastoral with forest margins, and population borders were demarcated by roads, forests or cliffs.

Spermatophore Consumption Time, Male Body Mass and Spermatophore Mass

All measurements on spermatophore consumption of novel species were taken from field observations of marked (*P. erimanthos*) or non marked animals (*P. laevisissimus* and *P. thessalicus*) throughout their mating season. Captured animals were paired in containers or hanging mesh cages in the field. Male and female *P. laevisissimus* were captured as sub-adults and allowed to mature for around seven days before pairing to allow for full development of the accessory glands (males) and full receptivity (Heller and Helversen, 1991; see Reinhold & Heller, 1993; McCartney *et al.*, 2008 for discussion on cage and laboratory effects in *Poecilimon*). To minimize disturbance of females, observations of spermatophore consumption progress were made at intervals rather than continuously. *Poecilimon laevisissimus* and *P. thessalicus* were only used in observations after witnessing the onset of spermatophore consumption, whereas we estimated onset for *P. erimanthos* as half of the interval between the first observation of a female without a spermatophore, and again with a spermatophore (females observed about every hour). Spermatophore consumption times of all species were also estimated as half of the interval between the observation of the female last seen with a spermatophore, and subsequently without a spermatophore.

Spermatophore consumption time and male body mass were measured in the 1997 and 1998 breeding seasons and pooled for *P. thessalicus* (data did not differ significantly between years; spermatophore consumption time, $t_{14} = -0.561$; $p = 0.584$; male body mass $t_{66} = -1.501$; $p = 0.138$). Spermatophore masses for *P. thessalicus* are reported from 1998. Measurements of spermatophore consumption time, male body mass and spermatophore mass for *P. laevisissimus* are reported from 1997. Spermatophore consumption times were recorded for *P. erimanthos* in 1990 and the male body mass and spermatophore mass were recorded in 1997.

Sperm Transfer

Poecilimon thessalicus and *P. laevisissimus* were observed in 1998 whereas *P. erimanthos* was observed in 1997 and 1998. *Poecilimon erimanthos* (in 1997) and *P. thessalicus* (in 1998) were observed at the locations where they were collected. In 1998, we collected approximately 50 sub-adult *P. laevisissimus* and *P. erimanthos* east of the village of Kumani, N. Elia and took them to Central Greece, where we made further caged observations.

All bush-crickets taken from the field were sub-adults and were stored separately by sex and species, then allowed to mature for at least seven days. We allowed mating of 20 to 30 virgin pairs of each species. Mated females were allocated randomly to predetermined spermatophore attachment times that were set at intervals relative to the spermatophore

consumption time in order to determine the rate of sperm transfer. For each species, the duration of the first sperm transfer trial was set to equal the average spermatophore consumption time for that species (see Table 1). All mated females, except some *P. erimanthos* in 1997, were assigned randomly to a pre-determined transfer time for examination. For *P. laevisissimus* and *P. thessalicus* we tested sperm transfer times at appropriately equal periods either side of the average spermatophore consumption time, and repeated this until we had adequately covered the full period from no transfer until (near) full transfer (*P. thessalicus* = 120, 240, 480, 780, 1020, 1260 min intervals, *P. laevisissimus* = 60, 120, 180, 240 min. intervals). The spermatophores of *P. erimanthos* in 1997 were removed at various intervals between 30-80 min., with two distinct modes of 35 min and 75 min. This meant the mean number of sperm that had transfer in six observations between 30-35 min., and six observations between 45-80 min. were pooled into two groups at 35 min. and 80 min. and the mean sperm transfer value was used for each. The spermatophores of *Poecilimon erimanthos* in 1998 were removed at 1 min., 120 min., and 240 min and combined with the data of 1997 (35 min. and 80 min.).

Immediately after mating, each female was placed head-first into a large scintillation tube to prevent her from bending to remove the ampulla. We then stored the females in a cool, shaded area and males were returned to cages. After the assigned period, each female was removed from her vial and the spermatophore removed by grasping the ampulla at its base with dissecting forceps and pulling it carefully from her genital pore and the spermatheca was excised. The female was killed and the spermatheca and the ampulla were stored in separate vials with a known volume of water for sperm counting (1-5 ml depending on the structure's size). If sperm ampullae became semi-detached or sperm had drained outside the female these data were not used in the analysis.

Each ampulla and spermatheca was macerated with a scalpel and mixed by passing it repeatedly through a syringe until the sperm had been suspended in the water and the sample homogenised. A sub-sample was placed on a haemocytometer slide (Swift: Neubauer improved). Sperm from a minimum volume of 50 μ l (or up to 200 μ l) were counted and multiplied by the appropriate dilution factor to give the total number of sperm per spermatheca. Five sub-samples were taken and the solution was remixed before each new sub-sample was taken. From total sperm (ampulla and spermatheca) we derived the percentage of sperm within each mating transferred from the spermatophore into the spermatheca.

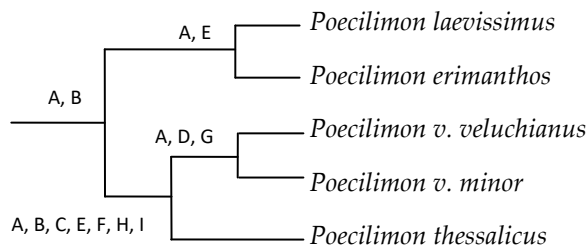
ANALYSIS

Sperm Transfer and Spermatophore Consumption

In order to compare the match/mismatch of complete sperm transfer and spermatophore consumption of all novel species, average spermatophore consumption times of all species were overlaid on a time-course chart of sperm transfer. In an attempt to compare the sperm transfer profiles of the three species presented here we spent considerable effort fitting regression models to sperm transfer patterns, and were not convinced that they could either reliably resolve the shape of sperm transfer curves, or validly explain the behaviour of sperm transferring into the female. Ultimately, no model we used could clarify the sperm transfer relationship between different species (see

In each species there were mating attempts resulting in no sperm transferring. These data were not included in analyses but are discussed further. Data were analysed using SAS 9.1. The analyses of *P. v. veluchianus* and *P. v. minor* were as recorded in Reinhold and Heller (1993) and Heller and Reinhold (1994).

Regression analyses on relative spermatophore mass against the proportion of sperm that had transferred into the female were first performed across taxa. All proportion data were arcsine (square root) transformed and tested for normality. In conjunction with this regression analysis, corresponding regression analyses were also performed on transformed proportion data with phylogenetic independent contrasts in order to control for relatedness (Felsenstein, 1985). While this method is typically preferred over standard linear regression analyses across species, sample sizes are reduced further using contrasts (n-1) and so have less power.



Phylogenetic Independent Comparisons

A cladogram of the species used in this study was constructed using the literature on the phylogeny for these *Poecilimon* taxa (see Figure 1 for references) and the computer package PDAP (Maddison & Maddison, 2006) (Figure 1). The proportion of sperm and relative spermatophylax data were added to the tree in order to calculate phylogenetically-independent contrasts. In all cases, branch lengths were set to 1. The contrasts were then standardised by dividing them by the variance (square root of the sum of the branch length, Felsenstein (1985)). Generalised linear models were then used to regress the standardised independent contrasts of

relative spermatophore size against the standardised independent contrasts of the proportion of transferred sperm response variable. All inferential regressions involving phylogenetically-independent contrasts were forced through the origin (Garland *et al.* 1992).

RESULTS

Spermatophore Consumption and Sperm Transfer

The spermatophores (spermatophylax and ampulla) were consumed in 101 ± 10.7 min (range 30-165 min., $n = 14$) for *P. laevissimus* (Table 1), a period too short to allow more than a small portion of the sperm to transfer into the female (Figure 2). Only about 15% of available sperm had transferred during any of the observations made prior to the last observations at 240 min (2.4 times longer than the mean spermatophore consumption period). The four observations at this longest interval revealed that a large amount of sperm still remaining in the ampulla and therefore the spermatophylax appears to be much smaller than is necessary to ensure complete sperm transfer.

Spermatophores in *P. erimanthos* (1990) were consumed in 84 ± 3.5 min (range 55-130 min, $n=39$, Table 1). This corresponds with a peak in sperm transfer, and more than 50% of sperm had transferred to the spermatheca by this time (Figure 2). After the time usually required for spermatophore consumption, sperm transfer seemed to slow down reaching about 75% of the total transfer after 240 min. Thirty five minutes after mating no sperm had been transferred ($n=9$) yet after this time all except one female (that was discarded because she was found with no sperm after 4 hours) contained over 50% ($n=18$) of the available sperm. So a fast transfer process occurs in this species and takes between 35 and 60 min, and is complete just before the spermatophore is normally consumed, indicating that the spermatophore may be of about the correct size for optimum sperm transfer (about 60% of total sperm).

Pooled data for *P. thessalicus* from both years gave a spermatophore consumption time of 15.7 h (943 ± 47.6 min, $n=16$) (Table 1). The sperm transfer pattern of *P. thessalicus* differs from that in the two previous species, in that peak transfer occurred between 13-25% of mean spermatophore consumption time (240 min., Figure 2), and 93% of sperm had transferred by the end of spermatophore consumption. There was a clear plateau in sperm transfer in *P. thessalicus* at around 90-95% of total available sperm and therefore females were inseminated nearly four times more quickly than required for spermatophylax consumption. Even the fastest spermatophore consumption, of about 710 min, would have allowed around 93% of the sperm to transfer by the time one third of the spermatophore was consumed. Five out of 26 matings (19.2%) did not release any sperm into the female after transfer onset (one female at each of 240, 780, and 1260 min and two females at 480 min) and, since all other pairings resulted in close to, or above, 90% sperm transfer, these were not included in calculations of the means or standard errors. Interestingly, the average number of sperm in the ampullae that failed to transfer any sperm was only 8.3 million ($n=5$, S.E.=2.9 million, range = 2.3-17.8 million), significantly fewer than the 22.6 million ($n=22$, S.E.=21 million, range = 0.05-37.3) in spermatophores that did transfer (Mann-Whitney rank analysis $U=27$, $P < 0.007$).

Table 1. Male body mass, spermatophore consumption time (min) and absolute and relative spermatophore mass in three species of *Poecilimon* studied here (mean $\pm$ S.E. (range: n); upper three rows) and two sub species taken from Reinhold & Heller (1993), (lower two rows)

Species	Spermatophore consumption time (min) (range: n)	Spermatophore mass (mg)	Male body mass (mg)	Relative spermatophore size
<i>P. erimanthos</i>	84 $\pm$ 3.5 (55-135: 39)	47 $\pm$ 3 (n=11)	640 $\pm$ 4 (n=25)	7.2% (n=11)*
<i>P. laevisissimus</i>	101 $\pm$ 10.7 (30-165:14)	47 $\pm$ 6 (n=9)	781 $\pm$ 13 (n=50)	6.1% (n=9)*
<i>P. thessalicus</i>	943 $\pm$ 47.6 (710-1380: 16)	112 $\pm$ 8 (n=28)	440 $\pm$ 7 (n=68)	33% $\pm$ 2..34% (n=17)
<i>P. veluchianus veluchianus</i>	570	162	640	24.9%
<i>P. v. minor</i>	200	74	365	19.1%

*No S.E. available because relative spermatophore mass was taken from dividing the average of pooled spermatophore mass from the average of pooled male body mass.

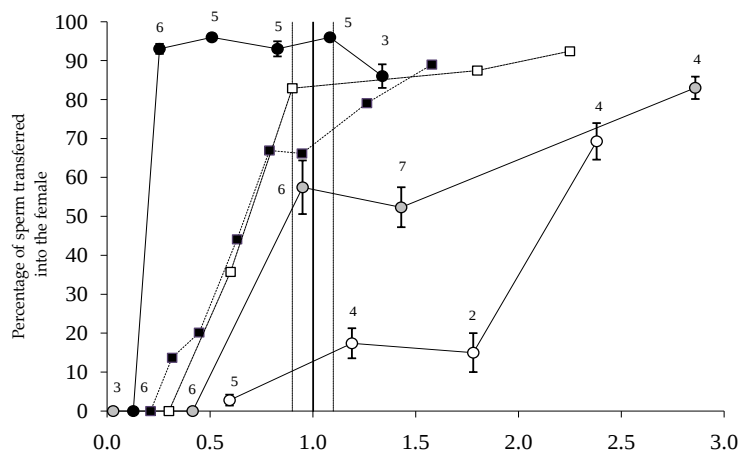


Figure 2. Percentage of sperm transferred after copulation from the male ampulla to the female spermatheca ($\pm$ S.E.) plotted relative to the mean spermatophore consumption time, (numbers above points = n). Novel species are represented by unbroken lines: *P. laevisissimus* (open circles), *P. erimanthos* (grey circles), and *P. thessalicus* (black circles). Broken lines represent *P. v. veluchianus* (closed squares) and *P. v. minor* (open squares) calculated from the published data (S.E and n not presented; for details see Heller and Reinhold (1994)). Dashed vertical lines show one SD in consumption time for *P. thessalicus* (the species with the largest SD).

Relative Spermatophore Mass and Proportion of Sperm Transferred

No significant relationship was found between spermatophore size and the proportion of sperm that had transferred into the female by the spermatophore consumption time ($F_{1,4} = 7.69$, $p = 0.069$, $r^2 = 0.72$). While this was not strengthened while controlling for relatedness ($F_{1,3} = 3.06$, $p = 0.179$), a strong relationship is apparent (Figure 3). An increase in sample size is likely to produce a significant effect; males of larger spermatophore-producing *Poecilimon* taxa are likely to transfer a greater proportion of sperm than species producing smaller spermatophores.

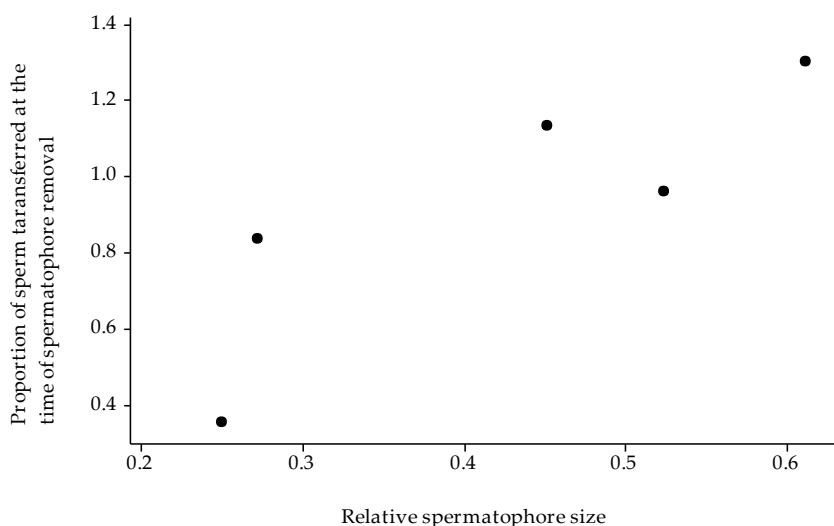


Figure 3. Spermatophore size compared to the proportion of sperm that have transferred into the female at the time of spermatophore consumption among five *Poecilimon* taxa. NB. Data are arcsine (square root) transformed.

DISCUSSION

The percentage of sperm that had transferred into the female by the time it took her to consume and remove the spermatophore differed markedly between the five species. The match/mismatch between complete sperm transfer and spermatophore consumption time found across our species correspond to our predictions that nuptial gifts of different sizes are affected by ejaculate protection and paternal investment to different degrees. Large nuptial gifts in *Poecilimon* are apparently either of correct size or larger than necessary to allow for a full transfer of sperm, whereas small nuptial gifts seem to be less than capable of protecting the ejaculate and allowing the complete complement of sperm to be transferred. Sexual selection for larger spermatophores in *Poecilimon* is likely to increase male confidence in sperm transfer (McCartney *et al.*, 2008, 2010) and correspond to greater level of courtship related female mating investment (McCartney *et al.*, 2012).

Poecilimon laevis and *P. erimanthos* have small spermatophylaxes which seem to be either smaller than necessary for sperm transfer, or have consumption times marginally correspondent with the time it takes for sperm to transfer into the female; thus they are likely to function primarily as ejaculate protection. *Poecilimon thessalicus*, on the other hand, has one of the largest spermatophores reported (McCartney *et al.*, 2008), and a nuptial gift almost four times larger than necessary for complete sperm transfer. It is therefore likely to function as both ejaculate protection and paternal investment.

It may be assumed that *Poecilimon* species with a similarly large spermatophore size as *P. thessalicus* will also have similarly long consumption times. This, however, does not appear to be the case in either of the subspecies of *P. veluchianus* which also have large nuptial gifts but comparatively quick consumption times (Table 1, Figure 2). While *P. thessalicus* and *P. veluchianus* indeed have larger gifts, the difference does not seem to lie in the speed at which

the sperm transfers into the female, but rather with the extended period over which female *P. thessalicus* consume nuptial gifts. This point is important when understanding a key assumption of the paternal investment hypothesis; males must invest in their own offspring. Lengthened consumption time increases ejaculate transfer which delays the speed at which a female remates and increases the time in which nutrients of the donating male's gift can be incorporated in his offspring. Reasons for an extended feeding time in *P. thessalicus* are thus far unknown, however the study population was at relatively high altitude (ca. 1,100 m a.s.l.), with night time temperature often at around 10-15°C, compared to *P. v. veluchianus* and *P. v. minor* (330 m a.s.l.; Reinhold and Heller, 1993) and *P. erimanthos* and *P. laevissimus* (around 600 m a.s.l.) where night temperatures are typically 20°C or above (unpubl. data). The metabolism of *P. thessalicus* at these temperatures is likely to be lower than that of the other species, resulting in consumption duration and digestion times of nuptials gift being significantly slower. However, temperature differences are unlikely to have affected our results because spermatophores are costly to produce and are evolutionary labile (McCartney *et al.*, 2008); males would be expected to allocate fewer resources to gift production – to a size more appropriate to ejaculate protection – if there were no fitness benefits to having a proportionately large spermatophylax gift. A further explanation for the slow gift consumption time of *P. thessalicus* may be related to possible bitter substances in the spermatophylax of *P. thessalicus*. These may affect the speed at which females are able to consume the nuptial gift (as suggested by Heller *et al.*, 1998) but further work is needed in order to verify the substances, their palatability, and the effect they have on females.

While there is a match between nuptial gift consumption and sperm transfer times in *P. v. veluchianus* and *P. veluchianus minor* these species have nuptial gift sizes that further correspond to our predictions that larger gifts are influenced by paternal investment. *Poecilimon v. veluchianus* has a substantial last male mating advantage (Achmann *et al.*, 1992) and nutrients from the nuptial gift of the donating male are likely invested into his own offspring (McCartney 2010). Evidence of paternal investment in this species comes from a correlation of nuptial gift size on the dry mass of the donating males' offspring, and a greater lifespan of starved offspring (immediately after eclosion) fathered by males with large spermatophores (Reinhold, 1999).

Typically it takes 3-4 days for the nutrients in nuptial gifts to be incorporated in egg batches in the female (Bowen *et al.*, 1984; Gwynne and Brown, 1994; Wedell, 1993b; Simmons and Gwynne, 1993; Reinhold, 1999, although see Voigt *et al.*, 2006, 2008). While sperm precedence patterns have only been analysed in two *Poecilimon* species (*P. veluchianus*; Achmann *et al.*, 1992 and *Poecilimon hoelzeli*; Achmann, 1996), both show a last male mating advantage. The combination of these factors in both *P. erimanthos* and *P. laevissimus* is, however, likely to exclude the possibility of paternal investment; both species remate, on average, every 1-2 days and lay eggs every two days (McCartney 2010). It is therefore unlikely that there is sufficient time for gift nutrients to be incorporated into the donating male's offspring. In contrast, field observations from *P. thessalicus* suggest that females may have extended inter-mating refractory periods of about 7-8 days (and up to 19 days) and lay eggs every 1-2 days (McCartney, 2010), so males are likely to have their nutrients incorporated into the majority of eggs before females remate.

Transfer of a full ejaculate is necessary to ensure optimum fertilisation for males especially in polyandrous species (Smith, 1984). It is difficult therefore to understand why, in the two species that produce smaller spermatophores, males do not protect their ejaculate with larger

nuptial gifts. In *P. erimanthos* and *P. laevisissimus* females consumed 48% and 87% respectively of the sperm that males produced. While *P. laevisissimus* seemed to remove and eat the spermatophore nearly eight times faster than expected for maximum sperm transfer, sperm from *P. erimanthos* transferred into the female at a rate that arguably approximated spermatophore consumption time, but still resulted in a waste of a large portion of sperm. Similarly, 9% of spermatophores are estimated to be prematurely consumed in *P. v. veluchianus* (Reinhold and Heller, 1993). It is likely that there is conflict between the sexes over optimum sperm number and resulting spermatophore attachment duration. Premature removal of the ampulla may constitute a form of post-copulatory female discrimination (Sakaluk and Eggert, 1996), but it is unlikely that such a high number of matings observed here resulted in removal discrimination. Sperm loading, the adjustment of copulation duration and ejaculate size according to the risk of sperm competition (Parker *et al.*, 1990), has been observed in some other insects species (see for example Dickenson, 1986; Garcia-Gonzalez and Gomendio, 2004) including bush-crickets *Uromenus rugosicollis* (Vahed, 1997), and may well be a feature in some *Poecilimon*. Males may produce an optimum number of sperm ideal for sperm competition but in *P. laevisissimus*, females may “have the edge” over this conflict by being able to consistently consume and remove the nuptial gift and sperm ampulla before the sperm is fully transferred (reviewed in Vahed, 2007b; Gwynne, 2008). This assertion of a conflict between the sexes is further corroborated by evidence in *P. laevisissimus* where the pairs struggle for some time as the females appear to try and escape the clasp of the male’s cerci, and may additionally represent a form of female discrimination that leads to ‘fit’ males transferring more sperm overall (Eberhard, 1996).

In a different form, large quantities of sperm and spermatophore material are also wasted in *P. thessalicus*. We found that a large proportion of males did not transfer any sperm (18.5%; $n=27$). Spermatophores are expensive to produce (Dewsbury, 1982; Drummond, 1984; Simmons, 1990, 1995a; Heller and von Helversen, 1991; Vahed, 2007b; Lehmann, 2012), so those that fail to initiate represent a considerable waste in time and energy to *P. thessalicus* males. It may be that constriction of the females in scintillation tubes affected the onset of sperm transfer in *P. thessalicus*, although this is unlikely as onset was not affected in *P. laevisissimus*, *P. erimanthos* and a previously studied species, *P. hoelzeli* (R. Achmann *pers. comm.*). Importantly, total sperm numbers in ampullae that did not transfer were much less (by 63%) than the total number of sperm in ampulla that did transfer. While these data suggest that the mechanical initiation of sperm transfer may be dependent on the internal pressure or volume of sperm or ejaculate, mechanisms behind the sperm transfer process are poorly understood in bush-crickets. Future studies would benefit by further assessment of sperm transfer initiation during this critical onset period (Achmann *et al.*, 1992; Reinhold and Heller, 1993; Simmons and Achmann, 2000; Simmons, 2001).

Ultimately, no model we used for analysis could clarify the sperm transfer relationship between different species. Vahed (1994), however, previously fitted models using data from Gwynne *et al.* (1984) and Gwynne (1986) and showed that there was no difference between the sperm transfer curves for *Leptophyes punctatissima* and *Requena verticalis*, two species with varied sperm transfer profiles. Vahed (1994) suggested that the variation found within the sperm transfer among individuals of each species may be too large to easily detect a difference among species, although ultimately concluding that the function of the spermatophylax in *R. verticalis* is likely the same as that for *L. punctatissima*; to protect the ejaculate. As a comparison, we adopted the model used by Vahed (1994) and similarly found no difference

between the sperm transfer curves of the two most different *Poecilimon* species (i.e., *P. thessalicus* and *P. laevissimus*, $S=0.261$, $P=0.61$). We therefore suggest that the variation found within the sperm transfer among individuals of each species is too large to detect a difference between species and that the curves are unlikely to be considered the same.

It is important to keep in mind that the function of the nuptial gift is influenced by substances in the ampulla, other than sperm, that are transferred during mating (McCartney *et al.* 2008; McCartney, *et al.* 2010). Some of these substances are known to influence female intermating refractory period (Heller & Helversen, 1991; Heller & Reinhold, 1994; Lehmann and Lehmann, 2000b; Vahed, 2007b), the timing of oviposition (Arnqvist and Rowe, 2005; Vahed, 2007b), and the share of eggs that are laid with the donating male's nutritional investment (Simmons 1990; Vahed, 2003). Indeed, the positive relationship we found between spermatophore size and the proportion of sperm transferred may tie closely to the total volume of ejaculate substances transferred. If these substances affect fertilisation success or the incorporation of nutrients into offspring, the size or function of the nuptial gift may instead vary in accordance with these and be an important factor governing gift size (Vahed, 2003, 2007a; McCartney *et al.*, 2008).

It is unlikely that male *P. erimanthos* or *P. laevissimus* make significant paternal investments in their offspring in terms of nutrients. While paternal investment has been directly observed in *P. v. veluchianus* (Reinhold 1999), the disparity in time between complete sperm transfer and spermatophore consumption in *P. thessalicus* is also best explained by paternal investment. Larger spermatophores apparently increase male confidence in sperm transfer – and perhaps total ejaculate transfer – and are likely to ensure a greater level of paternal assurance. Furthermore, a recent study has shown that females of *Poecilimon* species that have males that invest more in spermatophore production will compete for access to males, invest relatively more in mating effort, and take greater risks in finding mates than species with smaller spermatophores (McCartney *et al.* 2012).

In terms of ejaculate protection and paternal investment, we present evidence that both sexual selection and natural selection influence spermatophore size within the single bush-cricket genus *Poecilimon*. However, irrespective of function, it is clear that sperm is wasted in all species presented here, and a better understanding is needed of the cost of sperm production as well as the mechanisms which affect sperm transfer if we are to fully understand the relationship between nuptial gift size, paternal investment, and ejaculate protection. Future studies would do well to assess how other substances in the ejaculate may control female re-mating, ova production and oviposition rate, and how the transference of these substances relate to gift consumption time.

ACKNOWLEDGMENTS

We thank K. Teltscher, L. Penny, M. Volleth and K. Witt for help in the field, K. Reinhardt, K. Reinhold and W. Edrich for discussion and, K. Vahed and M. Rossiter for helpful comments on earlier versions of the manuscript. Our research was supported by Massey University and D.F.G. (Deutsche Forschungsgemeinschaft) scholarships.

REFERENCES

- Achmann R. 1996. *Individueller Paarungs- und Fortpflanzungserfolg bei der Laubheuschrecke Poecilimon hoelzeli: Ein molekulargenetischer Ansatz* [Individual mating and reproduction success in the grasshopper *Poecilimon hoelzeli*: A molecular genetic approach]. PhD thesis. Friedrich-Alexander-Universität.
- Achmann R, Heller K-G, Epplen JT. 1992. Last-male sperm precedence in the bushcricket *Poecilimon veluchianus* (Orthoptera, Tettigoniidae) demonstrated by DNA fingerprinting. *Molecular Ecology* 1: 47-54.
- Arnqvist G, Rowe C. 2005. *Sexual conflict (Monographs in Behavior and Ecology)*. Princeton, NJ: Princeton University Press.
- Boldyrev BT. 1915. *Contributions à l'étude de la structure des spermatophores et des particularités de la copulation chez Locustodea et Gryllodea* [Contributions to the study of spermatophore structure and copulation characteristics in Locustodea and Gryllodea]. *Horae Societatis Entomologicae Rossicae* 6: 1-245.
- Bowen BJ, Codd CG, Gwynne DT. 1984. The katydid spermatophore (Orthoptera: Tettigoniidae): male investment and its fate in the mated female. *Australian Journal of Zoology* 32: 23-31.
- Dewsbury DA. 1982. Ejaculate cost and male choice. *American Naturalist* 119: 601-610.
- Dickinson JL. 1986. Prolonged mating in the milkweed leaf beetle *Labidomera clivicollis* (Coleoptera: Chrysomelidae): a test of the "sperm-loading" hypothesis. *Behavioral Ecology and Sociobiology* 18: 331-338.
- Drummond BA ed. 1984. *Multiple mating and sperm competition in the Lepidoptera*. New York: Academic Press.
- Eades DC, Otte D. 2008. *Orthoptera Species File Online*. Version 2.0/3.4. <http://Orthoptera.SpeciesFile.org>.
- Eberhard WG. 1996. *Female control: sexual selection by cryptic female choice*. Princeton, NJ: Princeton University Press.
- Felsenstein J. 1985. Phylogenies and the comparative method. *American Naturalist* 125: 1-15.
- Garcia-Gonzalez F, Gomendio M. 2004. Adjustment of copula duration and ejaculate size according to the risk of sperm competition in the golden egg bug (*Phyllomorpha laciniata*). *Behavioral Ecology* 15: 23-30.
- Gerhardt U. 1913 Copulation und Spermatophoren von Grylliden und Locustiden I [Copulation and spermatophore of Grylliden and Locustiden I]. *Zoologische Jahrbucher (Syst.)* 35: 415-532.
- Gwynne DT. 1984. Sexual selection and sexual differences in mormon crickets (Orthoptera: Tettigoniidae, *Anabrus simplex*). *Evolution* 38: 1011-1022.
- Gwynne DT. 1986. Courtship feeding in katydids (Orthoptera, Tettigoniidae) - investment in offspring or in obtaining fertilizations. *American Naturalist* 128: 342-352.
- Gwynne DT. 1988a. Courtship feeding and the fitness of female katydids (Orthoptera: Tettigoniidae). *Evolution* 42: 545-555.
- Gwynne DT. 1988b. Courtship feeding in katydids benefits the mating males Offspring. *Behavioral Ecology and Sociobiology* 23: 373-377.

- Gwynne DT. 1990. The katydid spermatophore: evolution of a parental investment. In: Bailey WD and Rentz DC, eds. *The tettigoniidae: biology, systematics and evolution*. Bathurst: Crawford house press. 27-240.
- Gwynne DT. 1997. The evolution of edible 'sperm sacs' and other forms of courtship feeding in crickets, katydids and their kin (Orthoptera: Ensifera). In: Choe JC and Crespi BJ, eds. *Evolution of mating systems in insects and arachnids*: University of Cambridge. 110-129.
- Gwynne DT. 2001. *Katydids and bush-crickets: reproductive behavior and evolution of the tettigoniidae*. Cornell University Press: Ithaca.
- Gwynne DT. 2008. Sexual conflict over nuptial gifts in insects. *Annual Review of Entomology*. 53: 83-101.
- Gwynne DT, Bowen BJ, Codd CG. 1984. The function of the katydid spermatophore and its role in fecundity and insemination ((Orthoptera: Tettigoniidae). *Australian Journal of Zoology* 32: 15-22.
- Gwynne DT, Brown WD. 1994. Mate feeding, offspring investment, and sexual differences in katydids (Orthoptera: Tettigoniidae). *Behavioral Ecology* 5: 267-272.
- Harvey PH. 1991. Comparing uncertain relationships: The Swedes in revolt. *Trends in Ecology & Evolution* 6: 38 - 39.
- Harvey PH, Pagel MD. 1991. *The comparative method in evolutionary biology*. Oxford University Press: Oxford.
- Heller K-G. 1984. To the bioacoustics and phylogeny of the genus *Poecilimon* (Orthoptera, Tettigoniidae, Phaneropterinae). *Zoologische Jahrbücher. Systemologie* 111: 69-117.
- Heller K-G. 1990. Evolution of song pattern in east Mediterranean Phaneropteridae: constraints by the communication system. In: Bailey WJ and Rentz DCF, eds. *The Tettigoniidae: Biology, Systematics and Evolution*. Bathurst, Australia: Crawford House. 130-151.
- Heller K-G. 1992. Risk shift between males and females in the pair-forming behavior of bush-crickets. *Naturwissenschaften* 79: 89-91.
- Heller K-G. 2004. Fauna europaea: insecta: phaneropteridae. fauna europaea version 1.1 *Fauna Europaea*. 1.1 ed. <http://www.faunaeur.org>.
- Heller K-G, Faltin S, Fleischmann P, von Helversen O. 1998. The chemical composition of the spermatophore in some species of phaneropterid bush crickets (Orthoptera: Tettigoniidae). *Journal of Insect Physiology* 44: 1001-1008.
- Heller K-G, Korsunovskaya O, Sevgili H, Zhantiev RD. 2006. Bioacoustics and systematics of the *Poecilimon heroicus* group (Orthoptera: Phaneropteridae: Barbitistinae). *European Journal of Entomology*. 103: 853-865.
- Heller K-G, Reinhold K. 1994. Mating effort function of the spermatophore in the bushcricket *Poecilimon veluchianus* (Orthoptera, Phaneropteridae): support from a comparison of the mating behaviour of two subspecies. *Biological Journal of the Linnean Society* 53: 153-163.
- Heller K-G, von Helversen D. 1991. Operational sex-ratio and individual mating Frequencies in two bush-cricket species (Orthoptera, Tettigoniidae, *Poecilimon*). *Ethology* 89: 211-228.
- Lehmann AW. 1998. *Artbildung, akustische Kommunikation und sexuelle Selektion bei griechischen Laubheuschrecken der Poecilimon propinquus Gruppe (Orthoptera: Phaneropteridae)*. [Speciation, acoustic communication and sexual selection in Greek grasshoppers of the *Poecilimon propinquus* group (Orthoptera: Phaneropteridae)].

- Lehmann, GUC. Lehmann, A.W. 2000. Female bushcrickets mated with parasitized males show rapid remating and reduced fecundity (Orthoptera: Phaneropteridae: *Poecilimon mariannae*). *Naturwissenschaften* 87: 404-407.
- Lehmann, GUC. 2012. Weighing costs and benefits of mating in bushcrickets (Insecta: Orthoptera: Tettigoniidae) with an emphasis on nuptial gifts, protandry and mate density. *Frontiers in Zoology* 9: 19.
- McCartney J, Heller K-G, Potter MA, Robertson AW, Telscher K, Lehmann G, Lehmann A, Helversen D v, Reinhold K, Achmann R. 2008. Understanding nuptial gift size in bushcrickets: an analysis of the genus *Poecilimon* (Tettigoniidae; Orthoptera). *Journal of Orthoptera Research* 17: 231-242.
- McCartney J, Lehmann AW, Lehmann GUC. 2010. Lifetime spermatophore investment in natural populations of two closely related bush-cricket species (Orthoptera: Tettigoniidae: *Poecilimon*). *Behaviour* 147: 285-298.
- McCartney J. 2010. *Spermatophore size variation in the bush-cricket genus Poecilimon*. PhD, Massey University, New Zealand: pp 297.
- McCartney J, Kokko H., Heller K-G, Gwynne D. 2012. The evolution of sex differences in mate searching when females benefit: new theory and a comparative test. *Proceedings of the Royal Society of London (B)*. 279: 1225 - 1232.
- Parker GA, Simmons LW, Kirk H. 1990. Analysing sperm competition data: simple models for predicting mechanisms. *Behavioral Ecology and Sociobiology* 27: 55-65.
- Quinn JS, Sakaluk SK. 1986. Prezygotic male reproductive effort in insects: why do males provide more than sperm. *Florida Entomologist* 69: 84-94.
- Reinhold K. 1999. Paternal investment in *Poecilimon veluchianus* bushcrickets: beneficial effects of nuptial feeding on offspring viability. *Behavioral Ecology and Sociobiology* 45: 293-299.
- Reinhold K, Heller K-G. 1993. The ultimate function of nuptial feeding in the bushcricket *Poecilimon veluchianus* (Orthoptera: Tettigoniidae: Phaneropterinae). *Behavioral Ecology and Sociobiology* 32: 55-60.
- Sakaluk SK. 1984. Male crickets feed females to ensure complete sperm transfer. *Science* 223: 609-610.
- Sakaluk SK, Eggert AK. 1996. Female control of sperm transfer and intraspecific variation in sperm precedence: Antecedents to the evolution of a courtship food gift. *Evolution* 50: 694-703.
- Simmons LW. 1990. Nuptial feeding in Tettigoniids: male costs and the rates of fecundity increase. *Behavioral Ecology & Sociobiology* 27: 43-47.
- Simmons LW. 1995a. Courtship feeding in Katydids (Orthoptera: Tettigoniidae): Investment in offspring and in obtaining fertilizations. *The American Naturalist* 146: 307-315.
- Simmons LW. 1995b. Male bushcrickets tailor spermatophores in relation to their remating intervals. *Functional Ecology* 9: 881-886.
- Simmons LW. 2001. *Sperm competition and its evolutionary consequences in the Insects*. Princeton University Press: Princeton, NJ.
- Simmons LW, Parker PG. 1989. Nuptial feeding in insects: Mating effort versus paternal investment. *Ethology* 81: 332-343.
- Simmons LW, Gwynne DT. 1991. The refractory period of female katydids (Orthoptera: Tettigoniidae): sexual conflict over the remating interval? *Behavioral Ecology* 2: 276-282.

- Simmons LW, Craig M, Llorens T, Schinzig M, Hosken D. 1993. Bushcricket spermatophores vary in accord with sperm competition and parental investment theory. *Proceedings for the Royal Society of London* 251: 183-186.
- Simmons LW, Gwynne DT. 1993. Reproductive investment in bushcrickets: the allocation of male and female nutrients to offspring. *Proceedings of the Royal Society of London (B)* 252: 1-5.
- Simmons LW, Beesley L, Lindhjem P, Newbound D, Norris J, Wayne A. 1999. Nuptial feeding by male bushcrickets: an indicator of male quality? *Behavioral Ecology* 10: 263-269.
- Simmons LW, Achmann R. 2000. Microsatellite analysis of sperm-use patterns in the bushcricket *Requena verticalis*. *Evolution* 54: 942-952.
- Simmons LW, Beveridge M, Kennington WJ. 2007. Polyandry in the wild: temporal changes in female mating frequency and sperm competition intensity in natural populations of the tettigoniid *Requena verticalis*. *Molecular Ecology* 16: 4613-4623.
- Smith RJ. 1984. Sperm utilisation strategies of females. In: Smith RL, ed. *Sperm Competition and the Evolution of Animal Mating Systems*: Academic Press.
- Thornhill R. 1976. Sexual selection and paternal investment in insects. *American Naturalist* 110: 153-163.
- Thornhill R, Alcock J. 1983. *The Evolution of Insect Mating Systems*. Harvard University Press: Cambridge, MA.
- Trivers RL. 1972. Parental investment and Sexual Selection. In: Campbell B, ed. *Sexual Selection and the Descent of Man*. Chicago.: Aldine.
- Ulrich B, Reinhold K, Niehuis O., Misof B. 2010. Secondary Structure and Phylogenetic Analysis of the Internal Transcribed Spacers 1 and 2 of Bush Crickets (Orthoptera: Tettigoniidae: Barbitistini). *Journal of Zoological Systematics and Evolutionary Research*: 48, 219-228.
- Vahed K. 1994. *The evolution and function of the spermatophylax in bushcrickets (Orthoptera: Tettigoniidae)*. PhD Thesis, University of Nottingham.
- Vahed K. 1997. Copulation and spermatophores in the Ephippigerinae (Orthoptera: Tettigoniidae): prolonged copulation is associated with a smaller nuptial gift in *Uromenus rugosicollis* Serville. *Journal of Orthoptera Research* 6: 83-89.
- Vahed K. 1998. The function of nuptial feeding in insects: review of empirical studies. *Biological Reviews* 73: 43-78.
- Vahed K. 2003. Increases in egg production in multiply mated female bushcrickets *Leptophyes punctatissima* are not due to substances in the nuptial gift. *Ecological Entomology* 28: 124-128.
- Vahed K. 2007a. Comparative evidence for a cost to males of manipulating females in bushcrickets. *Behavioral Ecology* 18: 499-506.
- Vahed K. 2007b. All that glitters is not gold: sensory bias, sexual conflict and nuptial feeding in insects and spiders. *Ethology* 113: 105-127.
- Vahed K, Gilbert FS. 1996. Differences across taxa in nuptial gift size correlate with differences in sperm number and ejaculate volume in bushcrickets (Orthoptera: Tettigoniidae). *Proceedings of the Royal Society of London: B* 263: 1257-1265.
- Voigt CC, Kretzschmar AS, Speakman JR, Lehmann GUC. 2008. Female bushcrickets fuel their metabolism with male nuptial gifts. *Biology Letters* 4: 476-478.
- Voigt CC, Lehmann GUC, Michener RH, Joachimski MM. 2006. Nuptial feeding is reflected in tissue nitrogen isotope ratios of female katydids. *Functional Ecology* 20: 656-661.

- Warchalowska-Sliwa E, Heller KG, Maryanska-Nadachowska A, Lehmann A. 2000. Chromosome evolution in the genus *Poecilimon* (Orthoptera, Tettigoniioidea, Phaneropteridae). *Folia Biologica-Krakow* 48: 127-136.
- Wedell N. 1991. Sperm competition selects for nuptial feeding in a bushcricket. *Evolution* 45: 1975-1978.
- Wedell N. 1993a. Spermatophore size in bushcrickets: comparative evidence for nuptial gifts as a sperm protection device. *Evolution* 47: 1203-1212.
- Wedell N. 1993b. Mating effort or paternal investment? Incorporation rate and cost of male donations in the wartbiter. *Behavioral Ecology and Sociobiology* 32: 239-246.
- Wedell N. 1994a. Variation in nuptial gift quality in bush crickets (Orthoptera: Tettigoniidae). *Behavioral Ecology* 5: 418-425.
- Wedell N. 1994b. Dual function of the bushcricket spermatophore. *Proceedings of the Royal Society of London* 258: 181-185.
- Wedell N, Arak A. 1989. The wartbiter spermatophore and its effect on female reproductive output (Orthoptera: Tettigoniidae, *Decticus verrucivorus*. *Behavioral Ecology and Sociobiology* 24: 117-125.
- Willemse F, Heller K-G. 1992. Notes on Systematics of Greek species of *Poecilimon* Fischer, 1835 (Orthoptera: Phaneropteridae). *Tijdschrift voor Entomologie* 135: 299-315.

Chapter 36

MATE CHOICE COPYING IN BOTH SEXES OF THE GUPPY: THE ROLE OF SPERM COMPETITION RISK AND SEXUAL HARASSMENT

***Claudia Zimmer^{1,*}, Alexandros S. Gavalas¹,
Benjamin Kunkel¹, Janina Hanisch¹, Sara Martin¹,
Svenja Bischoff¹, Martin Plath¹ and David Bierbach^{1,2}***

¹Department of Ecology and Evolution, University of Frankfurt,
Frankfurt am Main, Germany

²Leibniz-Institute of Freshwater Ecology and Inland Fisheries,
Berlin, Germany

ABSTRACT

Mate choice copying was mostly described as a strategy employed by females to assess the quality of potential mates, but also males can copy other males' mate choice. In both cases, focal individuals show an increased propensity to copulate with a potential mating partner they could observe interact sexually with another individual (the 'model'). Sexual interactions, however, convey additional—partly conflicting—information to the choosing individual: females may try to avoid sexually active males due to a rougher courtship or coercive mating attempts, and males may respond to an increased sperm competition risk (SCR). How do females and males respond to different copying situations, in which the model individual either resides in the vicinity of a potential mating partner without physical contact (i.e., no harassment, and low SCR), or interacts sexually with the potential mating partner? Do individuals copy less in the latter situation? We investigated these questions in the guppy (*Poecilia reticulata*), a livebearing fish with internal fertilization, strong SCR among males, and frequent sexual harassment of females. Focal individuals could choose to associate with a large or a small stimulus fish, and mate choice tests were repeated after the previously non-preferred stimulus fish could be seen associating (low harassment and low SCR) or physically interacting (high harassment and high SCR) with a model individual. In a control treatment, no model was presented. We found both males and

* Corresponding Author's Email: cla.zim.uni@gmail.com (C.Z.).

females to copy similarly in both copying treatments, while no response was observed in the control. This contrasts with a study reporting that Atlantic molly (*P. mexicana*) males copy less under elevated SCR. Even though we lack a compelling explanation as to why both congeners might differ, we are tempted to argue that strong(er) benefits arising from copying may have selected guppies to copy in a broader range of contexts, including situations where the choosing individual incurs harassment or SCR.

Keywords non-independent mate choice, sexual selection, public information, social learning, sperm competition, mate choice copying

INTRODUCTION

Sexual selection, e.g., through mate choice, is a powerful evolutionary force, and individuals may integrate a wide range of information in their assessment of mate quality (Andersson 1994; Candolin 2003). Besides phenotypic traits like body coloration (Houde 1997; Price et al. 2008) and body size (Reznick and Endler 1982), dynamic traits like courtship displays (Darrow and Harris 2004; Rosenthal 2007) and parental care (Warner et al. 1995), also social interactions with conspecifics other than the choosing individual may serve as cues indicating mate quality (Ophir & Galef jr 2004; Ziege et al. 2012; Bierbach et al. 2013a). Especially in group-living animals that spend most of their lives surrounded by conspecifics, mating takes place not in privacy, but in a public domain (Valone and Templeton 2002; Danchin 2004; Dabelsteen 2005; Valone 2007; Druen and Dugatkin 2011). McGregor and Peake (2000), introduced the term ‘communication networks’ to account for the multi-directional nature of communicative interactions in animal societies. In such communication networks, by-standing individuals not involved in the sender-receiver dyad (but residing within a physical range allowing them to detect a signal), have the possibility to obtain information by observing communicatory interactions amongst other group members.

The process of extracting information from social interactions among other individuals within a communication network has been termed ‘social eavesdropping’ (Naguib et al. 2004; Dabelsteen 2005; Peake 2005; Valone 2007; Fitzsimmons et al. 2008). A widespread form of social eavesdropping in the context of mate choice involves the observation of aggressive male conflicts by females and subsequent choice of one of the rivals based on its performance (Ophir and Galef jr 2003; Aquiloni et al. 2008; Bierbach et al. 2013a). Another mechanism involves the observation of sexual interactions between males and females and subsequent choice of one of the involved individuals; termed ‘mate choice copying’ (Dugatkin 1992; Pruett-Jones 1992; Nordell and Valone 1998; Westneat et al. 2000; Bonnie and Earley 2007; Witte and Nöbel 2011).

Mate choice copying has been reported for various vertebrate taxa like birds (Höglund et al. 1990; Galef jr and White 1998; White and Galef jr 1999; Swaddle et al. 2005; Freed-Brown and White 2009) and mammals (Clutton-Brock and McComb 1993; Galef jr et al. 2008), including humans (Uller and Johansson 2003; Waynforth 2007; Yorzinski et al. 2010), but also in invertebrates (Mery et al. 2009; but see Auld et al. 2009). However, following Darwin’s groundbreaking work on sexual selection (Darwin 1871) most studies assumed females to be the choosing sex due to their higher investment in oocyte production compared to males (Andersson 1994). As a consequence, mate choice copying was most extensively studied in

females, especially in teleost fishes like guppies, *Poecilia reticulata* (Dugatkin 1992, 1996, 1998; Dugatkin and Godin 1992, 1993; Brooks 1999; Godin and Hair 2009), sailfin mollies, *Poecilia latipinna* (Schlupp et al. 1994; Witte and Ryan 1998, 2002; Witte and Noltemeier 2002; Witte and Ueding 2003), Hispaniola limia, *Limia perugiae* (Applebaum and Cruz 2000), Japanese medaka, *Oryzias latipes* (Grant and Green 1996) and Atlantic mollies, *Poecilia mexicana* (Heubel et al. 2008).

In theory, females can benefit from mate choice copying in at least two ways (after Gibson and Höglund 1992): (1) The accuracy of their own mate assessment may be increased, which provides a good opportunity for young females to copy decisions of older—more experienced—females (Dugatkin and Godin 1993). (2) Moreover, the time spent in assessing mate quality may be reduced (see Witte and Nöbel 2011 for discussion). The importance of female mate choice copying in shaping mating preferences can be illustrated in sailfin mollies (*P. latipinna*), in which genetically-based female preferences, such as the innate preference for large male body size, are reduced or even reversed through mate choice copying (Witte and Noltemeier 2002).

In promiscuous species, sexual activity (or the number of mating partners) may serve as an indicator of male quality. For example, females of the Atlantic molly were recently shown to have an increased propensity of interacting with males they saw mating with either a female or with another male, and the authors concluded that male sexual activity, whether in a heterosexual or homosexual context, is indeed a trait females find attractive (Bierbach et al. 2013b). However, poeciliid males exhibit high frequencies of sexual behaviors (Farr 1989; Magurran and Seghers 1994; Bisazza and Marin 1995; Houde 1997; Magurran 2001; Plath et al. 2003, 2007), while females can store sperm to fertilize a succession of broods and require only few copulations (Constantz 1984, 1989; Magurran and Seghers 1994; Pilastro et al. 2007). High frequencies of male sexual behavior are known to cause physical damage to the female genital tract (Magurran 2011), and females have less time available for feeding in presence of a harassing male as females spent considerable time trying to avoid sexual harassment (Magurran and Seghers 1994; Plath et al. 2007; Köhler et al. 2011). So, associating with a male that proofed its sexual activity (or at least its motivation to mate through its proximity to another female) obviously brings about an increased risk of sexual harassment. In the current study we asked whether females of two populations of guppies copy less under an increased level of sexual harassment (e.g., through observing sexual interactions between a male and a model female vs. mere associations between both).

For some species, also male mate choice copying has been described [sailfin mollies, *P. latipinna* (Schlupp and Ryan 1997; Witte and Ryan 2002), pipefish, *Syngnathus typhle* (Widemo 2005) and three-spined stickleback, *Gasterosteus aculeatus* (Frommen et al. 2009)]. However, male mate copying in internally fertilizing species—like mammals or poeciliid fishes—remains a conundrum as males incur an increased risk of sperm competition when choosing another male's previous mate (sensu Parker 1979; e.g., Constantz 1984; Becher and Magurran 2004). Nevertheless, poeciliid females are most receptive as virgins or for few days postpartum (Rosenthal 1952; Constantz 1984; Liley 1996) and, accordingly, only a small proportion of females in a population is receptive at the same time (Houde 1997; Magurran 2005). Furthermore, males evaluate female receptivity through so-called nipping behavior, where a male touches the female's genital opening with its snout (Parzefall 1969). Alternatively, males can observe sexual interactions of other males with females, and copying other males' mate choice may allow saving considerable energetic and opportunity costs

associated with mate searching and testing different females via nipping (sensu Schlupp and Ryan 1997). Additionally, in some poeciliids (including the guppy), last male sperm precedence is known (Farr 1989; Evans and Magurran 2001; Becher and Magurran 2004), which gives a competitive edge to the copying male even if the copied one has previously inseminated the female. Nevertheless, sperm competition plays a crucial role in poeciliid mating behavior (see also Evans and Magurran 2001; Dosen and Montgomerie 2004; Wong and McCarthy 2009; Evans and Pilastro 2011; Jeswiet et al. 2011), and a recent study by Bierbach et al. (2011a) found Atlantic molly (*P. mexicana*) males to copy less when sperm competition risk was high (e.g., when males observed sexual interactions between the model male and a stimulus female rather than mere associations without copulation).

While mate choice copying in general appears to be beneficial for both, males and females, risk of sexual harassment (for females) or risk of sperm competition (for males) ought to render certain copying situations less attractive. Specifically, we would expect a fine-tuning of mate choice copying behavior according to the nature of the observed interactions. In the current study, we compared male and female mate choice copying in guppies between two treatments (following the protocol of Bierbach et al. 2011a) during which focal individuals could either observe mere associations between a stimulus fish and a model (representing no sexual harassment or sperm competition) or direct sexual interactions (high risk of sexual harassment or sperm competition). Specifically, we asked (a) whether males and females show differences in their mate choice copying behavior and (b) whether the copying situation (no vs. high risk of sexual harassment or sperm competition) affects the strength of copying. In another experiment, we asked if males and females have a predilection for either of the two copying situations. We thus presented another set of males and females with video sequences showing a male or a female in association or engaging in sexual behavior and measured their preferences for both types of copying situations.

MATERIALS AND METHODS

Study Species and Animal Maintenance

Test fish came from two populations of guppies (*P. reticulata*): individuals from the first population were second to third generation descendants of wild-caught fish originating from Venezuela and were imported to Germany by ‘Aquarium Dietzenbach’ (coded as ‘wild-type’ in the following sections), while fish from the second population were regular aquarium fish (coded as ‘domestic’ in the later). Fish were reared in mixed-sex, single-population stock tanks (150–200 L) under a 12:12 light: dark cycle at a constant temperature of 28°C. All tanks were equipped with natural gravel and artificial plants to provide shelter. We fed the fish to satiation with commercially available flake food (TetraMin®) and food tablets (TetraTabs®) twice daily, and frozen chironomid larvae and *Artemia salina* nauplii once a week.

Experiment 1

To compare male and female mate choice copying in different copying situations (see Introduction), we largely followed the experimental protocol of Schlupp and Ryan (1997) that was recently modified and used to investigate male mate choice copying under sperm competition risk in the Atlantic molly (*P. mexicana*, Bierbach et al. 2011a). All focal fish were isolated prior to the tests in groups of < 20 individuals in 100-L aquaria for at least 24 hours. Focal and stimulus individuals stemmed from different isolation groups to prevent effects of familiarity (Bierbach et al. 2011b). A standard test tank (80 cm length × 30 cm height × 30 cm depth) was divided into five sections of equal size: The two lateral compartments were divided by transparent Plexiglas partitions and contained the stimulus fish (Fig. 1). The remaining compartment was visually divided in two lateral preference zones and a central neutral zone by marks drawn on the front of the tank. The tank was filled to a height of 15 cm with aged tap water of 27–29°C and was aerated and heated whenever no experiment was conducted.

As both sexes of the guppy are known to have an intrinsic mating preference for large body size (Reynolds and Gross 1992; Herdman et al. 2004), we took two different-sized stimulus individuals from one isolation tank and introduced them into one of the two lateral compartments each. Stimulus individuals were exchanged between subsequent trials. Afterward, a focal fish of the opposite sex was introduced into a clear Plexiglas cylinder (10 cm diameter), placed centrally in the neutral zone, and was left undisturbed for 5 min. After this habituation period, we gently lifted the cylinder and measured the amount of time the focal fish spent in each of the two preference zones, i.e., near the large and small stimulus fish during a 5-min observation period. To account for potential side biases, the focal fish was then gently placed back into the cylinder and stimulus fish were interchanged. After 5 min of habituation, measurement was repeated for another 5 min. Like in previous studies, we decided *a priori* to discard trials in which focal fish spent more than 80% of their choice time during both parts of the experiment in the same compartment as side biases (e.g., Plath et al. 2004), but no cases of side-bias were observed. This initial preference test consisting of two measurements is henceforth called the first part of a trial (Figure 1a).

Times spent near each stimulus fish during the two test units were summed, and if the focal fish spent more than 66% of the time with one of the stimuli they were considered to have shown a preference and were used for the subsequent test phases. Trials in which focal fish did not show a strong preference were terminated at this point (Schlupp and Ryan 1997). Out of 81 initially tested wild-type males, 15 did not show a preference for one of the two stimulus females; similarly, 17 out of 85 females showed no preference. In the domestic stock, 17 out of 61 initially tested males did not show a preference and out of 58 females, 16 tests were terminated after the first part of a trial.

Directly after the first part, the focal fish was retransferred into the Plexiglas cylinder in the center of the neutral zone (“observation phase”; Figure 1b). We created situations during which the focal fish either had no opportunity to copy (control), or it could observe visual interactions and associations between the model individual and the formerly non-preferred mate, or physical (i.e., sexual) interactions of the latter two. To do so, focal fish were randomly assigned to one of the following three treatments: In the control treatment (1), an empty transparent Plexiglas cylinder was placed in the front portion of the non-preferred individual’s compartment. To control for any changes in the behavior of the stimulus fish due to this procedure, we introduced another Plexiglas cylinder also into the compartment of the formerly

preferred individual but covered this portion of the stimulus compartment by an opaque Plexiglas sheet and thus prevented the focal fish from seeing it (Figure 1b).

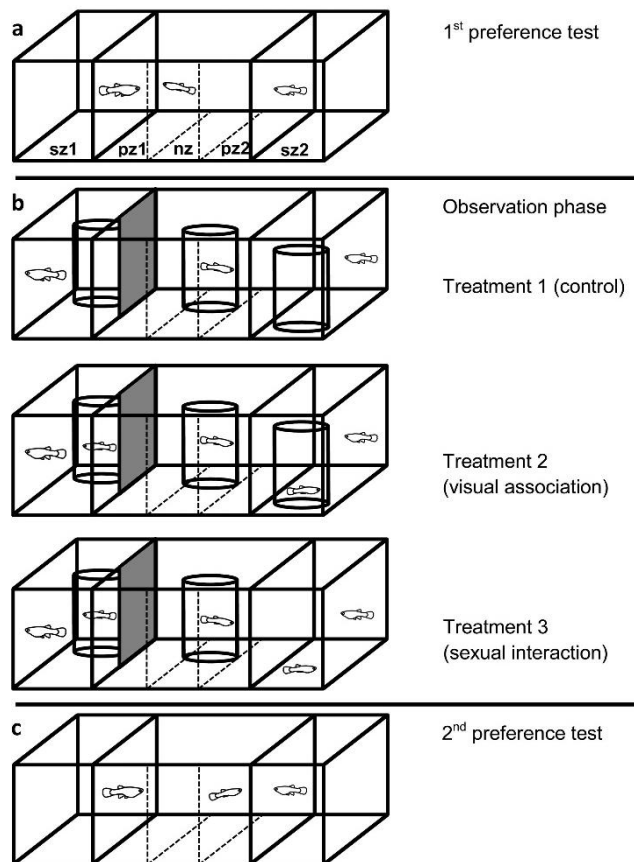


Figure 1. Experimental setup used to examine mate choice copying in *P. reticulata*. In the first preference test (a), we measured the time the focal individual spent in the preference zones (pz1 or pz2) near either stimulus fish (sz1 or sz2). During the observation phase (b), focal fish were randomly assigned to one of three treatments: treatment 1, control without a model individual; treatment 2: visual association, where the focal fish could observe a model individual visually interacting with the formerly non-preferred stimulus, and treatment 3: sexual interactions, during which they could observe physical interactions between the model individual and the formerly non-preferred stimulus. The second preference test (c) was analogous to the first and enabled us to compare preferences before and after the focal fish had the possibility to copy the model's mate choice. Depicted is the test scheme for male mate choice copying; tests for female mate choice copying were conducted analogously.

Treatment (2) was similar to previous mate copying tests in *P. latipinna* (Schlupp and Ryan 1997). Again, two Plexiglas cylinders were placed in the front portion of both compartments holding the stimulus fish (see above), but we now introduced one model individual per cylinder, so the focal fish could observe a model of the same sex visually interact with the formerly non-preferred mate, but was prevented from seeing the other model interact with the formerly preferred mate.

To allow physical interactions between the model individual and the formerly non-preferred stimulus, we introduced the model individual without a Plexiglas cylinder into the

partition of the formerly non-preferred mate [treatment (3)]. Hence, the model individual could interact sexually with the stimulus fish. The previously preferred mate was able to interact visually with a fish of the opposite sex in a Plexiglas cylinder just like in treatment (2) (the cylinder with this fish was again hidden behind an opaque plastic sheet as described before). We quantified pre-copulatory behavior (nipping, mean $\pm$ SE, wild-type 10 ± 8 ; domestic 8 ± 10) and copulations (gonopodial thrusting, wild-type 3 ± 5 ; domestic 1 ± 3) in the stimulus-model pair during the 20 min of observation. After the 20-min observation phase, the model individuals were removed and measurement of association preferences (including switching of side assignments of the stimulus fish between the two test units) was carried out as described for the first part of the experiment (Figure 1c). After the tests, the standard lengths of all fish were measured to the nearest millimeter (Table 1a).

Table 1. Body size of the test fish. Given are mean ($\pm$ SE) standard lengths (SL in mm) of the focal, stimulus (large and small) as well as model fish ('preferred' = model interacted with preferred stimulus; 'non-preferred' = model interacted with non-preferred stimulus)

Population/Sex	focal	large	small	model preferred	model non-preferred
a) Experiment 1					
Domestic, females ($N = 42$)	33.0 ± 0.7	27.5 ± 0.3	21.5 ± 0.2	33.6 ± 1.0	34.4 ± 0.7
Domestic, males ($N = 44$)	22.7 ± 0.4	35.5 ± 0.5	23.8 ± 0.4	23.3 ± 0.4	22.9 ± 0.4
Wild-type, females ($N = 68$)	28.9 ± 1.1	24.8 ± 0.6	18.2 ± 0.4	27.3 ± 1.3	28.3 ± 1.3
Wild-type, males ($N = 66$)	23.7 ± 0.3	36.4 ± 0.4	29.4 ± 0.6	24.3 ± 0.5	24.6 ± 0.6
b) Experiment 2					
Wild-type, females ($N = 21$)	28.0 ± 2.8				
Wild-type, males ($N = 22$)	21.0 ± 1.5				

Experiment 2

To test if males and females have a predilection for either of the two copying situations, we used dichotomous association preference tests in conjunction with video playback techniques (Makowicz et al. 2010; Bierbach et al. 2013b). We created video sequences showing (1) a male and a female separated by an (invisible) PVC partition or (2) a pair of fish allowed to interact freely. We used only wild-type fish in this experiment. Males for the production of video recordings were isolated for one week in 10-L tanks to ensure that they were motivated to mate. The recording tank was divided into two same-sized compartments by an opaque PVC partition and a camcorder (Panasonic HDC-TM60EG-K) was set up 9 cm from the front wall to capture both compartments simultaneously. For the videos showing 'visual association', one female (SL $\pm$ SD, 29.2 ± 0.9 mm) and one male (22.6 ± 0.7 mm) were placed in either compartment, and after a 5-min acclimatization phase were filmed for at least 6 minutes. For videos showing direct interactions, we removed the PVC-partition after the acclimatization phase so males could mate with the female. We recorded numbers of nipping (8 ± 8) and mating

attempts (8 ± 5). For each treatment and prepared 10 videos (resolution 800 x 600 px, saved as avi-files) that were later displayed via MS Powerpoint (Microsoft Office 2010).

For the binary association preference tests we followed the protocol of Bierbach et al. (2011a). The test tank (60 x 30 x 30 cm) was visually divided into three sections: two lateral preference zones of 10 cm width and a central, neutral zone of 40 cm width. The tank was filled with aged tap water to a height of 25 cm. Water was heated and aerated between trials, but heater and air stone were removed prior to testing. The bottom of the test tank was covered with fine white quartz sand. Illumination was provided by two 60 W fluorescent lamps on the ceiling of the test room. We placed two identical video screens (Samsung SyncMaster P2470LHD, 24 in.; resolution 1,280 × 768 pixels) at both short sides of the test tank and covered all remaining parts of the test tank's walls (including the front wall) with grey cardboard to reduce disturbances from the outside. The test fish was observed from above by use of a mirror.

To initiate a trial, we introduced a focus fish into a transparent Plexiglas cylinder placed centrally in the neutral zone and gave the test subject 5 min for acclimatization, during which the monitors showed a uniformly gray display. Afterwards, we released the focal fish and started the video playback, showing the two types of video sequences (see above) on the left and right side (side-assignment was altered between trials). We measured the time the focal individual spent in each preference zone for 5 min. Focal fish were then reintroduced into the Plexiglas cylinder and given another 5 min of habituation while both screens displayed gray background only. Subsequently, measurement of association times was repeated for another 5 min with sides of the video presentation interchanged. We summed up association times from both parts and calculated the strength of preference (SOP) as follows:

$$SOP = \frac{t_{sex\ int} - t_{vis\ assoc}}{t_{sex\ int} + t_{vis\ assoc}},$$

where $t_{sex\ int}$ is the time spent in the compartment next to the sexual interaction-sequence and $t_{vis\ assoc}$ the time spent next to the visual association-sequence. Upon completion of a trial, we measured SL of the focal fish to the nearest millimeter (Table 1b).

Statistical Analysis

All data were tested for normal distribution using Kolmogorov-Smirnov tests and were presented as mean $\pm$ standard error (SE) throughout. Statistical analyses were conducted using IBM SPSS 18.

Experiment 1

In a first step, we evaluated the focal individuals' preference for large body size. We compared the amount of time focal fish spent near large and small stimulus individuals during the initial preference test (first part) and during the second part of the tests for all three treatments separately, using paired sample *t*-tests.

Our main question was whether focal fish would alter their individual mate choice decisions after they had observed a model individual visually [treatment (2)] or sexually [treatment (3)] interact with the formerly non-preferred stimulus. We, therefore, calculated a score expressing the change of mating preference as the difference between individuals'

relative association time with the initially non-preferred female during the second part and relative association time near the same female during the first part ('copying score'; see Heubel et al. 2008; Bierbach et al. 2011a, 2013b), such that no change would lead to a score of zero, and positive values would indicate a relative increase in time spent near the initially non-preferred stimulus due to copying. Scores were compared using a full-factorial univariate General Linear Model (GLM) with 'treatment', 'sex' and 'population' as fixed factors. As mate choice copying in other poeciliids depends on the size difference between the two stimulus individuals (Witte and Ryan 1998), we initially included stimulus body size difference (SL large – SL small) and focal individuals' body size as covariates but removed them as neither the covariates themselves nor any of their interactions had a significant effect (body size difference: $F_{1, 198} = 2.11$, $P = 0.15$; focal body size: $F_{1, 198} = 0.22$, $P = 0.64$). We also removed the non-significant 3-way interaction of all main factors from the final model ($F_{2, 208} = 2.55$, $P = 0.081$) as well as non-significant 2-way interactions (treatment $\times$ sex: $F_{2, 209} = 2.1$, $P = 0.16$; sex $\times$ population: $F_{1, 210} = 0.59$, $P = 0.44$). We used Fisher's LSD tests to *post hoc* analyze differences between treatments.

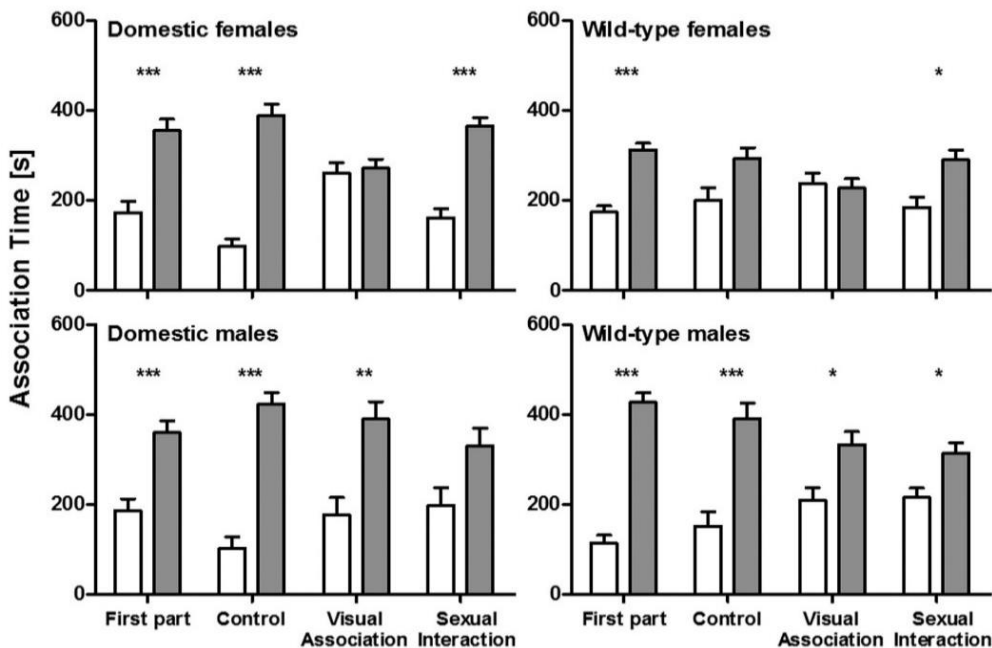


Figure 2. Time focal fish spent in association with the large and the small stimulus individuals during the first part of the choice tests (left) and during the second part in the three copying treatments. During the observation phase preceding the second part of the choice tests either no copying was possible (control), or a model fish could interact visually or physically with the previously non-preferred stimulus. Depicted are association times ($\pm$ SE); significant P -values are from paired t -tests (* $P < 0.05$; ** $P < 0.01$; *** $P < 0.001$).

To test for a potential effect of body size differences between model and focal individuals on the strength of mate choice copying (copying score), we included body size difference (SL focal – SL model) as a covariate in another GLM while analyzing only the subset of data from treatments (2) and (3).

To evaluate whether the number of sexual interactions that a focal fish observed during the observation phase affected the strength of copying in treatment 3, we included numbers of nippings and mating attempts as covariates in another GLM analyzing the subset of data from treatment (3).

Experiment 2

We compared times males and females spent near the video sequences showing sexual interactions and visual associations for each sex separately using paired *t*-tests. Furthermore, we compared the strength of preference (SOP) between males and females in a GLM including ‘video ID’ as random factor and focal individuals’ body size (SL) as a covariate.

RESULTS

Experiment 1: Mate Choice Copying in Males and Females

Preference for Large Mating Partners

Males and females of both populations spent significantly more time in association with larger stimulus fish during the first part of the preference tests (Fig. 2). This preference was largely maintained during the second part of the tests (i.e., after the observation phase) with two notable exceptions: females of both populations ceased expressing a preference for the larger stimulus male in treatment (2) (visual association) and spent almost equal amounts of time near the two stimuli. Also, no significant preference was seen in males from the domestic stock in treatment (3) (sexual interaction), but qualitatively males still tended to spent more time associating with the larger female ($P = 0.11$; Fig. 2).

Changes in Individual Preferences

The final GLM detected a significant effect of the main factor ‘treatment’ and *post hoc* LSD tests indicate that, overall, test subjects showed consistent mate choice in the control treatment (copying scores close to zero in Fig. 3; LSD: $P < 0.001$ compared to both other treatments), but copied to a similar extent in treatment (2) (positive copying score for ‘visual association’) and treatment (3) [positive copying score for ‘sexual interaction’; LSD test, treatment (2) vs. treatment (3): $P = 0.52$]. Moreover, a significant interaction effect of ‘treatment by population’ was uncovered, suggesting that the two populations differed in their responses to the different copying treatments, even though effect size (estimated as partial η^2) was lower for the interaction term compared to the main effect ‘treatment’ (Table 2). Nevertheless, estimated marginal means derived from GLM revealed that domesticated guppies, overall, copied more than wild-type guppies (Fig. 3). Sexes did not differ statistically in their copying behavior (Table 2).

We also tested for an effect of the focal and model individuals’ body size difference on the strength of copying within the subset of data from treatments (2) and (3). However, body size difference had no statistically significant effect (GLM; $F_{1, 142} = 0.40$, $P = 0.53$). In another GLM, we could not detect an effect of the amount of sexual behavior shown by the stimulus-model pair in treatment (3) and strength of copying (nipping behavior: $F_{1, 64} = 0.85$, $P = 0.36$; copulations: $F_{1, 64} = 0.18$, $P = 0.68$; courtship displays: $F_{1, 64} = 0.27$, $P = 0.60$).

Table 2. Results from a univariate GLM with ‘copying score’ as the dependent variable and ‘treatment’ (control, visual association and sexual interaction), ‘sex’, ‘population’ (domestic or wild-type) and their two-way interaction terms as fixed factors. Significant effects are in bold face

Source	Mean Square	<i>df</i>	<i>F</i>	<i>P</i>	Partial η^2
Treatment	1.67	2	19.63	<0.001	0.17
Sex	0.17	1	2.17	0.16	<0.01
Population	0.04	1	0.29	0.64	<0.01
Treatment * Population	0.41	2	4.56	0.01	0.05
Error	0.09	213			

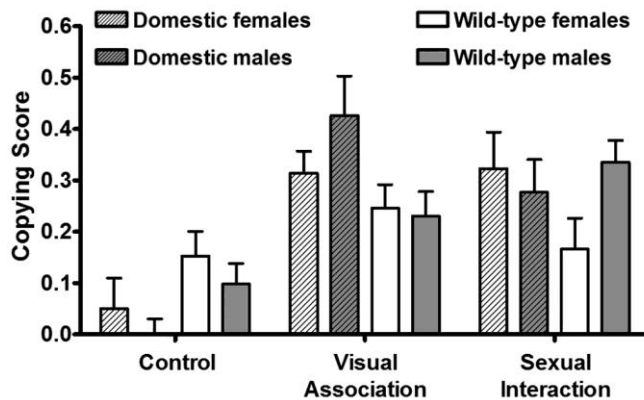


Figure 3. Changes in individual mating preferences when focal fish were given an opportunity to copy another individual's mate choice. During the observation phase preceding the second part of the choice tests either no copying was possible (control, left) or a model fish could interact visually (middle) or physically (right) with the previously non-preferred stimulus fish. Depicted are copying scores (see main text), whereby positive values indicate that preferences for the initially non-preferred stimulus individuals increased in strength.

Experiment 2: Preference for Different Copying Situations

Neither females nor males showed a preference for either type of video (males, visual association: 194.7 ± 17.3 s; sexual interactions: 172.6 ± 16.6 s; paired *t*-test: $t_{21} = -0.91$, $P = 0.37$; females, visual association: 205.5 ± 20.6 s; sexual interactions 161.6 ± 22.6 s; $t_{20} = -1.16$, $P = 0.26$). Also, there was no significant effect of the factor ‘sex’ ($F_{1,30} = 0.14$, $P = 0.71$) when comparing SOP-values using GLM. Likewise, the covariate (focal individuals’ body size) had no significant effect ($F_{1,30} = 0.20$, $P = 0.66$).

DISCUSSION

Copying the mate choice decisions of other individuals provides benefits to both males and females, as copying helps reduce costs of mate searching (Gibson and Höglund 1992; Schlupp and Ryan 1997; Witte and Noltemeier 2002). At the same time, mate choice copying is

associated with costs, namely increased sperm competition risk for males that copy other males' mate choice (see Evans and Magurran 1999) and increased sexual harassment for copying females (see Plath et al. 2007). Cost-benefit trade-offs are predicted to shape mate choice decisions in general (Kokko et al. 2003), and similar trade-offs ought to affect copying behavior as well.

Guppies have a promiscuous mating system characterized by strong mating competition among males, leading to high levels of sperm competition and sexual harassment (Magurran and Seghers 1994; Evans and Magurran 1999; Plath et al. 2007), and so the costs associated with mate choice copying should be high. Our current study was motivated by the hypothesis that individuals copy opportunistically when costs associated with mate choice copying are moderate or low (i.e., in the visual association treatment), but cease copying other individuals' mate choice decisions when costs are high (in the direct interaction treatment). Support for this hypothesis comes from a study that found males of the Atlantic molly (*P. mexicana*), another member of the family Poeciliidae, to reduce mate choice copying when sperm competition risk was experimentally increased using the same experimental approach we employed here (Bierbach et al. 2011a).

Our study revealed that male and female guppies copy the mate choice decisions of conspecifics, and there was no significant difference between sexes in strength of mate choice copying. Contrary to prediction, however, we found no significant difference between the copying situation involving mere associations between stimulus and model individuals [treatment (2)] as compared to the situation involving sexual interactions [treatment (3)]. We do not have a compelling explanation at hand for the apparent differences between Atlantic mollies (Bierbach et al. 2011a) and guppies (this study). It seems tempting to argue though that copying provides greater benefits to the copier in guppies than Atlantic mollies, such that male and female guppies are willing to accept costs arising from sperm competition and harassment and thus, copy in a broader range of social contexts. Future studies should make an attempt to shed light on this question. For example, video animation techniques could be employed to create virtual copying situations that gradually increase the amount of sexual interactions between stimulus and model individuals. Comparing Atlantic mollies and guppies in such an experimental design would be an elegant test to our new hypothesis, as Atlantic mollies are predicted to cease copying at a lower threshold value of sexual behavior. This experiment, of course, still leaves open the important question of *why* species of poeciliids might differ in their propensity to copy the mate choice of others. Comparative approaches using different species and genera of poeciliid fishes might shed light on the question of how differences in social organization, mate encounter rate, male aggression, and other factors, affect the evolution of copying behavior.

Beside the strong treatment effect, we also found a weak but statistically significant interaction effect of 'population by treatment', indicating differences in the magnitude of mate copying between populations, and domestic guppies copied more than the descendants of wild-caught animals. Noteworthy, several studies reported on a lack of female mate choice copying behavior in feral and wild-type pet shop guppies (Brooks 1996, 1999; Lafleur et al. 1997), resulting in some controversy about the importance of mate choice copying for sexual selection (Dugatkin 1998). The majority of studies suggest that female mate choice copying is a universal and widespread behavior—at least in guppies (Dugatkin 1992, 1996; Dugatkin and Godin 1992, 1993; Godin & Hair 2009)—while Brooks (1998) argued that population differences may exist. Our results indeed suggest some variation in mate choice copying among populations. We have

no compelling explanation at hand as to the question of why domestic guppies copied more than wild type guppies. The exact founder population(s) from which the artificial strain was derived is unknown, so it remains possible that the ancestors of our domestic strain already exhibited stronger copying; likewise, we cannot rule out the possibility that the domestication process has selected for increased copying, even though the latter scenario seems unlikely.

ACKNOWLEDGMENTS

We like to thank Holger Geupel for help with animal care. Financial support came from the German Research Foundation (Deutsche Forschungsgemeinschaft, DFG) to MP (PL 470/3-1).

REFERENCES

- Andersson MB (1994) *Sexual selection*. Princeton University Press, Princeton NJ.
- Applebaum SL, Cruz A (2000) The Role of Mate-Choice Copying and Disruption Effects in Mate Preference Determination of *Limia perugiae* (Cyprinodontiformes, Poeciliidae). *Ethology* 106 (10):933–944.
- Aquiloni L, Buřić M, Gherardi F (2008) Crayfish females eavesdrop on fighting males before choosing the dominant mate. *Curr Biol* 18 (11):R462.
- Auld HL, Punzalan D, Godin JJ, Rundle HD (2009) Do female fruit flies (*Drosophila serrata*) copy the mate choice of others? *Behavioural Processes* 82 (1):78–80.
- Becher SA, Magurran AE (2004) Multiple mating and reproductive skew in Trinidadian guppies. *P Roy Soc B-Biol Sci* 271 (1543):1009–1014.
- Bierbach D, Girndt A, Hamfler S, Klein M, Mücksch F, Penshorn M, Schwinn M, Zimmer C, Schlupp I, Streit B, Plath M (2011b) Male fish use prior knowledge about rivals to adjust their mate choice. *Biology Letters* 7 (3):349–351.
- Bierbach D, Jung CT, Hornung S, Streit B, Plath M (2013b) Homosexual behaviour increases male attractiveness to females. *Biology Letters* 9 (1):20121038.
- Bierbach D, Kronmarck C, Hennige-Schulz C, Stadler S, Plath M (2011a) Sperm competition risk affects male mate choice copying. *Behav Ecol Sociobiol* 65 (9):1699–1707.
- Bierbach D, Sassmannshausen V, Streit B, Arias-Rodriguez L, Plath M (2013a) Females prefer males with superior fighting abilities but avoid sexually harassing winners when eavesdropping on male fights. *Behav Ecol Sociobiol* 67 (4):675–683.
- Bisazza A, Marin G (1995) Sexual selection and sexual size dimorphism in the eastern mosquitofish *Gambusia holbrooki* (Pisces Poeciliidae). *Ethol Ecol Evol* 7 (2):169–183.
- Bonnie KE, Earley RL (2007) Expanding the scope for social information use. *Anim Behav* 74 (2):171–181.
- Briggs SE, Godin JJ, Dugatkin LA (1996) Mate-choice copying under predation risk in the Trinidadian guppy (*Poecilia reticulata*). *Behavioral Ecology* 7 (2):151–157.
- Brooks R (1996) Copying and the repeatability of mate choice. *Behav Ecol Sociobiol* 39 (5):323–329.

- Brooks R (1998) The importance of mate copying and cultural inheritance of mating preferences. *Trends in Ecology & Evolution* 13 (2):45–46.
- Brooks R (1999) Mate choice copying in guppies: females avoid the place where they saw courtship. *Behaviour* 136 (4):411–421.
- Candolin U (2003) The use of multiple cues in mate choice. *Biol Rev* 78 (4):575–595.
- Clutton-Brock T, McComb K (1993) Experimental tests of copying and mate choice in fallow deer (*Dama dama*). *Behav Ecol* 4 (3):191–193.
- Constantz GD (1984) Sperm competition in Poeciliid fishes. In: Smith RL (ed) *Sperm Competition and the Evolution of Animal Mating systems*. Elsevier Science, Oxford, pp 465–485.
- Constantz GD (1989) Reproductive biology in poeciliid fishes. In: Meffe GK (ed) *Ecology and evolution of livebearing fishes* (Poeciliidae). Prentice Hall, Englewood Cliffs, NJ, pp 33–50.
- Dabelsteen T (2005) Public, private or anonymous? Facilitating and countering eavesdropping. In: McGregor PK (ed) *Animal communication networks*. Cambridge University Press, Cambridge, pp 38–62.
- Danchin É (2004) *Public Information: From Nosy Neighbors to Cultural Evolution*. *Science* 305 (5683):487–491.
- Darrow KO, Harris WA (2004) Characterization and Development of Courtship in Zebrafish, *Danio rerio*. *Zebrafish* 1 (1):40–45.
- Darwin C (1871) *The descent of man, and Selection in relation to sex*. J. Murray, London.
- Dosen LD, Montgomerie R (2004) Mate preferences by male guppies (*Poecilia reticulata*) in relation to the risk of sperm competition. *Behav Ecol Sociobiol* 55 (3):266–271.
- Druen M, Dugatkin LA (2011) Communication networks. In: Evans JP, Pilastro A, Schlupp I (eds) *Ecology and evolution of poeciliid fishes*. University of Chicago Press, Chicago, IL, pp 84–113.
- Dugatkin LA (1992) Sexual Selection and Imitation: Females Copy the Mate Choice of Others. *Am Nat* 139 (6):1384–1389.
- Dugatkin LA (1996) Copying and Mate Choice. In: Heyes CM, Galef BG, jr (eds) *Social learning in animals: the roots of culture*. Academic Press, New York, pp 85–105.
- Dugatkin LA (1998) A comment on Lafleur et al.'s re-evaluation of mate-choice copying in guppies. *Anim Behav* 56:513–514.
- Dugatkin LA (1998) Genes, copying, and female mate choice: shifting thresholds. *Behav Ecol* 9 (4):323–327.
- Dugatkin LA, Godin JJ (1992) Reversal of Female Mate Choice by Copying in the Guppy (*Poecilia reticulata*). *P Roy Soc B-Biol Sci* 249 (1325):179–184.
- Dugatkin LA, Godin JJ (1993) Female mate copying in the guppy (*Poecilia reticulata*): age-dependent effects. *Behav Ecol* 4 (4):289–292.
- Evans JP, Magurran AE (1999) Male mating behaviour and sperm production characteristics under varying sperm competition risk in guppies. *Anim Behav* 58 (5):1001–1006.
- Evans JP, Magurran AE (2001) Patterns of sperm precedence and predictors of paternity in the Trinidadian guppy. *P Roy Soc B-Biol Sci* 268 (1468):719–724.
- Evans JP, Pilastro A (2011) Postcopulatory sexual selection. In: Evans JP, Pilastro A, Schlupp I (eds) *Ecology and evolution of poeciliid fishes*. University of Chicago Press, Chicago, IL, pp 197–208.

- Farr JA (1989) Sexual selection and secondary sexual differentiation in poeciliids: determinants of male mating success and the evolution of female choice. In: Meffe GK (ed) *Ecology and evolution of livebearing fishes* (Poeciliidae). Prentice Hall, Englewood Cliffs, NJ, pp 91–123.
- Fitzsimmons LP, Foote JR, Ratcliffe LM, Mennill DJ (2008) Eavesdropping and communication networks revealed through playback and an acoustic location system. *Behav Ecol* 19 (4):824–829.
- Freed-Brown G, White DJ (2009) Acoustic mate copying: female cowbirds attend to other females' vocalizations to modify their song preferences. *P Roy Soc B-Biol Sci* 276 (1671):3319–3325.
- Frommen JG, Rahn AK, Schroth SH, Waltschyk N, Bakker TCM (2009) Mate-choice copying when both sexes face high costs of reproduction. *Evol Ecol* 23 (3):435–446.
- Galef BG, jr, Lim TC, Gilbert GS (2008) Evidence of mate choice copying in Norway rats, *Rattus norvegicus*. *Anim Behav* 75 (3):1117–1123.
- Galef BG, jr, White DJ (1998) Mate-choice copying in Japanese quail, *Coturnix coturnix japonica*. *Anim Behav* 55 (3):545–552.
- Gibson RM, Höglund J (1992) Copying and sexual selection. *Trends Ecol Evol* 7 (7):229–232.
- Godin JJ, Hair K (2009) Mate-choice copying in free-ranging Trinidadian guppies (*Poecilia reticulata*). *Behaviour* 146 (10):1443–1461.
- Grant JWA, Green LD (1996) Mate copying versus preference for actively courting males by female Japanese medaka (*Oryzias latipes*). *Behav Ecol* 7 (2):165–167.
- Heubel KU, Hornhardt K, Ollmann T, Parzefall J, Ryan MJ, Schlupp I (2008) Geographic variation in female mate-copying in the species complex of a unisexual fish, *Poecilia formosa*. *Behaviour* 145 (8):1041–1064.
- Höglund J, Lundberg A, Alatalo RV (1990) Copying the Mate Choice of Others? Observations On Female Black Grouse. *Behaviour* 114 (1):221–231.
- Houde AE (1997) *Sex, color, and mate choice in guppies*. Princeton University Press, Princeton NJ.
- Jeswiet SB, Lee-Jenkins SSY, Godin JJ (2011) Concurrent effects of sperm competition and female quality on male mate choice in the Trinidadian guppy (*Poecilia reticulata*). *Behav Ecol* 23 (1):195–200.
- Köhler A, Hildenbrand P, Schleucher E, Riesch R, Arias-Rodriguez L, Streit B, Plath M (2011) Effects of male sexual harassment on female time budgets, feeding behavior, and metabolic rates in a tropical livebearing fish (*Poecilia mexicana*). *Behav Ecol Sociobiol* 65 (8):1513–1523.
- Lafleur DL, Lozano GA, Sclafani M (1997) Female mate-choice copying in guppies, *Poecilia reticulata*: a re-evaluation. *Anim Behav* 54 (3):579–586.
- Liley NR (1996) Ethological isolating mechanisms in four sympatric species of Poeciliid fishes. *Behaviour* 13 (suppl.) (1):197.
- Magurran AE (2001) Sexual conflict and evolution in Trinidadian guppies. *Genetica* 112:463–474.
- Magurran AE (2005) *Evolutionary ecology: the Trinidadian guppy*. Oxford University Press, Oxford, UK.
- Magurran AE (2011) Sexual coercion. In: Evans JP, Pilastro A, Schlupp I (eds) *Ecology and evolution of poeciliid fishes*. University of Chicago Press, Chicago, IL, pp 209–217.

- Magurran AE, Seghers BH (1994) A Cost of Sexual Harassment in the Guppy, *Poecilia reticulata*. *P Roy Soc B-Biol Sci* 258 (1351):89–92.
- Makowicz A, Plath M, Schlupp I (2010) Using video playback to study the effect of an audience on male mating behavior in the Sailfin molly (*Poecilia latipinna*). *Behavioural Processes* 85 (1):36–41.
- McGregor PK, Peake TM (2000) Communication networks: social environments for receiving and signalling behaviour. *Acta Ethol* 2 (2):71–81.
- Mery F, Varela SA, Danchin É, Blanchet S, Parejo D, Coolen I, Wagner RH (2009) Public Versus Personal Information for Mate Copying in an Invertebrate. *Curr Biol* 19 (9):730–734.
- Naguib M, Amrhein V, Kunc HP (2004) Effects of territorial intrusions on eavesdropping neighbors: communication networks in nightingales. *Behav Ecol* 15:1011–1015.
- Nordell SE, Valone TJ (1998) Mate choice copying as public information. *Ecol Lett* 1 (2):74–76.
- Ophir AG, Galef BG, jr (2003) Female Japanese quail that ‘eavesdrop’ on fighting males prefer losers to winners. *Anim Behav* 66 (2):399–407.
- Ophir AG, Galef BG, jr (2004) Sexual experience can affect use of public information in mate choice. *Anim Behav* 68:1221–1227.
- Parker GA (1979) Sexual selection and sexual conflict. In: Blum MS, Blum NA (eds) *Sexual selection and reproductive competition in insects*. Academic Press, New York, San Francisco, London, pp 123–166.
- Parzefall J (1969) Zur vergleichenden Ethologie verschiedener *Mollienesia*-Arten einschliesslich einer Höhlenform vom *M. sphenops*. *Behaviour* 33 (1/2):1–37.
- Peake TM (2005) Eavesdropping in communication networks. In: McGregor PK (ed) *Animal communication networks*. Cambridge University Press, Cambridge, pp 13–37.
- Pilastro A, Gasparini C, Boschetto C, Evans JP (2007) Colorful male guppies do not provide females with fecundity benefits. *Behav Ecol* 19 (2):374–381.
- Plath M, Makowicz AM, Schlupp I, Tobler M (2007) Sexual harassment in live-bearing fishes (Poeciliidae): comparing courting and noncourting species. *Behav Ecol* 18 (4):680–688.
- Plath M, Parzefall J, Körner KE, Schlupp I (2004) Sexual selection in darkness? Female mating preferences in surface- and cave-dwelling Atlantic mollies, *Poecilia mexicana* (Poeciliidae, Teleostei). *Behav Ecol Sociobiol* 55 (6):596–601.
- Plath M, Parzefall J, Schlupp I (2003) The role of sexual harassment in cave and surface dwelling populations of the Atlantic molly, *Poecilia mexicana* (Poeciliidae, Teleostei). *Behav Ecol Sociobiol* 54 (3):303–309.
- Price AC, Weadick CJ, Shim J, Rodd FH (2008) Pigments, Patterns, and Fish Behavior. *Zebrafish* 5 (4):297–307.
- Pruett-Jones S (1992) Independent versus nonindependent mate choice: do females copy each other? *Am Nat* 140 (6):1000–1009.
- Reynolds JD, Gross MR (1992) Female Mate Preference Enhances Offspring Growth and Reproduction in a Fish, *Poecilia reticulata*. *P Roy Soc B-Biol Sci* 250 (1327):57–62.
- Reznick D, Endler JA (1982) The Impact of Predation on Life History Evolution in Trinidadian Guppies (*Poecilia reticulata*). *Evolution* 36:160–177.
- Rosenthal GG (2007) Spatiotemporal Dimensions of Visual Signals in Animal Communication. *Annu. Rev. Ecol. Evol. Syst.* 38 (1):155–178.

- Rosenthal HL (1952) Observations on Reproduction of the Poeciliid *Lebistes reticulatus* (Peters). *Biological Bulletin* 102 (1):30.
- Schlupp I, Marler C, Ryan MJ (1994) Benefit to male sailfin mollies of mating with heterospecific females. *Science* 263 (5145):373–374.
- Schlupp I, Ryan MJ (1997) Male Sailfin mollies (*Poecilia latipinna*) copy the mate choice of other males. *Behav Ecol* 8 (1):104–107.
- Swaddle JP, Cathey MG, Correll M, Hodgkinson BP (2005) Socially transmitted mate preferences in a monogamous bird: a non-genetic mechanism of sexual selection. *P Roy Soc B-Biol Sci* 272 (1567):1053–1058.
- Tobler M, Schlupp I, Plath M (2011) Costly interactions between the sexes: combined effects of male sexual harassment and female choice? *Behav Ecol* 22 (4):723–729.
- Uller T, Johansson LC (2003) Human mate choice and the wedding ring effect. *Hum Nat* 14 (3):267–276.
- Valone TJ (2007) From eavesdropping on performance to copying the behavior of others: a review of public information use. *Behav Ecol Sociobiol* 62 (1):1–14.
- Valone TJ, Templeton JJ (2002) Public information for the assessment of quality: a widespread social phenomenon. *Philos T Roy Soc B* 357 (1427):1549–1557.
- Warner RR, Wernerus F, Lejeune P, van den Berghe E (1995) Dynamics of female choice for parental care in a fish species where care is facultative. *Behav Ecol* 6 (1):73–81.
- Waynforth D (2007) Mate Choice Copying in Humans. *Hum Nat* 18 (3):264–271.
- Westneat DF, Walters A, McCarthy TM, Hatch MI, Hein WK (2000) Alternative mechanisms of nonindependent mate choice. *Anim Behav* 59 (3):467–476.
- White DJ, Galef BG, jr (1999) Mate choice copying and conspecific cueing in Japanese quail, *Coturnix coturnix japonica*. *Anim Behav* 57 (2):465–473.
- White DJ, Galef BG, jr (2000) Social effects on mate choices of male Japanese quail, *Coturnix japonica*. *Anim Behav* 59:975–979.
- Widemo MS (2005) Male but not female pipefish copy mate choice. *Behavioral Ecology* 17 (2):255–259.
- Witte K, Nöbel S (2011) Learning and Mate Choice. In: Brown C, Laland K, Krause J (eds) *Fish cognition and behavior*, 2nd edn. Wiley-Blackwell, Ames, IA, pp 81–107.
- Witte K, Noltemeier B (2002) The role of information in mate-choice copying in female sailfin mollies (*Poecilia latipinna*). *Behav Ecol Sociobiol* 52 (3):194–202.
- Witte K, Ryan MJ (1998) Male body length influences mate-choice copying in the sailfin molly *Poecilia latipinna*. *Behav Ecol* 9 (5):534–539.
- Witte K, Ryan MJ (2002) Mate choice copying in the sailfin molly, *Poecilia latipinna*, in the wild. *Anim Behav* 63 (5):943–949.
- Witte K, Ueding K (2003) Sailfin molly females (*Poecilia latipinna*) copy the rejection of a male. *Behav Ecol* 14 (3):389–395.
- Wong BBM, McCarthy M (2009) Prudent male mate choice under perceived sperm competition risk in the eastern mosquito fish. *Behav Ecol* 20 (2):278–282.
- Yorzinski JL, Platt ML, Reby D (2010) Same-Sex Gaze Attraction Influences Mate-Choice Copying in Humans. *PLoS ONE* 5 (2):e9115.
- Ziege M, Hennige-Schulz C, Mücksch F, Bierbach D, Tiedemann R, Streit B, Plath M (2012) A comparison of two methods to assess audience-induced changes in male mate choice. *Current Zoology* 58 (1):84–94.

Chapter 37

SEXUAL SELECTION UNDER PARENTAL CHOICE: MATING STRATEGIES AND REPRODUCTIVE SUCCESS

Menelaos Apostolou and Marianna Zacharia

University of Nicosia, Nicosia, Cyprus

ABSTRACT

Across pre-industrial human societies mating is regulated, with arranged marriage, in which male parents select spouses for their female relatives, being the primary mode of long-term mating. This chapter reviews the model of sexual selection under parental choice that offers a good account for these patterns of human mating. This model postulates that parent-offspring conflict over mating induces parents to control the mate choices of their children, and the spouses they select for them are individuals who conform best to their preferences. By doing so, parents become a significant evolutionary force affecting the course of sexual selection. The model is applied in order to achieve a comprehension of the evolution of specific mating strategies. In particular, it is argued that in a context where mate choice is regulated, at least three mating strategies can thrive: addressing parental choice, addressing female choice and circumventing parental and female choice by force.

INTRODUCTION

When Darwin introduced his theory of sexual selection, which postulates that certain traits have evolved to permit individuals to attract and retain mates, he identified female choice to be a primary sexual selection force (Darwin, 1871). In particular, he argued that since it is men who strive to gain access to women and not the other way round, traits which make a man to be more attractive to women are selected and increase in frequency in the population.

Darwin did not provide an explanation of why it is typically men who struggle to gain access to women and not the opposite. This theoretical gap that was successfully addressed and closed a hundred years later by Robert Trivers. Trivers (1972) emphasized that women devote, supply and provide more to their offspring than men which makes them the scarce reproductive resource over which men strive to gain access. Trivers' argument contributed considerably to

the theory of sexual selection under female choice, as it offered more solid foundations. The aforementioned, combined with the observation that in Western societies men seem to sacrifice substantial resources to gain access to women, initiated and enhanced intense theorizing and empirical work centered on the model of sexual selection under female choice.

For example, Zahavi (1975) suggested that costly traits, such as the peacock's tail, evolved to reliably communicate male fitness to females. This theory was utilized by Zahavi & Zahavi (1997) and Miller (2000) who articulated that men have evolved a tendency to engage in costly behaviors like in sports in order to communicate their abilities to women in a cheat-proof fashion. Likewise, since sexual selection is directed by female choice, female preferences direct its course. Accordingly, Buss (2003) and other researchers examined mate preferences, with a focus on female mate preferences.

The model of female choice, although it bears some good theoretical basis and appears to apply to a Western context, it has a fundamental weakness namely, it is not consistent with the patterns of human mating found in modern and historical pre-industrial societies. In these societies, women are not free to exercise choice, as their mating decisions are regulated by their parents, typically their fathers, who select spouses for them (Apostolou, 2010b; Blood, 1972). Since the way of life over the previous two million years of human evolution more closely bears a resemblance to the way of life of modern pre-industrial societies compared to that of Western societies, it is reasonably possible that these patterns of mating portrayed mate choice during most of human evolution (Apostolou, 2010b). This delineates that the model of sexual selection under female choice is insufficient in explaining how sexual selection works in our species.

A different model has been recently put forward which better fits the mating patterns observed in pre-industrial human societies (Apostolou, 2007b, 2010b). In this model, conflicting interests over mating induce parents to regulate mate choice and choose individuals who comply best with their preferences, as spouses for their children. The purpose of this chapter is to review this model and apply it in understanding the evolution and the prevalence of male mating strategies. The fundamental constituent of the model is parent-offspring conflict over mating, which will be discussed next.

PARENT-OFFSPRING CONFLICT OVER MATING

All children's genes come from their parents, yet not all of parents' genes are inside their children. Consequently, as parents and children are not genetically identical, the traits of a mating candidate which are most beneficial and desirable to children are not essentially so for their parents. For this reason, parents and children have evolved asymmetrical preferences over traits, which offer them asymmetrical benefits (Apostolou, 2007a; 2008; Buunk, Park, & Dubbs, 2008; Trivers, 1974).

Perhaps the most characteristic case is genetic quality. The coefficient of relatedness of parents to children is 0.5, but the coefficient of relatedness of grandparents to grandchildren is half as much that is, 0.25. Consequently, the probability of a particular gene of an individual being passed into the next generation by a spouse or in-law would be 50% or 25%, respectively. This means that parents obtain less genetic benefits from a prospective mate of high genetic quality than their offspring (Apostolou, 2007a; Buunk et al., 2008). Beauty is a proxy of genetic

quality (Thornhill & Gangestad, 1993), and therefore, is preferred more in a spouse than in an in-law (Apostolou, 2008; Buunk et al., 2008). Preference divergence is not limited solely to beauty, as research has demonstrated that parents and offspring have diverse preferences concerning family and religion background, and exciting personality (Apostolou, 2008; Buunk et al., 2008; Perilloux, Fleischman, & Buss, 2011).

Asymmetrical preferences lead to conflict among parents and children as a result of the trade-off nature of mating: constrained by their own mate value, parents and offspring have to make compromises with respect to a mating candidate's desirable qualities. However, since the two parties do not share identical preferences, they are going to make dissimilar compromises (Apostolou, 2008; Buunk et al., 2008). Therefore, if children solely exercise mate choice, they will compromise more than their parents would like over certain traits, like industriousness and social status, in order to get an attractive spouse. Conversely, genetic quality associated with beauty is less important to parents, thus the advantage from this trait is insufficient to compensate for the loss of other desirable traits. For that reason, the children's mate choice imposes a cost upon parents in the form of loss in desirable traits (Apostolou, 2008).

A study examining the tradeoffs of in-law and mate choice, asked parents to design an ideal spouse for their children, and their children to design an ideal spouse for themselves (Apostolou, 2011). Participants were given a budget of mate points and they were asked to allocate them across several desirable traits. The study demonstrated that children assigned fewer points in traits, such as similarity in religion and good family background in order to get more of exciting personality and beauty. Their parents, on the other hand, assigned fewer points on exciting personality and beauty in order to get more of good family background and similarity in religion.

These findings indicate that if children are left on their own to exercise mate choice, this choice will impose a cost on their parents with regard to loss in desirable traits. This stimulates and encourages parents to control their children's mate choices, which in turn has an impact on the course of sexual selection.

THE MODEL OF PARENTAL CHOICE

Parent-offspring conflict over mating mandates that mate choice left in the hands of offspring does not maximize the fitness of parents. As a consequence, substantial evolutionary pressure is exercised on the parents to regulate the mating decisions of their children.

The prolonged period during which children rely on parental investment for survival and reproduction and the fact that parents and their kin are physically stronger than their children permit parents to control mate choice (Apostolou, 2007b). By doing so, they become effectively a sexual selection force, as traits that make an individual more likely to be chosen as an in-law are selected and increase in frequency in the population compared to less desirable ones.

Since females invest more in their offspring, they become the scarce reproductive resource to which males are seeking access (Trivers, 1972). Thus, by controlling this "resource", parents effectively control mating (Apostolou, 2007b). Furthermore, lenient parental control over female children possibly bears detrimental consequences, such as unwanted pregnancy (Perilloux, Fleischman, & Buss, 2008). Consequently, parental control is biased against female offspring, something that is also facilitated by the fact that daughters are physically weaker than

sons. Parental choice is also asymmetrical in favor of male parents. By means of greater physical strength, exclusive use of weaponry and control of political institutions, male parents have more influence over their offspring's mate choices than female parents (Apostolou, 2007b).

Nevertheless, parental control over mating is not absolute. Parents cannot always effectively "protect" their children, and their control diminishes as they become older and physically weaker and their offspring stronger and less dependent on parental investment (Apostolou, 2007b). Lastly, control over mating produces evolutionary pressures on children to evolve adaptations, such as psychological manipulation of their parents to accomplish their goals concerning mate choice (Trivers, 1974).

SEXUAL SELECTION UNDER PARENTAL CHOICE IN HUMAN SOCIETIES

Sexual Selection under Parental Choice in Pre-Industrial Societies

In pre-industrial societies offspring depend on their parents and close kin for food, support and protection. Accordingly, we expect that parents in these societies dominate mate choice. Apostolou (2007b), in an attempt to examine whether this is the case, he coded the mating patterns of a sample of 190 foraging societies spread across the globe. In these societies, parents exercise considerable control over the mating behaviour of their offspring, as the most common type of marriage is arranged marriage, since in 70% of the societies in the sample marriages were arranged. Marriage as the result of free courtship is found only in a small number of cases (4.3%). In arranged marriage, men, usually fathers, regulate the selection process, but it is not rare for women to be also influential. Finally, there is still the possibility for offspring to exercise choice by escaping, divorcing or forming extramarital affairs.

Moreover, Apostolou (2010b) analysed data from the Standard Cross Cultural Sample and found that in agropastoral societies the most frequent marriage type is arranged marriage, and marriage arrangements are controlled by males. Additionally, in agropastoral societies women are more likely, compared to men, to have an arranged marriage and to be married at a younger age. Nevertheless, children can still exercise choice by employing various means, such as divorce and extramarital affairs. Moreover, comparisons between agropastoral and foraging societies revealed that parents exercise more control over the mating decisions of their children in the former than in the latter societies. Male parents are also more frequently reported to dominate marriage arrangements in agropastoral societies than in hunting and gathering ones.

These patterns of mating have significant evolutionary implications. More specifically, the separation between agropastoral and foraging pre-industrial societies resembles the two most important stages of human evolution: humans lived as hunters and gatherers for the most part of human evolution, than over the last 10,000 years, there was a shift to a mode of subsistence based on agriculture and animal husbandry (Lee & DeVore, 1968). Therefore, by studying modern pre-industrial societies, we can make valid inferences concerning mating patterns in ancestral human societies (Ember, 1978; Lee & DeVore, 1968). Thus, as parental choice is dominant in modern foraging societies, we can infer that it was also dominant in ancestral

foraging ones (Apostolou, 2007b). This is supported by research based on phylogenetic analysis which endeavors to recreate the conditions of ancestral societies (Walker et al., 2011).

Moreover, it can be further inferred that the agricultural revolution at the onset of the Holocene (10,000 years ago) gave rise to a subsequent increase in the influence of parents, and especially male parents, over the mating choices of their children (Apostolou, 2010b). As a matter of fact, we do not need to make assumptions for pre-modern agropastoral societies, as data for the mating patterns in these societies is already available from historical sources. This evidence indicates that parents, primarily male ones, had been arranging the marriages of their offspring, and in particular of their daughters, in ancient Greece (Vrissimtzis, 1997), ancient Rome (Balsdon, 1962), Imperial China (Gernet, 1970) and in pre-Victorian England (Stone, 1990).

Sexual Selection under Parental Choice in Post-Industrial Societies

In contrast to pre-industrial societies, the extensive educational needs of technology-based societies require marriage to be delayed, in anticipation of the conclusion of one's training. For that reason, individuals are married later in life, at a period when they are relatively independent from their families. Moreover, social protection systems, such as the welfare state, police, human rights etc., alleviate individuals' dependence on their family. Similarly, the legal system prevents the use of physical force, so parents are restrained from freely imposing their will by means of physical punishment. Since parents cannot use their children's reliance on their investment and their physical strength as means of coercion, they cannot directly dominate their mating choices. Consequently, individuals in post-industrial societies have the benefit of autonomy and lack of restrictions concerning mate choice.

However, parental behaviour has been formed by evolution in order to attempt to dominate the mating behaviour of children, which denotes that even if parents cannot do so directly, they will attempt to do so by employing indirect means. To begin with, parents regulate considerable resources, for which their children are not unconcerned thus, these can be used to affect mate choice. For instance, manipulation of inheritance rights can give parents, particularly rich ones, a lever for effective manipulation (Apostolou, in press). Moreover, parents also employ means, like 'cajolery, persuasion, appeals to loyalty, and threats' (Apostolou, in press; Sussman, 1953) to influence the mating behaviour of their offspring.

In the same vein, contemporary Chinese parents cannot inflict their choices on their children, but they still persist to exercise the function of the facilitator through their own social systems (Ikels, 1985). Chinese parents in the USA attempt to create situation in which their children can get together with other Chinese children of a desirable background. For example, they may stage a barbecue when an eligible relative from out of state comes to visit (Ikels, 1985). Although the real potency of parental choice in post-industrial societies remains to be estimated, the evidence currently existing indicates that parents are influential over mating choices.

IN-LAW PREFERENCES

Marriage adds a new member to the family unit, who contributes significantly to its survival and reproductive efforts. In-laws integrate into their new family, supporting its subsistence activities (hunting, gathering, animal herding etc.), offering political support through their family relations, physical support in case of external threats (raids, wars, feuds etc.), their genes, and so on. However, in-laws differ. For example, some are diligent while others are indolent; some enjoy vigorous health while others suffer from chronic illness; some come from wealthy and powerful families while others have humble origins. Consequently, since prospective in-laws vary in their qualities, they also vary in the advantages they can offer to parents.

This reveals that parents have been experiencing strong evolutionary pressures to evolve preferences that permit them to select those in-laws who are most valuable to them (Apostolou, 2007a, 2010a). In contrast, parents who do not share such preferences are just as likely to choose a hard-working in-law as a lazy one. Therefore, the former gain a selective advantage over the latter, successfully augmenting the frequency of the genes that confer the disposition for these preferences in the population (Apostolou, 2007a). The result of this process is that parents have been equipped with well-defined in-law preferences.

Given that not all the traits in an in-law provide parents with the same benefits, parents are likely to have stronger preferences for qualities which are more beneficial for them. Moreover, parental preferences are expected to be dependent upon the sex of the in-law. Specifically, the division of labour assigns or attributes diverse tasks and roles to men and women in a given society. For instance, in all recorded human cultures, men dominate over women in controlling resources (Whyte, 1978). In modern Western post-industrial societies, the wealthiest self-made individuals are men, while the wealthiest women are in the 'Rich List' because they inherited, or married into wealth. Accordingly, parents should value traits related to the capacity to acquire resources more in a son-in-law than in a daughter-in-law.

Two studies have been conducted to date which endeavour specifically to identify parental preferences. To begin with, Apostolou (2007a) requested from a group of British parents to rate the desirability of several traits in a prospective son-in-law and daughter-in-law. Parents appeared to have a hierarchy of preferences in which specific traits were deemed to be more essential than others. In particular, traits such as kind and understanding, good health, and good earning capacity were located at the top of the parental hierarchy, followed by traits such as education/intelligence, good cook/housekeeper, and good family background located in the middle of the parental hierarchy. Traits such as physically attractive and chastity were located at the bottom of the parental preference hierarchy.

Moreover, parents rated traits in a different way in a daughter-in-law and a son-in-law. In particular, qualities such as ambition/industriousness, good financial prospects, and wealth were preferred more in a son-in-law than in a daughter-in-law, while traits, such as good housekeeper and good looks were preferred more in a daughter-in-law than in a son-in-law. Finally, both mothers and fathers were found to have the same opinion in terms of what they seek in an in-law of either sex.

The second study, using anthropological evidence from 67 pre-industrial societies, identified 13 desirable qualities that parents seek in an in-law (Apostolou, 2010a). Among the most frequently reported traits were good character, good family background, industry and

good worker, and subsequently favourable social status, wealth, and similar family social status. The least commonly reported traits were good looks and chastity. It was also found that traits such as good economic prospects, wealth and favourable social status were regarded highly more in a son-in-law than a daughter-in-law, while chastity was regarded highly more in a daughter-in-law than a son-in-law.

Although not purposely intended to pinpoint in-law preferences, other studies provide information on the traits parents desire in an in-law. More specifically, Hynie et al. (2006) investigated in-law preferences of Chinese immigrants in North America and results reveal that good social status and understanding are among the most preferred traits. Borgerhoff Mulder (1988) reported that, among the pastoral Kipsigies in Kenya, parents prefer as sons-in-law individuals who enjoy high social status, are wealthy, educated, have a good character and are diligent. Yu, Proulx, and Shepard (2007) found that Matsigenka women in Peru prefer men with masculine faces as sons-in-law and interpreted this finding as a preference for good providers, as masculine men are, on average, perceived as better resource providers. Finally, Apostolou (2007b) found that among foragers, parents seek sons-in-law who are good hunters and have a good family background, while they prefer daughters-in-law who are industrious and come from good families.

The knowledge of parental preferences, and how these deviate from the preferences of their children, can provide us with an understanding of the prevalence of certain mating strategies in the population.

MALE MATING STRATEGIES

Polymorphism in Mate Choice

Men seek to gain reproductive access to women, but the gatekeepers to this access are other men (usually fathers) who search for sons-in-law with particular traits. This creates substantial evolutionary pressures for men to attract interest as in-laws from parents and particularly from male parents. Nevertheless, women's mate choices are not entirely restricted by their parents, since they can engage in informal relationships before marriage and in extramarital relationships subsequent to marriage, while they can also divorce the spouses their parents have chosen for them (Apostolou, 2010b). Consequently, there is freedom for women to exercise choice, which delineates that there have also been evolutionary pressures on men to discover wide-ranging ways to appeal to women as mates.

To put it differently, there have been at least two mating niches that men can address: The parental niche (being selected through parental choice) and the female niche (being selected through female choice). The evolutionary effect of the existence of two niches is that men may have evolved to attend to both parental choice and female choice, or some men may have evolved to attend to parental choice and others to attend to female choice.

The latter scenario is more likely because, in cases where a species occupies multiple niches, a polymorphic equilibrium (e.g., two or more distinct specializations) is usually more optimal than a monomorphic one (e.g., a single specialization), as specialists are more efficient than generalists (Wilson, 1994). This is due to the fact that a male specializing in appealing to parental choice obtains reproductive advantages from marriage without having to bear the cost

of appealing to female choice. A male specializing in appealing to female choice may obtain reproductive advantages by engaging in different casual relationships without bearing the costs of appealing to parental choice.

As a species we are a polymorphic: Some individuals have blond hair, while others have dark hair; some individuals have blue eyes and others have green eyes. We are also polymorphic in terms of behavior. Some individuals are hard-working and others are lazy; some are kind and others are evil and so on. Approximately half of the variations between people with respect to personality are due to the fact that they share diverse genes (Plomin et al., 2008). Accordingly, polymorphism, being frequent in our species, should also exist in mating behavior, mainly in light of the fact that there is more than one mating niche.

Our knowledge of in-law and female preferences permits us to hypothesize how the two male morphs are likely to be structured. To begin with, men specializing in attracting interest from parents are likely to give emphasis to their personality and family background, and pay less attention to their physical appearance. They are also likely to be concerned about impressing and gaining the respect of other men. Conversely, men specializing in attracting the interest of women are likely to accentuate more their physical appearance, being charming, but pay less attention in communicating their good family background.

Assuming that parental choice has been dominant throughout human evolution, we would expect that the male morph that addresses parents should have been much more frequent than the male morph that addresses women. To put it in a different way, the parental niche has been much bigger than the female niche, leading to an equilibrium where the majority of men specialize in attracting interest from parents, and a minority in attracting interest from women.

Overall, men are likely to have evolved two strategies in order to obtain access to the opposite sex, one through females themselves and one through their parents. It is also of paramount importance to stress that there is a third approach, which permits men to force mating access to women.

THE EVOLUTION OF RAPE

Circumventing Parental Choice

When they look for sons-in-law, parents are in search for men who are endowed with qualities which are advantageous for them. Evidence from modern and historical societies reveals that parents are interested in finding sons-in-law who have a high resource-generating capacity, have a good family background, control wealth, and have a good character (Apostolou, 2010a, 2012; Borgerhoff Mulder, 1988; Koster, 2011). Men who lack desirable qualities, such as social status and find themselves in a context where mating is regulated, endure a substantial reproductive loss, since they are inclined to be excluded from mating or have to compromise by settling for a woman of low mate value.

To be more precise, parents with daughters of elevated mate value (e.g., attractive, young, etc.) are not going to be willing to give them as wives to men of an inferior mate value (e.g., men who lack desirable traits). Therefore, if a man of inferior mate value seeks to gain parental approval, he needs to address parents whose daughters also have an inferior mate value (e.g., are older, unattractive, have children from previous marriages, etc.), and who are thus less

reluctant to accept him as a son-in-law. In fact, this man will endure substantial fitness losses, as he will have to either settle for a woman of low mate value or opt out from the reproductive process.

In addition to parental choice, there are two related factors which also operate to alleviate the ability of men, who have a low mate value, to attract desirable wives. One is polygyny, which is practiced in most of the pre-industrial societies and acts to exclude low status men from gaining reproductive access (Blood, 1972). The reason is that high mate value men are able to attract multiple wives, leaving those at the bottom of the hierarchy single. In the same vein, hypergyny, that is, women marrying up the social hierarchy, a common practice in many societies, pulls women out of the lower classes, leaving a lot of low-status men without wives (Boone, 1986).

In this context, then, a forced-sex mating strategy can decrease the reproductive costs that a man would probably suffer by circumventing parental choice, permitting sexual access to women of high mate value, women that these low mate value men could not have accessed in any other way. This predicts that rapists are mostly young men of low mate value. This prediction is consistent to the findings of several studies (Thornhill & Palmer, 2000; Thornhill & Thornhill, 1983). This predicts also that the victims of rape should predominantly be high mate value women, who would not be accessible by these rapists in any other way. Consistent with this prediction, the majority of rape victims are frequently young women at the peak of their fertility (Greenfield, 1997; Kilpatrick et al., 1992; Thornhill & Palmer, 2000).

Circumventing Female Choice

In societies where mating is controlled, women still have some space to exercise mate choice. For instance, they can exercise mate choice by forming extramarital relationships or taking a divorce and getting married later, when their parents are no longer alive (Apostolou, 2010b). Consequently, a man can address female choice directly. However, women are in search of men with desirable traits, such as high resource acquisition capacity and physical attractiveness (Buss, 2003). This indicates that a man who lacks these traits is unlikely to be able to successfully appeal to female choice if he addresses high mate value women. He will be more effective, however, if he addresses low mate value women, such as older ones who are less likely to receive offers from men of high mate value. Nevertheless, getting married to an older woman will not enhance his fitness much, as the reproductive years of that woman have already been or will soon be exhausted.

A man will inevitably need to address female choice within marriage. In particular, a man's mate value may decrease if for instance he suffers an injury from a fight, or experiences a status loss. This will make his wife perceive him as a less desirable mate. Additionally, in a context where mating choice is regulated, a man may effectively pass through parental choice, but this does not mean that he will instantly obtain sexual access to his wife, as the latter may decline her husband's requests to permit him to gain sexual access.

Consequently, a forced-sex mating strategy can offer fitness benefits by permitting men to circumvent female choice. The female choice within marriage cases predicts further that rape would also occur within marriage. Consistently with this prediction, between 10% and 26% of women report experiencing marital rape (Russell, 1990; Watts, Keogh, Ndlovu, & Kwaramba, 1998). It is further predicted that rape will be more likely to occur when the mate value of the

husband diminishes (e.g., he loses his job) or the mate value of the wife augments (e.g., she loses weight).

Finally, in societies which practice arranged marriage, rape within marriage enhances the strength of parental choice as a sexual selection force because it makes the choices of parents on their daughters consequential. This predicts that in these societies, the parents of a man may prompt him to force sex on his wife if she rejects his requests to obtain sexual access, and the parents of the bride may also encourage or at least accept such action. This prediction also remains to be addressed by future research.

In a pre-industrial context, a forced-sex mating strategy can endow substantial reproductive advantages to inferior mate value men by enabling them to circumvent parental and female choice. This strategy is not of course cost-free, otherwise nearly all men would adopt it most of the times. In particular, a rapist is likely to meet the fierce reaction of a woman's parents and other kin, he is likely to meet the resistance of the woman, which might result in physical injury, and he is likely to face the vengeance of the woman, particularly in the case of marital rape.

CONCLUSION

The model of parental choice predicts that for the greatest part of the period of human evolution, parental choice constituted a significant sexual selection force, with individual mate choice still likely to be exercised. This prediction is consistent with evidence from the anthropological and historical records which demonstrate that across human pre-industrial societies mate choice is regulated, but there is also room for individuals to exercise their own choice.

Based on the abovementioned, it has been argued that men have evolved at least three diverse strategies in order to obtain sexual access to the opposite sex: they can address female choice, parental choice, or attempt to circumvent both selection mechanisms by force i.e., through rape. Future research needs to explore each strategy and assess their relevant fitness payoffs.

REFERENCES

- Apostolou, M. (2007a). Elements of parental choice: the evolution of parental preferences in relation to in-law selection. *Evolutionary Psychology*, 5, 70-83.
- Apostolou, M. (2007b). Sexual selection under parental choice: the role of parents in the evolution of human mating. *Evolution and Human Behavior*, 28, 403-409.
- Apostolou, M. (2008). Parent-offspring conflict over mating: The case of beauty. *Evolutionary Psychology*, 6, 303-315.
- Apostolou, M. (2010a). Parental choice: What parents want in a son-in-law and a daughter-in-law across 67 pre-industrial societies. *British Journal of Psychology*, 101, 695-704.
- Apostolou, M. (2010b). Sexual selection under parental choice in agropastoral societies. *Evolution and Human Behavior*, 31, 39-47.
- Apostolou, M. (2011). Parent-offspring conflict over mating: Testing the tradeoffs hypothesis. *Evolutionary Psychology*, 9, 470-495.

- Apostolou, M. (2012). Sexual selection under parental choice: Evidence from sixteen historical societies. *Evolutionary Psychology*, 10, 504-518.
- Apostolou, M. (in press). Do as we wish: Parental tactics of mate choice manipulation. *Evolutionary Psychology*.
- Balsdon, J. P. V. D. (1962). *Roman women: their history and habits*. London: The History Book Club.
- Blood, R. O. (1972). *The family*. New York: The Free Press.
- Boone, J. L. III (1986). Parental investment and elite family structure in preindustrial states: A case study of late medieval-early modern Portuguese genealogies. *American Anthropologist, New Series*, 88, 859-878.
- Borgerhoff Mulder, M. (1988). The relevance of the polygyny threshold model to humans. In C. Mascie-Taylor & A. Boyce (Eds.), *Human mating patterns* (pp. 209-230). Cambridge: Cambridge University Press.
- Buss, D. M. (2003). *The evolution of desire: Strategies of human mating* (2nd ed.). New York: Basic Books.
- Buunk, A. P., Park, J. H. & Dubbs, S. L. (2008). Parent-offspring conflict in mate preferences. *Review of General Psychology*, 12, 47-62.
- Darwin, C. (1871). *The descent of man and selection in relation to sex*. London: Murray.
- Ember, C. R. (1978). Myths about hunter-gatherers. *Ethnology*, 17, 439-448.
- Gernet, J. (1970). *Daily life in China on the eve of the Mongol invasion 1250-1276*. Stanford: Stanford University Press.
- Greenfield, L. (1997). *Sex offenses and offenders: An analysis of data on rape and sexual assault*. Washington, DC: Bureau of Justice Statistics, U.S. Department of Justice.
- Hynie, M., Lalonde, R. N. & Lee, N. (2006). Parent-child value transmission among Chinese immigrants to North America: The case of traditional mate preferences. *Cultural Diversity and Ethnic Minority Psychology*, 12, 230-244.
- Ikels, C. (1985). Parental perspectives on the significance of marriage. *Journal of Marriage and the Family*, 47, 253-264.
- Kilpatrick, D., Edmunds, C. & Seymour, A. (1992). *Rape in America: A report to the nation*. Arlington, VA: National Victim Center.
- Koster, J. (2011). Hypothetical rankings of prospective husbands for female kin in lowland Nicaragua: Consensus analysis indicates high agreement and associations with wealth and hunting skill. *Evolution and Human Behavior*, 32, 356-363.
- Lee, R. B. & Devore, I. (1968). *Man the hunter*. New York. Aldine.
- Miller, G. (2000). *The mating mind*. London: BCA.
- Perilloux, C., Fleischman, D. S. & Buss, D. M. (2008). The daughter guarding hypothesis: Parental influence on, and emotional reaction to, offspring's mating behavior. *Evolutionary Psychology*, 6, 217-233.
- Perilloux, C., Fleischman, D. S. & Buss, D. M. (2011). Meet the parents: Parent-offspring convergence and divergence in mate preferences. *Personality and Individual Differences*, 50, 253-258.
- Plomin R., DeFries, J. C., McClearn, J. C., McGuffin, P. (2008). *Behavioral genetics* (5th ed.). London: Worth Publishers.
- Russell, D. E. H. (1990). *Rape in marriage* (Rev.ed.). Indianapolis: Indiana University Press.
- Stone, L. (1990). *The family, sex and marriage in England 1500-1800*. London: Penguin Books.

- Sussman, M. B. (1953). Parental participation in mate selection and its effect upon family continuity. *Social Forces*, 1, 76-81.
- Thornhill, R. & Gangestad, S. (1993). Human facial beauty: Averageness, symmetry and parasite resistance. *Human Nature*, 4, 237-269.
- Thornhill, R. & Palmer, C. T. (2000). *A natural history of rape: Biological bases of sexual coercion*. Cambridge, Massachusetts: The MIT Press.
- Thornhill, R. & Thornhill, N. (1983). Human rape: An evolutionary analysis. *Ethology and Sociobiology*, 4, 137-173.
- Trivers, R. L. (1972). Parental investment and sexual selection. In B. Campell (Ed.), *Sexual selection and the descent of man: 1871-1971* (136-179). Chicago: Aldine.
- Trivers, R. (1974). Parent offspring conflict. *American Zoologist*, 24, 249-264.
- Vrissimtzis, N. A. (1997). *Love, sex & marriage in ancient Greece* (3rd ed.). Athens: Polygrama.
- Walker, R. S., Hill, K. R., Flinn, M. V. & Ellsworth, R. M. (2011). Evolutionary history of hunter-gatherer marriage practices. *PLoS ONE*, 6, e19066.
- Watts, C., Keogh, E., Ndlovu, M. & Kwaramba, R. (1998). Withholding of sex and forced sex: Dimensions of violence against Zimbabwean women. *Reproductive Health Matters*, 6, 57-65.
- Whyte, M. K. (1978). *The status of women in preindustrial societies*. Princeton, NJ: Princeton University Press.
- Wilson, D. S. (1994). Adaptive genetic variation and human evolutionary psychology. *Ethology and Sociobiology*, 15, 219-235.
- Yu, D., Proulx, S. R. & Shepard, G. H. (2007). Masculinity, culture and the Paradox of the Lek. In V. Swami & A. Furnham (Eds.), *Body beautiful: Evolutionary and sociocultural perspectives* (88-107). London: Palgrave Macmillan.
- Zahavi, A. (1975). Mate selection: A selection for a handicap. *Journal of Theoretical Biology*, 53, 205-214.
- Zahavi, A. & Zahavi, A. (1997). *The handicap principle: A missing piece of Darwin's puzzle*. New York: Oxford University Press.

Chapter 38

MELON GERMPLASM CHARACTERISTICS, DIVERSITY, PRESERVATION AND USES

C. Mallor and A. Díaz*

Unidad de Hortofruticultura, Instituto Agroalimentario de Aragón (IA2)
(CITA-Universidad de Zaragoza), Zaragoza, Spain

ABSTRACT

The need for germplasm banks that safeguard the melon genetic resources is more than justified by the genetic erosion aggravated in the last few decades, not only in the cultivated materials, but also in traditional landraces and wild relatives. The classical and new technologies employed in all stages, from the sample prospecting to resources management, with the conservation and evaluation of the plant resources in between, are described. An added value to these germplasm collections is the use of the genetic resources there preserved in the melon breeding programs. The access to wild genetic resources make possible to exploit them, for instance, for germplasm enhancement incorporating disease resistances in elite varieties. In this sense, conventional breeding methods have rendered unquestionable benefits to agriculture, which can be accelerated by the appearance of new biotechnological tools and genomic resources as a result of the increasing number of vegetable species whose genomes have been or are being sequenced. All germplasm banks, and those preserving vegetable resources like melon in particular, will have to face up to the challenge of characterizing genetically their collections in order to maximize their use in breeding strategies assisted by molecular tools, like molecular markers. At the same time, the use of molecular markers can help to efficiently manage the resources by the creation of core collections. This would alleviate the problems derived from the high number of entries in many of them that is compromising some of the purposes for which they have been created, the conservation and the use of the genetic diversity originally prospected. The advantages in terms of labor and economical investment of preserving, characterizing and using a reduced subset of samples without a considerable sacrifice of the genetic diversity, are undeniable.

* Corresponding Author's Email: adiazb@cita-aragon.es.

1. GENERAL INTRODUCTION: THE NEED FOR GERMPLASM BANKS

The need for germplasm banks that safeguard the vegetable genetic resources is more than justified by the genetic erosion aggravated in the last few decades, not only in the cultivars, but also in traditional landraces and wild relatives. The case of melon (*Cucumis melo* L.) serves as an example of this dwindling in variety. At the beginning of the twentieth century, the commercial seed houses in USA offered more than 300 melon varieties and only eight decades later, less than 30 of them could be found in the National Center for Genetic Resources Preservation (NCGRP, USA). On top of this preservation of the plant genetic biodiversity, germplasm banks can also be considered as a gene reservoir to breed crops.

Until 2009, there were 21,268 germplasm banks included in the World Information and Early Warning System (WIEWS) on Plant Genetic Resources for Food and Agriculture (PGRFA) spread all around the world (FAO, 2016), though not uniformly, as for instance in the African continent, there are many countries without any germplasm bank (Figure 1).

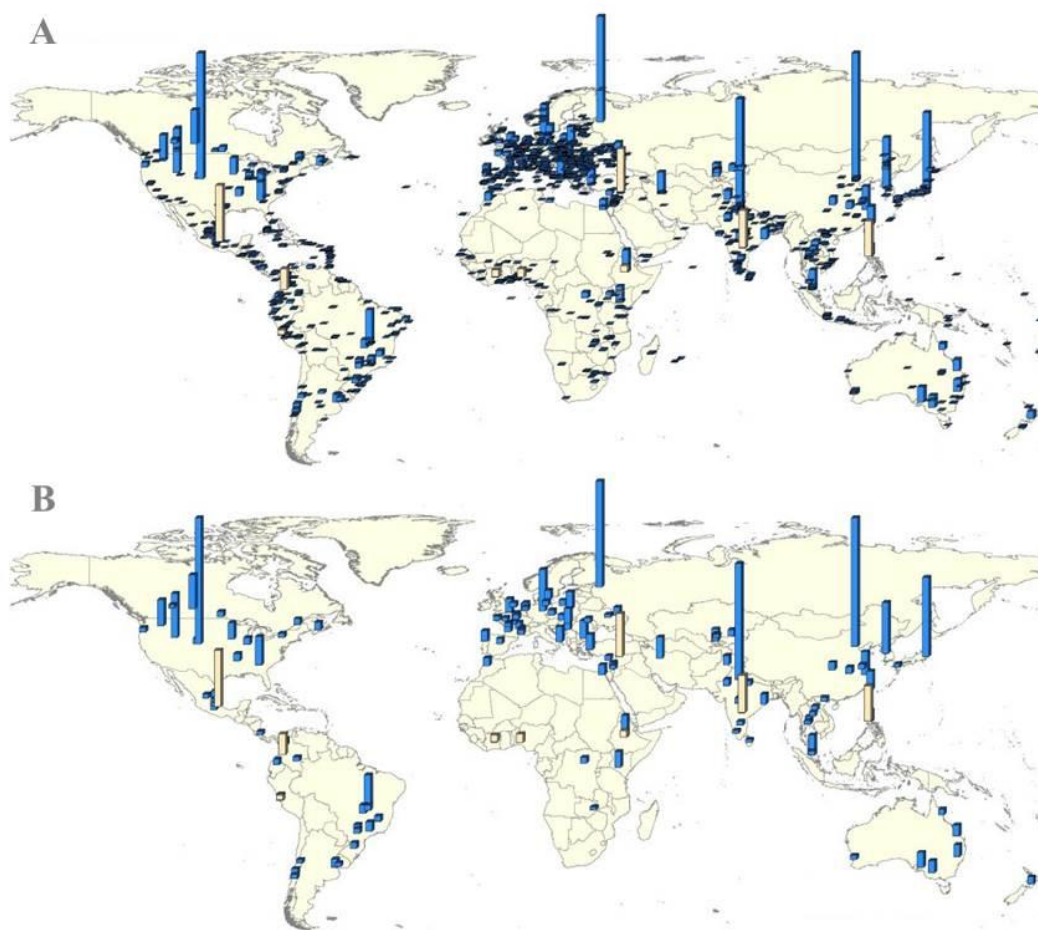


Figure 1. Distribution of all the germplasm banks around the world (A); and only those with more than 10,000 accessions (B) surveyed until 2009 (Source: Wiews, 2009. FAO).

There are some supranational organizations aimed at improving the accessibility to the information gathered by these germplasm banks on their respective collections. At a global level, Genesys is a portal that not only offers characterization and evaluation information about PGRFA from germplasm banks around the world, but it also supplies the seeds (Genesys, 2015). Genesys manages information of 8,679 accessions of *C. melo* coming from germplasm banks of 24 countries, mainly the United States (4,184 accessions), Spain (1612) and Russia (931) (Figure 2A). At the European level, the European Cooperative Programme for Plant Genetic Resources (ECPGR) maintains a catalogue of 4,274 accessions belonging to the *C. melo* species in the database named EURISCO (Eurisco, 2016), in which their passport data can be consulted. Twenty-one countries contribute to it with passport data of melon accessions, Spain being the one with the most significant participation (Figure 2B).

In many countries, the collections are centralized in national inventories. In the case of melon, the national collections keeping the highest number of accessions are those from Spain (1,555), Russia (931), United States (500), Germany (443), and Ukraine (406). The Spanish national inventory is managed by the National Plant Genetic Center (CRF). Among its tasks, the CRF keeps a replicate of the accessions maintained in the participant institutions (Table 1) and develops, publishes and keeps updated the inventory of the collections. In many of these banks, a great effort has been made to rescue landraces that would otherwise have disappeared by now, frequently being collected from local farmers, though the collections include all sort of accessions, from wild melon relatives to breeding lines (Figure 3).

Table 1. *Ex situ* collections of *C. melo* maintained in the institutes that participate in the Spanish national inventory, managed by the National Plant Genetic Center (CRF)

FAO code	Institute name	Organization and location	No of <i>C. melo</i> accessions	Web site
ESP026	Institute for Conservation & Improvement of Valencian Agrodiversity (COMAV)	Polytechnic University of Valencia (UPV). Valencia	759	http://www.comav.upv.es
ESP058	Institute for Mediterranean and Subtropical Horticulture (IHSM) "La Mayora"	Spanish National Research Council (CSIC); University of Málaga (UMA). Málaga	493	http://www.ihsm.uma-csic.es/
ESP027	Vegetable Germplasm Bank of Zaragoza (BGHZ)	Agrifood Research and Technology Center of Aragon (CITA). Zaragoza	235	http://sites.cita-aragon.es/BGHZ/
ESP200	Institute of Agricultural and Fishing Research and Training (IRFAP) of the Balearic Islands	Government of the Balearic Islands. Palma de Mallorca	39	http://www.caib.es/govern/organigrama/area.do?lang=es&coduo=1964
ESP198	Local Varieties Bank of the Madrid Research Institute for Rural, Agricultural and Food Development (IMIDRA)	Madrid autonomous government. Alcalá de Henares, Madrid	18	http://www.madrid.org/imidra
ESP003	Plant Germplasm Bank "César Gómez-Campo"	Technical University of Madrid (UPM). Madrid	10	http://www.bancodegermoplasma.upm.es/
ESP172	Centre for the Conservation of Agricultural Biodiversity in Tenerife (CCBAT)	Island Council of Tenerife. Santa Cruz de Tenerife	1	http://www.ccbat.es

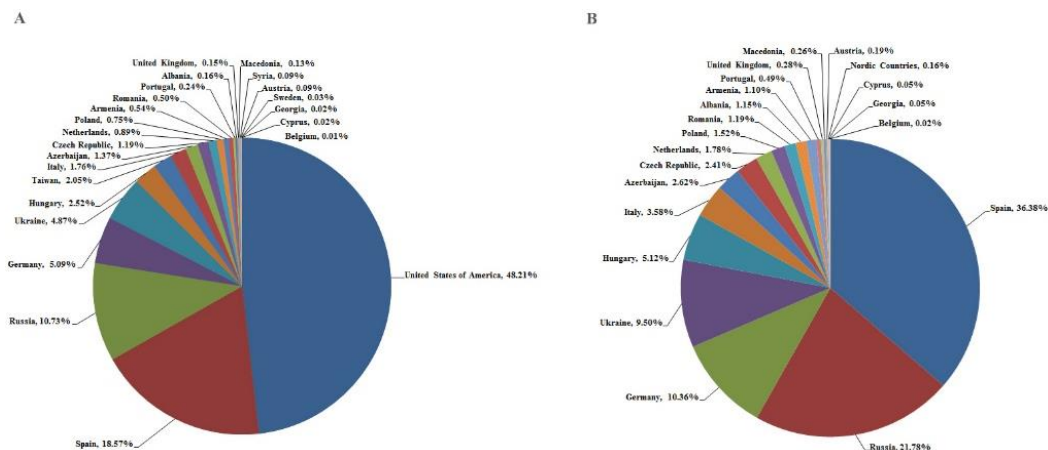


Figure 2. Contribution of each participant country to the total number of *Cucumis melo* accessions at the global database Genesys (A); and at the European database EURISCO (B).



Figure 3. Representative fruits of several melon (*Cucumis melo* L.) accessions preserved at the Vegetable Germplasm Bank of Zaragoza (BGHZ): (A) a wild relative (*C. melo* subsp. *agrestis*); (B) an exotic cultivar (*C. melo* subsp. *melo* var. *flexuosus* Greb.); (C) a landrace (Orange flesh melon from Pedrezuela, Madrid, Spain); (D) an elite cultivar ('Piel de Sapo'); (E) a breeding line (TP-21 cantaloup).

2. COLLECTION, CONSERVATION, CHARACTERIZATION AND EVALUATION OF MELON RESOURCES

A number of expeditions to collect *C. melo* accessions in an attempt to preserve the maximum variability within the species have been set on around the world in the last decades. A record of many of these prospecting works carried out between 1974 and 2012 is available at the Biodiversity International webpage (Biodiversity International, 2016; Figure 4). A total of 836 samples belonging to 664 different accessions of *C. melo* (apart from many other wild relatives within the genus *Cucumis*), collected in 42 expeditions in 30 countries, were prospected. More than the 20% of these samples were surveyed in Spain.

Melon plants produce fruits containing a type of seed called orthodox (Roberts, 1973), that is, they undergo maturation drying. This facilitates enormously their *ex situ* conservation, as they can tolerate extensive desiccation and also chilling, being stored dry at low temperatures (around -20°C) for long periods of time, maintaining an acceptable viability. This conventional storage method is easy to perform and affordable, what makes unnecessary to implement more sophisticated methods, like *in vitro* slow growth tissue culture or cryopreservation, though they can be used in melon with other purposes, as it is the case of the *in vitro* procedures.

The characterization and evaluation of melon accessions for agronomic and morphological traits promote their utilization in breeding programs. Many germplasm banks are making an immense effort to carry out the characterization of the material that is multiplied routinely. This work makes possible the identification of accessions with desirable characteristics and, hence, very interesting from a breeding point of view, but it also contributes to the proper management of the banks, for instance, allowing the identification and discard of duplicates.

As it happens in many other species, *C. melo* is integrated by a wide and very dissimilar array of forms. Firstly, it is divided into two subspecies, *melo* and *agrestis* (Jeffrey, 1980). Several horticultural groups compose each of them. Their number vary depending on the author consulted, though one of the latest and more accepted classifications (Pitrat, 2008) divides the subspecies *melo* into ten groups (*cantalupensis*, *reticulatus*, *adana*, *chandalak*, *ameri*, *inodorus*, *chate*, *flexuosus*, *dudaim*, and *tibish*) and the subsp. *agrestis* into five (*momordica*, *conomon*, *chinensis*, *makuwa*, and *acidulus*). Foreseeably, these groups will suffer modifications, as some of them are heterogeneous and the inclusion of some accessions in them is difficult.



Figure 4. Geographical location of the 836 *Cucumis melo* samples prospected between 1974 and 2012 in collecting missions supported by Biodiversity International (Source: Biodiversity International, 2012).

The availability of a guideline for cucurbit descriptors (Esquinas-Alcazar and Gulik, 1983) allowed unifying criteria for the morphological characters of interest in melon. More recently, the UPOV descriptor list for melon, which includes up to 76 characters (UPOV, 2014), has aided in selecting the traits to evaluate (Szamosi et al., 2010). In this sense, the International Plant Genetic Resources Institute (IPGRI) has published a list of minimum highly discriminating descriptors for melon (Table 2) that become easier to handle along with the maintenance of the collections (IPGRI, 2003).

Table 2. List of a minimum number of descriptors that allows an unambiguous discrimination of melon accessions

Part of the plant	Character	Categories
Whole plant	Seedling vigor	Poor; intermediate; vigorous
	Plant growth habit	Compact; dwarf; determinate; indeterminate; multilateral; other
Branch	Internode length	Very short; short; short-intermediate; intermediate; long
	Number of plant branches	
Leaf	Leaf shape	Entire; trilobate; pentalobate; 3-palmately lobed; 5-palmately lobed; other
	Leaf lobes	Shallow; intermediate; deep
	Central leaf lobe shape	Broadly ovate; shallowly oblong; narrowly oblong; elliptic; other
	Leaf color	Light green; green; dark green; variable; other
	Leaf petiole hairiness	Sparsely hispid; hispid; hispidulous; retrorse strigose; lanate; other
	Leaf persistence	Low; moderate; high
	Leaf senescence	Slight visual senescence; moderate senescence; conspicuous concurrent senescence
Inflorescence	Sex type	Monocious; andromonoecious; gynoeious female; male sterile; female sterile; other
	Days to 50% flowering	
	Flower color	White-yellow; yellow-cream; yellow; dark-yellow; orange; green; other
	Ovary pubescence length	Short; intermediate; long
	Ovary pubescence type	Spreading hairs; appressed hairs
Fruit	Fruit shape	Globular; flattened; oblate; elliptical; pyriform; ovate; acorn; elongate; scallop
	Fruit length/width ratio	
	Time of maturity	Early (<70 days); intermediate (70-90 days); late (91-110 days); very late (>110 days)
	Predominant fruit skin color	White; light-yellow; cream; pale green; green; dark green; blackish-green; orange; brown; grey; other
	Secondary fruit skin color	White; light-yellow; cream; pale green; green; dark green; blackish-green; orange; brown; grey; other
	Secondary skin color pattern	No color; speckled (spots <0.5 cm); spotted, blotchy (spots >0.5 cm); striped; short streaked; long streaked; other
	Fruit surface	Smooth; grainy; finely wrinkled; deeply wrinkled; shallowly wavy; rare warts; numerous warts; lightly corked/netted; heavily corked/netted; sutures; other
	Bitterness of mature fruit	Low; intermediate; high
Seed	Seed size	Very small; small; intermediate; large; very large
	Seed shape	Roundish; elliptical; oval; triangular; pine seed type; other
	100-seed weight	

Source: IPGRI, 2003.

With an efficient handling of the characterization and evaluation data, donor genotypes with outstanding characteristic for important traits can be identified. Furthermore, the transfer of both qualitative and quantitative traits from wild into domesticated forms has also become an attractive objective in breeding programs.

3. MELON GENETIC RESOURCES MANAGEMENT

As in any other crop, a deep knowledge of the genetic variation and structure of melon is a prerequisite for the conservation and exploitation of the biodiversity existing within the species. In this sense, the melon research community has been making use of the molecular markers available in each moment. All the markers listed below have played an important role in the identification and evaluation of the variability present in the species.

3.1. Morphological Markers

Before the development and routine use of molecular markers, a great number of morphological traits were used to identify and classify the variability present in melon. In this sense, characters related to the vegetative and flowering state of the plant, as well as to the fruit, were scored. Categories based on qualitative characters of the fruit, like aroma, taste and pulp or skin color, can also be established.

The first taxonomic classification of the species in “botanical groups” or “horticultural varieties” are based in these types of data, as those reported by Naudin (1859). One of the latest intraspecific classifications based on morphological traits is the one proposed by Pitrat (2008), who defined 15 botanical groups, belonging to two subspecies, *agrestis* and *melo*, as commented above. This subdivision has recently been confirmed with the use of molecular markers (Serres-Giardi and Dogimont, 2012; Esteras et al., 2013) though there are some discrepancies in the groups comprised in each subspecies (Serres-Giardi and Dogimont, 2012), what reveals the need of using easy to evaluate markers in taxonomic studies.

The evaluation of morphological traits has been frequently combined with agronomical (Escribano and Lazaro, 2009; Roy et al., 2012), physiological (Soltani et al., 2010) and biochemical data, like pH, total soluble solids (TSS) and the content of several polysaccharides (sucrose, glucose and fructose), organic acids (ascorbic acid) and vitamins (carotenoids) in the melon pulp (Stepansky et al., 1999; Dhillon et al., 2007; Fergany et al., 2011; Roy et al., 2012), among others. It is also common to complete the research on melon diversity using morphological traits with the study of agronomical characteristics, like yield-related parameters, and pest and disease resistance (Staub et al., 2004; Dhillon et al., 2007; Fergany et al., 2011; Roy et al., 2012). However, the use of morphological traits as markers is hindered by numerous factors, such as the influence of cultural management and environmental conditions on the phenotype. Furthermore, sometimes they reveal themselves insufficient to discriminate among different melon accessions and to assign them to the varietal groups described (Trimech et al., 2013). This is why morphological traits may only provide partial taxonomic information. To overcome this limitation, scientists have resorted to molecular markers to strengthen their results from morphological data in classification and diversity studies (Stepansky et al., 1999;

Luan et al., 2008; Sensoy et al., 2007; Phan et al., 2010; Parvathaneni et al., 2011; Sestili et al., 2011; Yildiz et al., 2014) and, the other way round, phenotypic characters has complemented molecular data, for instance, in the construction of genetic maps (Oliver et al., 2001; Perin et al., 2002; Cuevas et al., 2008), as it will be tackled in more detail in the following sections.

3.2. Biochemical Markers: Isozymes

Isozymes have been the biochemical markers most widely used in plant breeding. They represent the different biochemical forms of an enzyme codified by different alleles of the same gene (Soltis and Soltis, 1989).

The easy, quick and affordable methodology needed to employ these codominant markers favored their use in melon in areas like molecular breeding, to identify progenies in an early growth stage, what is applicable to seed purity tests (Issiki et al., 1991; Kato et al., 1998); in systematic, to determine the genetic distance with wild and cultivated relatives (Chen et al., 1997) and among melon accessions with different geographical origin (Akashi et al., 2002; McCreight et al., 2004); and in genetic mapping (Staub et al., 1998).

Isozymes have some disadvantages because they are gene products, so their expression is affected by the environment, sometimes it is tissue-specific and can be subjected to selective processes. For these reasons, scientists working on melon have been combining the use of isozymes with DNA-based markers, for instance, to explore the intraspecific variation of the species (Staub et al., 1997), or to construct a linkage map (Oliver et al., 2001).

3.3. Genetic Markers

Genetic markers offer information directly about the genotype, quite the opposite than the previous ones, which were based on the phenotypes and the gene products, respectively. This is the reason why they are presently the most extensively used markers. Among them, amenable to automation markers seem to be the most suitable ones for the management of genetic resources.

Virtually all sorts of DNA markers have been employed in melon and, in many cases, to explore, characterize and manage the diversity of the species. The markers most commonly used for this purpose have been: RFLPs (Restriction Fragment Length Polymorphisms, Botstein et al., 1980) and AFLPs (Amplified Fragment Length Polymorphisms, Vos et al., 1995), which detect the polymorphism created in the DNA by the alteration of the target site of a restriction enzyme either by punctual mutations or deletions and insertions; RAPDs (Random Amplified Polymorphic DNAs), which make use of arbitrary, short primers to amplify genomic DNA, what renders a band profile considered as a dominant marker (Welsh and McClelland, 1990; Williams et al., 1990); SCARs (Sequence Characterized Amplified Regions, Paran and Michelmore, 1993), which can be obtained from the markers mentioned above by cloning and sequencing one of the fragments obtained previously and designing specific primers to amplify the region by PCR, getting in most cases a more specific and robust dominant marker; microsatellites or SSRs (Simple Sequence Repeats, Hamada et al., 1982), which are multi-allelic, hypervariable, and codominant markers based on short tandemly repeated DNA motifs; and SNPs (Single Nucleotide Polymorphisms, Cooper et al., 1985), which consist of single

DNA base differences between homologous genomes. Other types of markers, like ISSRs (Inter-Simple Sequence Repeats, Zietkiewicz et al., 1994), derived from SSRs, that amplifies inter-microsatellite sequences, or SRAPs (Sequence-Related Amplified Polymorphism, Li and Quiros, 2001), consisting of the amplification of open reading frames (ORFs), have had a minority use in melon though have also thrown useful information on the genetic diversity of the species (Parvathaneni et al., 2011; Sestili et al., 2011; Yildiz et al., 2011).

RAPDs, being one of the first genetic markers to be available, were commonly used alone (Mliki et al., 2001) and together with morphological traits (López-Sesé et al., 2003; Luan et al., 2008; Yi et al., 2009) to explore melon germplasm diversity and to assist in curation tasks. RFLPs have been used, together with RAPDs (Silberstein et al., 1999), or RAPDs and AFLPs (Garcia-Mas et al., 2000), to dissect the molecular variation in different accessions within the species. Similarly, the use of different types of markers, like RAPDs and SSRs, has aided in the estimation of the genetic variation of melon germplasm to design strategies for large collection evaluation (López-Sesé et al., 2002). Among the studies on melon diversity using molecular markers recently published, SSRs (Akashi et al., 2001; Chiba et al., 2003; Monforte et al., 2003; Escribano et al., 2012; Guo et al., 2012; Kacar et al., 2012; Serres-Giardi et al., 2012; Hu et al., 2014; Ning et al., 2014; Raghami et al., 2014; Hu et al., 2015) or any of their variations, like EST-SSRs (Kong et al., 2011), are those more commonly chosen, what is a consequence of their great polymorphism and amenability to automation by PCR. In the case of SNPs, the melon genome (Garcia-Mas et al., 2012) and its transcriptome (Blanca et al., 2012) sequencing have contributed hugely to the availability of this kind of marker to the scientific community, turning them into good genome coverage supplier markers. In fact, SNPs have been recently used to dissect the variability present in a broad array of melon accessions, like commercial cultivars, breeding lines, landraces, feral types, and wild relatives, for trait as important from an economical perspective as the sugar accumulation and the climacteric behavior (Leida et al., 2015). SSR and SNP analyses have revealed themselves very useful to compare the genetic variability present in modern melon cultivars and landraces to an ancient extinct landrace (Szabo et al., 2005), making possible to trace the closest relative in the present melon germplasm to that medieval accession. Chloroplastic markers (one SSR and one SNP) have also been used to typify Chinese Hami melons by maternal lineages in comparison with occidental cultivars within group *inodorus* and germplasm coming from other countries in Central and South Asia (Aierken et al., 2011). The genetic relationship elucidated with genetic markers (Tanaka et al., 2007), together with the population structure of the different groups (Esteras et al., 2013), open up new opportunities for a more rational use of the melon germplasm richness in breeding programs.

Table 3. Databases containing information of molecular markers developed in melon

Database name	Website
CmMDb: <i>Cucumis melo</i> Microsatellites Database	http://65.181.125.102/cmmdb2/index.html
Cucurbit Genomics Database	http://www.icugi.org
Melonomics	https://melonomics.net/
VegMarks: a DNA marker database for vegetables	http://vegmarks.nivot.affrc.go.jp

The creation of databases with information about molecular markers developed in melon and other vegetables belonging or not to the Cucurbitaceae family (Table 3) facilitates access to information continuously being expanded.

As commented above, molecular markers, in general, have come to complement data from phenotypic features and, in most cases, both results support each other. Literature is riddled with examples of this synergic relationship between phenotypic and molecular data in studies about melon germplasm diversity as cited above, becoming patently clear however that molecular markers are more reliable and accurate identifying genetic diversity (Parvathaneni et al., 2011). The construction of a molecular database of the collections would assist managing them, for instance making decisions about new prospectings, based on the genetic diversity already collected, or determining if the new acquisitions are redundant.

3.4. Core Collections

Paradoxically, the intense work carried out prospecting, collecting and preserving plant genetic resources has resulted in many germplasm banks collapsing. The use of molecular markers can help to efficiently manage the resources by the creation of core collections. This would alleviate the problems derived from the high number of entries in many of them that is compromising some of the purposes for which they have been created, the conservation and the use of the genetic diversity originally prospected. The advantages in terms of labor and economical investment of preserving, characterizing and using a reduced subset of samples without a considerable sacrifice of the genetic diversity, are undeniable. Genetic markers have revealed themselves as the most appropriate tool for this task and, among them, SSRs seem to be the most useful thanks to their multiallelic nature and the easy methodology required. In melon, two core collections have been built up to now, making use of SSR markers (Hu et al., 2015), or AFLP markers for the actual selection of the core set and SSRs for a subsequent validation (Frary et al., 2013). In both cases, the accessions come mainly from Asia, what has been recently postulated as a center of origin of melon cultivation (Sebastian et al., 2010), which are conserved in the National Mid-term Genebank of Watermelon and Melon (Zhengzhou, China) (Hu et al., 2015), and the National Seed Genebank at the Aegean Agricultural Research Institute (AARI, Menemen, Izmir, Turkey). The core collections represented the 10% (Frary et al., 2013), and 19% (Hu et al., 2015) of the whole set of accessions, both being comprised in the range (5-20%) of most core collections of seed-propagated crops (van Hintum et al., 2000). Regarding the genetic variability present in the core collection respect to the source one, 87% and 100% of the SSR alleles were retained, respectively. Undoubtedly, these core sets will make easier the management of those particular collections and the access of breeders to the desirable alleles contained in them.

There is no panacea for many of the challenges that curators need to face. Even if classic and new techniques can complement each other, certain tools are more appropriate than others for some specific applications, like the examples exposed above in which genetic markers (AFLPs and SSRs) have been used to create core collections. Similarly, the perfect molecular marker does not exist, at least, for the moment. They all have advantages and disadvantages, either of technical, methodological, economic or even of analytical nature. Depending on the particular objective to be addressed, different types of markers can be the most suitable ones.

At the time of choosing the marker to be used, several factors need to be weighed, such as the degree of polymorphism required, and the time, labor and economical cost.

4. USES OF MELON GERMPLASM

Exotic and commercial melon germplasm has been extensively used in phenotypic screenings, for instance, to find resistance to diseases (Thomas and Jourdain, 1992; Boiteux et al., 1995; Pan and More, 1996; Wolff and Miller, 1998; Wolukau et al., 2007) and plagues (Garzo et al., 2004), but also tolerance to abiotic stresses, like dehydration (Pandey et al., 2011; reviewed in Kumar et al., 2012; Mundalia et al., 2015). To exploit all the possibilities that the rich melon germplasm offers to the breeding programs, especially the exotic accessions, it is necessary to characterize it and genotype it with molecular markers. Many scientists have started to make an effort in this sense, including accessions within the subspecies *agrestis* in their diversity studies, as it has been mentioned above (Esteras et al., 2013; Hu et al., 2014; Koffi et al., 2014). Molecular markers developed in melon have also been used extensively in genetic breeding programs to aid in the selection of individuals with desirable characteristics (Marker Assisted Selection, MAS) in an early stage (Kim et al., 2010; reviewed in Oumouloud et al., 2013; Díaz et al., 2014; Perpiñá et al., 2016). For this, obtaining linkage maps in melon and, even more, merging them into consensus ones (Díaz et al., 2011 and references in it contained; Díaz et al., 2015), has been determinant to identify markers linked to the traits of interest. Genetically distant melon sources, for instance, belonging to the two *C. melo* subspecies, *melo* and *agrestis*, have been commonly crossed to obtain breeding and/or mapping populations (Fita et al., 2006; reviewed in Díaz et al., 2011; Yuste-Lisbona et al., 2011; Hwang et al., 2014). Among the traits analyzed in those studies, there are not only resistances to diseases, but also other economically important ones, like those related to the quality (i.e., sugar and organic acid content), size and morphology of the fruits, or plant growing habits, directed to obtain dwarf plants.

A way of exploiting the variability present in the melon germplasm is by the construction of exotic libraries introgressing genomic regions from the wild and exotic accessions (i.e., from the subspecies *agrestis*) onto the background of elite cultivars. Obviously, for these tools to become useful, a genetic map, especially of the introgressed region, is required. The availability of markers amenable to genotype by high-throughput technologies, like SNPs, has sped-up the process, as in the case of the IL (Introgression Line) collection recently obtained in the cultivar Charentais background (Perpiñá et al., 2016). The exotic accession used in this case was PI 161375, named Songwan Charmi, which belongs to the subspecies *agrestis*, and bears many interesting characteristics from a breeding point of view, like resistance to several pests and diseases. The results reported there are very promising as pre-breeding lines for traits with a high economic impact, like high sugar content or delayed climacteric ripening, are being generated.

On the other hand, the use of wild or non-cultivated germplasm in breeding has some clear disadvantages, like the transference of undesirable characteristics to the elite cultivar background by linkage drag, making necessary several generations of backcrossing and selection to get rid of them completely. Again, the combined use of state-of-the-art and classical technologies has assisted scientists to overcome this obstacle. Recently, Sherman et

al., (2013) have employed microarrays together with bulk segregation analysis to finely map the region responsible for the acidity in the melon pulp (*pH* locus). Obviously, this is an irrelevant aspect in sweet melons, in which the pH is very high (around 6.0), though the consumption of many other types of melons is restricted to when they are still young due to the high accumulation of organic acids at ripe stage caused by their allelic composition at *pH* locus. The SNPs linked to the *pH* locus obtained by these authors will make possible to obtain melons with new taste combinations in a more directed way and, hence, investing less time and effort.

5. CHALLENGES AND PROSPECTS

The drastic price reduction in genotyping and next-generation sequencing technologies make feasible the genetic characterization of germplasm collections. In fact, the whole soybean collection (18,480 accessions) maintained at the United States Department of Agriculture (USDA), has already been genotyped with 52,041 SNPs (Song et al., 2013). These data, together with those coming from historical phenotypic characterization, has been exploited to develop prediction models for traits as important from an economic point of view, as protein and oil content, or yield (Jarquin et al., 2016). This approach undoubtedly raises the value of the samples and their data treasured at the germplasm banks. In this case, the collection consisted of domesticated accessions. It would be of great interest to include in these genotyping platforms wild crop relatives as the wild germplasm represents a valuable source of variability with a huge potential in breeding programs. In melon, some of the most important germplasm banks have already tackled this endeavor (Esteras et al., 2013), as it is the case of the Institute for Conservation and Improvement of Valentian Agrodiversity (COMAV, Valencia, Spain).

All germplasm banks, not only those preserving vegetable resources and melon, in particular, will have to face up to the challenge of characterizing genetically their collections in order to maximize their use in breeding strategies assisted by molecular tools, like genetic markers.

REFERENCES

- Aierken, Y., Akashi, Y., Phan, T. P. N., Halidan, Y., Tanaka, K., Long, B., Nishida, H., Long, C. L., Wu, M. Z. & Kato, K. (2011). Molecular analysis of the genetic diversity of Chinese Hami melon and its relationship to the melon germplasm from central and south Asia. *J Jpn Soc Hortic Sci*, 80, 52-65.
- Akashi, Y., Fukuda, N., Wako, T., Masuda, M. & Kato, K. (2002). Genetic variation and phylogenetic relationships in East and South Asian melons, *Cucumis melo* L., based on the analysis of five isozymes. *Euphytica*, 125, 385-396.
- Akashi, Y., Shiomi, S., Kubo, Y., Masuda, M. & Kato, K. (2001). Microsatellite and CAPS markers for ethylene-related genes, 1-aminocyclopropane-1-carboxylic acid (ACC) synthase and ACC oxidase genes, and their variation in melon (*Cucumis melo* L.) *Breeding Sci* 51:107-112.
- BGHZ. *Vegetable Germplasm Bank of Zaragoza*. (2016). <http://sites.cita-aragon.es/BGHZ/>.

- Biodiversity International. (2016). <http://www.biodiversityinternational.org/>.
- Blanca, J., Esteras, C., Ziarsolo, P., Perez, D., Fernandez-Pedrosa, V., Collado, C., Rodriguez de Pablos, R., Ballester, A., Roig, C., Canizares, J. & Pico, B. (2012). Transcriptome sequencing for SNP discovery across *Cucumis melo*. *BMC Genomics*, 13, 280.
- Boiteux, L. S., Reifschneider, F. J. B. & Pessoa, H. B. S. V. (1995). Phenotypic expression of quantitative and qualitative components of partial resistance to powdery mildew (*Sphaerotheca fuliginea* race 1) in melon (*Cucumis melo*) germplasm. *Plant breed*, 114, 185-187.
- Botstein, D., White, R. L., Skolnick, M. & Davis, R. W. (1980). Construction of a genetic linkage map using Restriction Fragment Length Polymorphisms. *Am J Hum Genet*, 32, 314-331.
- CCBAT. Centre for the Conservation of Agricultural Biodiversity in Tenerife. (2016). <http://www.ccbat.es>.
- Chen, J. F., Isshiki, S., Tashiro, Y. & Miyazaki, S. (1997). Biochemical affinities between *Cucumis hystris* Chakr. and two cultivated *Cucumis* species (*C. sativus* L. and *C. melo* L.) based on isozyme analysis. *Euphytica*, 97, 139-141.
- CmMdb: *Cucumis melo* Microsatellites Database. (2014). <http://65.181.125.102/cmmdb2/index.html>.
- Chiba, N., Suwabe, K., Nunome, T. & Hirai, M. (2003). Development of microsatellite markers in melon (*Cucumis melo* L.) and their application to major cucurbit crops. *Breeding Sci*, 53, 21-27.
- COMAV. Institute for Conservation & Improvement of Valentian Agrobiodiversity. (2016). <http://www.comav.upv.es>.
- Cooper, D. N., Smith, B. A., Cooke, H., Niemann, S. & Schmidtke, J. (1985). An estimate of unique sequence heterozygosity in the human genome. *Hum Genet*, 69, 201-205.
- Cuevas, H. E., Staub, J. E., Simon, P. W., Zalapa, J. E. & McCreight, J. D. (2008). Mapping of genetic loci that regulate quantity of beta-carotene in fruit of US Western Shipping melon (*Cucumis melo* L.). *Theor Appl Genet*, 117, 1345-1359.
- Dhillon, N. P. S., Ranjana, R., Singh, Kuldeep, Eduardo, I., Monforte, A. J., Pitrat, M., Dhillon, N. K. & Singh, P. P. (2007). Diversity among landraces of Indian snapmelon (*Cucumis melo* var. *momordica*). *Genet Resour Crop Ev*, 54, 1267-1283.
- Díaz, A., Fergany, M., Formisano, G., Ziarsolo, P., Blanca, J., Fei, Z., Staub, J. E., Zalapa, J. E., Cuevas, H. E., Dace, G., Oliver, M., Boissot, N., Dogimont, C., Pitrat, M., Hofstede, R., van Koert, P., Harel-Beja, R., Tzuri, G., Portnoy, V., Cohen, S., Schaffer, A., Katzir, N., Xu, Y., Zhang, H., Fukino, N., Matsumoto, S., Garcia-Mas, J. & Monforte, A. J. (2011). A consensus linkage map for molecular markers and Quantitative Trait Loci associated with economically important traits in melon (*Cucumis melo* L.). *BMC Plant Biol*, 11, 111.
- Díaz, A., Forment, J., Argyris, J. M., Fukino, N., Tzuri, G., Harel-Beja, R., Katzir, N., Garcia-Mas, J. & Monforte, A. J. (2015). Anchoring the consensus ICuGI genetic map to the melon (*Cucumis melo* L.) genome. *Mol Breeding*, 35, 188.
- Díaz, A., Zarouri, B., Fergany, M., Eduardo, I., Álvarez, J. M., Picó, B. & Monforte A. J. (2014). Mapping and introgression of QTL involved in fruit shape transgressive segregation into 'Piel de Sapo' melon (*Cucumis melo* L.). *Plos One* 9, e104188.
- Escribano, S. & Lazaro, A. (2009). Agro-morphological diversity of Spanish traditional melons (*Cucumis melo* L.) of the Madrid provenance. *Genet Resour Crop Ev*, 56, 481-497.

- Escribano, S., Lazaro, A., Cuevas, H. E., Lopez-Sese, A. I. & Staub, J. E. (2012). Spanish melons (*Cucumis melo* L.) of the Madrid provenance: a unique germplasm reservoir. *Genet Resour Crop Ev*, 59, 359-373.
- Esquinas-Alcazar, J. T. & Gulik, P. J. (1983). *Genetic resources of Cucurbitaceae: a global report*. Rome: IBPGR Secretariat.
- Esteras, C., Formisano, G., Roig, C., Díaz, A., Blanca, J., García-Mas, J., Gómez-Gillamón, M. L., López-Sessé, A., Lázaro, A. I., Monforte, A. & Pico, B. (2013). SNP genotyping in melons: Genetic variation, population structure and linkage disequilibrium. *Theor Appl Genet*, 126, 1285-1303.
- EURISCO. *European Internet Search Catalogue*. (2016). <http://eurisco.ipk-gatersleben.de/>.
- FAO. Food and Agriculture Organization of the United Nations. (2016). <http://www.fao.org/>.
- Fergany, M., Kaur, B., Monforte, A. J., Pitrat, M., Rys, C., Lecoq, H., Dhillon, N. P. S. & Dhaliwal, S. S. (2011). Variation in melon (*Cucumis melo*) landraces adapted to the humid tropics of southern India. *Genet Resour Crop Ev*, 58, 225-243.
- Fita, A., Picó, B., & Nuez, F. (2006). Implications of the genetics of root structure in melon breeding. *J Am Soc Hortic Sci*, 131, 372-379.
- Frery, A., Sigva, H. O., Tan, A., Taskin, T., Inal, A., Mutlu, S., Haytaoglu, M. & Doganlar, S. (2013). Molecular genetic diversity in the Turkish national melon collection and selection of a preliminary core set. *J Am Soc Hortic Sci*, 138, 50-56.
- Garcia-Mas, J., Benjak, A., Sanseverino, W., Bourgeois, M., Mir, G., Gonzalez, V. M., Henaff, E., Camara, F., Cozzuto, L., Lowy, E., Alioto, T., Capella-Gutierrez, S., Blanca, J., Cañizares, J., Ziarolo, P., Gonzalez-Ibeas, D., Rodriguez-Moreno, L., Droege, M., Du, L., Alvarez-Tejado, M., Lorente-Galdos, B., Mele, M., Yang, L. M., Weng, Y. Q., Navarro, A., Marques-Bonet, T., Aranda, M. A., Nuez, F., Pico, B., Gabaldon, T., Roma, G., Guigo, R., Casacuberta, J. M., Arús, P. & Puigdomenech, P. (2012) The genome of melon (*Cucumis melo* L.). *Proc Natl Acad Sci USA*, 109, 11872-11877.
- Garcia-Mas, J., Oliver, M., Gómez-Paniagua, H. & de Vicente, M. C. (2000). Comparing AFLP, RAPD and RFLP markers for measuring genetic diversity in melon. *Theor Appl Genet*, 101, 860-864.
- Garzo, E., Palacios, I., & Fereres, A. (2004). Characterization of melon germplasm resistant to *Aphis gossypii* Glover. In: J. J. Simon, C. A. Dedryver, C. Rispe, & M. Hulle (Eds.), *Aphids in a new millennium* (pp. 441-447). Paris, INRA.
- Genesys. *Gateway to Genetic Resources*. (2015). <https://www.genesys-pgr.org/welcome>.
- Guo, D. L., Zhang, J. P., Xue, Y. M., & Hou, X. G. (2012). Isolation and characterization of 10 SSR markers of *Momordica charantia* (Cucurbitaceae). *Am J bot*, 99, e182-183.
- Hu, J. B., Wang, P. Q., Li, Q. & Su, Y. (2014). Microsatellite analysis of genetic relationships between wild and cultivated melons in Northwest and Central China. *Mol Biol Rep*, 41, 7723-7728.
- Hu, J. B., Wang, P. Q., Su, Y., Wang, R. J., Li, Q. & Sun, K. L. (2015). Microsatellite diversity, population structure, and core collection formation in melon germplasm. *Plant Mol Biol Rep*, 33, 439-447.
- Hamada, H., Petrini, M. G. & Kakunaga, T. (1982). A novel repeated element with Z-DNA-forming potential is widely found in evolutionary diverse eukaryotic genomes. *Proc Natl Acad Sci USA*, 79, 6465-6469.

- Hwang, J., Oh, J., Kim, Z., Staub, J. E., Chung, S. M. & Park, Y. (2014). Fine genetic mapping of a locus controlling short internode length in melon (*Cucumis melo* L.). *Mol Breed*, 34, 949-961.
- ICuGI. *Cucurbit Genomics Database*. (2016). <http://www.icugi.org>.
- IHSM. Institute for Mediterranean and Subtropical Horticulture “La Mayora.” (2016). <http://www.ihsm.uma-csic.es/>.
- IMIDRA. *Local Varieties Bank of the Madrid Research Institute for Rural, Agricultural and Food Development*. (2016). <http://www.madrid.org/imidra>.
- IPGRI. (2003). *Descriptors for Melon (Cucumis melo L.)*. International Plant Genetic Resources Institute, Rome, Italy.
- IRFAP. Institute of Agricultural and Fishing Research and Training of the Balearic Islands. (2016). <http://www.caib.es/govern/organigrama/area.do?lang=es&coduo=1964>.
- Isshiki, S., Ohmi, C., & Okubo, H. (1991). Inheritance of glutamate oxaloacetate transaminase isozymes in kekiri (*Cucumis melo* kekiri group). *J Fac Agr Kyushu U*, 36, 79-82.
- Jarquín, D., Specht, J. & Lorenz, A. 2016. Prospects of genomic prediction in the USDA soybean germplasm collection: historical data creates robust models for enhancing selection of accessions. *G3-Genes Genom Genet*, 6, 2329-2341.
- Jeffrey, C. (1980) A review of the Cucurbitaceae. *Bot J Linn Soc*, 81:233-247.
- Kacar, Y. A., Simsek, O., Solmaz, I., Sari, N. & Mendi, Y. Y. (2012). Genetic diversity among melon accessions (*Cucumis melo*) from Turkey based on SSR markers. *Genet Mol Res*, 11, 4622-4631.
- Kato, K., Akashi, Y., Okamoto, A., Kadota, S. & Masuda, M. (1998). Isozyme polymorphism in melon (*Cucumis melo* L.), and application to seed purity test of F1 cultivars. *Breeding Sci*, 48, 237-242.
- Kim, H., Baek, J., Choi, Y. O., Lee, J. H., Sung, S. K., & Kim, S. (2010). Identification of a cluster of oligonucleotide repeat sequences and its practical implication in melon (*Cucumis melo* L.) breeding. *Euphytica*, 171, 241-249.
- Koffi, K. K., Vroh, B. I., Baudoin, J. P. & Zoro, B. I. A. (2014). Genetic relationships among accessions of African indigenous melons (*Cucumis melo* L. ssp. *agrestis*) using AFLP markers. *Biochem Syst Ecol*, 56, 213-217.
- Kong, Q., Xiang, C., Yang, J. & Yu, Z. (2011) Genetic variations of Chinese melon landraces investigated with EST-SSR markers. *Hortic Environ Biotechnol*, 52, 163-169.
- Kumar, R., Solankey, S. S., & Singh, M. (2012). Breeding for drought tolerance in vegetables. *Vegetable Sci*, 39, 1-15.
- Leida, C., Moser, C., Esteras, C., Sulpice, R., Lunn, J. E., de Langen, F., Monforte, A. J. & Pico, B. (2015). Variability of candidate genes, genetic structure and association with sugar accumulation and climacteric behavior in a broad germplasm collection of melon (*Cucumis melo* L.). *BMC Genet*, 16, 28.
- Li, G. & Quiros, C. F. (2001) Sequence-related amplified polymorphism (SRAP), a new marker system based on a simple PCR reaction: its application to mapping and gene tagging in *Brassica*. *Theor Appl Genet*, 103, 455-461.
- López-Sesé, A. I., Staub, J. E. & Gómez-Guillamón, M. L. (2003). Genetic analysis of Spanish melon (*Cucumis melo* L.) germplasm using a standardized molecular-marker array and geographically diverse reference accessions. *Theor Appl Genet*, 108, 41-52.
- López-Sesé, A. I., Staub, J., Katzir, N. & Gómez-Guillamón, M. L. (2002). Estimation of between and within accession variation in selected Spanish melon germplasm using RAPD

- and SSR markers to assess strategies for large collection evaluation. *Euphytica*, 127, 41-51.
- Luan, F., Delannay, I. & Staub, J. E. (2008). Chinese melon (*Cucumis melo* L.) diversity analyses provide strategies for germplasm curation, genetic improvement, and evidentiary support of domestication patterns. *Euphytica*, 164, 445-461.
- McCreight, J. D., Staub, J. E., López-Sesé, A. & Chung, S. M. (2004). Isozyme variation in Indian and Chinese melon (*Cucumis melo* L.) germplasm collections. *J Am Soc Hortic Sci*, 129, 811-818.
- Melonomics. (2016). <https://melonomics.net/>.
- Mliki, A., Staub, J. E., Sun, Z. Y. & Ghorbel, A. (2001). Genetic diversity in melon (*Cucumis melo* L.): An evaluation of African germplasm. *Genet Resour Crop Ev*, 48, 587-597.
- Monforte, A. J., Garcia-Mas, J. & Arús, P. (2003). Genetic variability in melon based on microsatellite variation. *Plant Breeding*, 122, 153-157.
- Mundalia, L., Joshi, U., Rampuria, S., Meghwal, R., Kumar, S., & Sharma, R. (2015). Effect of water stress on growth and antioxidative defence system in *Cucumis* spp. at seedling stage. *Indian J Agr Sci*, 85, 628-633.
- Ning, X. F., Xiong, L. M., Wang, X. L., Gao, X. W., Zhang, Z. G., Zhong, L. & Li, G. (2014). Genetic diversity among Chinese Hami melon and its relationship with melon germplasm of diverse origins revealed by microsatellite markers. *Biochem Syst Ecol*, 57, 432-438.
- Naudin C. V. (1859) Essais d'une monographie des espèces et des variétés du genre *Cucumis*. *Ann Sci Nat Bot*, 11, 5-87. [Essay of a monograph of the species and varieties of the genus *Cucumis*. *Ann Sci Nat Bot*, 11, 5-87].
- Oliver, M., Garcia-Mas, J., Cardús, M., Pueyo, N., López-Sesé, A., Arroyo, M., Gómez-Paniagua, H., Arús, P. & de Vicente, M. C. 2001. Construction of a reference linkage map for melon. *Genome*, 44, 836-845.
- Oumouloud, M. E., Chikh-Rouhou H., Garcés-Claver A., González-Torres R., Perl-Treves R. & Álvarez J. M. (2013). Breeding melon for resistance to Fusarium wilt: recent developments. *Euphytica*, 192, 155-169.
- Pan, R. S. & More, T. A. (1996). Screening of melon (*Cucumis melo* L.) germplasm for multiple disease resistance. *Euphytica*, 88, 125-128.
- Pandey, S., Ansari, W. A., Jha, A., Bhatt, K. V. & Singh, B. (2013). Evaluation of melons and indigenous *Cucumis* spp. genotypes for drought tolerance. *Acta Hortic*, 979, 335-339.
- Paran, I., & Michelmore, R. M. (1993). Development of reliable PCR-based markers linked to downy mildew resistance genes in lettuce. *Theor Appl Genet*, 85, 958-993.
- Parvathaneni, R. K., Natesan, S., Devaraj, A. A., Muthuraja, R., Venkatachalam, R., Subramani, A. P. & Laxmanan, P. (2011). Fingerprinting in cucumber and melon (*Cucumis* spp.) genotypes using morphological and ISSR markers. *J Crop Sci Biotech*, 14, 1-8.
- Perin, C., Hagen, L. S., De Conto, V., Katzir, N., Danin-Poleg, Y., Portnoy, V., Baudracco-Arnas, S., Chadoeuf, J., Dogimont, C. & Pitrat, M. (2002). A reference map of *Cucumis melo* based on two recombinant inbred line populations. *Theor Appl Genet*, 104, 1017-1034.
- Perpiñá, G., Esteras, C., Gibon, Y., Monforte, A. J. & Picó B. (2016). A new genomic library of melon introgression lines in a cantaloupe genetic background for dissecting desirable agronomical traits. *BMC Plant Biol*, 16, 154.
- Phan, T. P. N., Akashi, Y., Tran, T. M. H., Tanaka, K., Aierken, Y., Yamamoto, T., Nishida, H., Long, C. & Kato, K. (2010). Genetic diversity in Vietnamese melon landraces revealed by

- the analyses of morphological traits and nuclear and cytoplasmic molecular markers. *Breeding Sci*, 60, 255-266.
- Pitrat, M. (2008). Melon (*Cucumis melo* L.). In: J. Prohens, F. Nuez (Eds.), *Handbook of crop breeding*, Vol. I: Vegetables (pp. 283-315). New York, NY: Springer.
- Plant Germplasm Bank “César Gómez-Campo.” 2016. <http://www.bancodegermoplasma.upm.es/>.
- Raghami, M., López-Sesé, A. I., Hasandokht, M. R., Zamani, Z., Moghadam, M. R. F. & Kashi, A. (2014). Genetic diversity among melon accessions from Iran and their relationships with melon germplasm of diverse origins using microsatellite markers. *Plant Syst Evol*, 300, 139-151.
- Roberts, H. F. (1973). Predicting the viability of seeds. *Seed Sci Technol*, 1, 499-514.
- Roy, A., Bal, S. S., Fergany, M., Kaur, S., Singh, H., Malik, A. A., Singh, J., Monforte, A. J., Dhillon, N. P. S. (2012). Wild melon diversity in India (Punjab State). *Genet Resour Crop Ev*, 59, 755-767.
- Sebastian P., Schaefer H., Telford I. R. H. & Renner S. S. (2010). Cucumber (*Cucumis sativus*) and melon (*C. melo*) have numerous wild relatives in Asia and Australia, and the sister species of melon is from Australia. *Proc Natl Acad Sci USA*, 107, 14269-14273.
- Sensoy, S., Buyukalaca, S. & Abak, K. (2007). Evaluation of genetic diversity in Turkish melons (*Cucumis melo* L.) based on phenotypic characters and RAPD markers. *Genet Resour Crop Ev*, 54, 1351-1365.
- Serres-Giardi, L. & Dogimont, C. (2012) How microsatellite diversity helps to understand the domestication history of melon. In: N. Sari, I. Solmaz, & V. Aras (Eds.), *Cucurbitaceae 2012. Proceedings of the Xth EUCARPIA Meeting on Genetics and Breeding of Cucurbitaceae* (pp. 254-263). Antalya, Turkey: Çukurova University.
- Sestili, S., Giardini, A., Ficcadenti, N. (2011). Genetic diversity among Italian melon inodorus (*Cucumis melo* L.) germplasm revealed by ISSR analysis and agronomic traits. *Plant Genetic Resources*, 9, 214-217.
- Sherman, A., Eshed, R., Harel-Beja, R., Tzuri, G., Portnoy, V., Cohen, S., Rubinstein, M., Schaffer, A. A., Burger, J., Katzir, N. & Ophir, R. (2013). Combining bulk segregation analysis and microarrays for mapping of the pH trait in melon. *Theor Appl Genet*, 126, 349-358.
- Silberstein, L., Kovalski, I., Huang, R. G., Anagnostou, K., Jahn, M. M. K. & Perl-Treves, R. (1999). Molecular variation in melon (*Cucumis melo* L.) as revealed by RFLP and RAPD markers. *Scientia Hortic-Amsterdam*, 79, 101-111.
- Soltani, F., Akashi, Y., Kashi, A., Zamani, Z., Mostofi, Y. & Kato, K. (2010). Characterization of Iranian melon landraces of *Cucumis melo* L. groups Flexuosus and Dudaim by analysis of morphological characters and random amplified polymorphic DNA. *Breeding Sci*, 60, 34-45.
- Staub, J. E., Box, J., Meglic, V., Horejsi, T. F. & McCreight, J. D. (1997). Comparison of isozyme and random amplified polymorphic DNA data for determining intraspecific variation in *Cucumis*. *Genet Resour Crop Ev*, 44, 257-269.
- Staub, J. E., López-Sesé, A. I. & Fanourakis, N. (2004). Diversity among melon landraces (*Cucumis melo* L.) from Greece and their genetic relationships with other melon germplasm of diverse origins. *Euphytica*, 136, 151-166.
- Staub, J. E., Meglic, V. & McCreight, J. D. (1998). Inheritance and linkage relationships of melon (*Cucumis melo* L.) isozymes. *J Am Soc Hortic Sci*, 123, 264-272.

- Stepansky, A., Kovalski, I. & Perl-Treves, R. (1999). Intraspecific classification of melons (*Cucumis melo* L.) in view of their phenotypic and molecular variation. *Plant Syst Evol*, 171, 313-332.
- Soltis, D. E., & Soltis, P. S. (1989). *Isozymes in Plant Biology*. Portland, OR: Dioscorides Press.
- Song, Q., Hyten, D. L., Jia, G., Quigley, C. V., E. W. Fickus E. W., Nelson, R. L., & Cregan, P. B. (2013). Development and evaluation of SoySNP50K, a high-density genotyping array for soybean. *PLoS One* 8, e54985.
- Szabo, Z., Gyulai, G., Humphreys, M., Horvath, L., Bittanszky, A., Lagler, R. & Heszky, L. (2005). Genetic variation of melon (*C. melo*) compared to an extinct landrace from the Middle Ages (Hungary) I. rDNA, SSR and SNP analysis of 47 cultivars. *Euphytica*, 146, 87-94.
- Szamosi, C., Solmaz, I., Sari, N. & Barsony, C. (2010). Morphological evaluation and comparison of Hungarian and Turkish melon (*Cucumis melo* L.) germplasm. *Sci Hort-Amsterdam*, 124, 170-182.
- Tanaka, K., Nishitani, A., Akashi, Y., Sakata, Y., Nishida, H., Yoshino, H. & Kato, K. (2007). Molecular characterization of South and East Asian melon, *Cucumis melo* L., and the origin of Group Conomon var. *makuwa* and var. *conomon* revealed by RAPD analysis. *Euphytica*, 153, 233-247.
- Thomas, C. E., & Jourdain, E. L. (1992). Evaluation of melon germplasm for resistance to downy mildew. *Hortscience*, 27, 434-436.
- Trimech, R., Zaouali, Y., Boulila, A., Chabchoub, L., Ghezal, I., Boussaid, M. (2013). Genetic variation in Tunisian melon (*Cucumis melo* L.) germplasm as assessed by morphological traits. *Genet Resour Crop Ev*, 60, 1621-1628.
- UPOV. (2014). Melon, TG/104/5. *Guidelines for the conduct of tests for distinctness, uniformity and stability*. <http://www.upov.int/edocs/tgdocs/en/tg104.pdf>.
- van Hintum, T. J. L., Brown, A. H. D., Spillane, C. & Hodgkin, T. (2000) Core collections of plant genetic resources. *IPGRI Technical Bulletins* No. 3. International Plant Genetic Resources Institute, Rome, Italy.
- VegMarks: a DNA marker database for vegetables. (2016). <http://vegmarks.nivot.affrc.go.jp>.
- Vos, P., Hogers, R., Bleeker, M., Reijans, M., Van de Lee, T., Hornes, M., Frijtens, A., Pot, J., Peleman, J., Kuiper, M., & Zabeau, M. (1995). AFLP: a new technique for DNA fingerprinting. *Nucleic Acids Res*, 23, 4407-4414.
- Welsh, J. & McClelland, M. (1990). Fingerprinting genomes using PCR with arbitrary primers. *Nucleic Acids Res*, 18, 7213-7218.
- Williams, J. G. K., Kubelik, A. R., Livak, K. J., Rafalski, J. A. & Tingey, S. V. (1990). DNA polymorphisms amplified by arbitrary primers are useful as genetic markers. *Nucleic Acids Res*, 18, 6531-6535.
- Wolukau, J. N., Zhou, X. H., Li, Y., Zhang, Y. B., Chen, J. F. (2007). Resistance to gummy stem blight in melon (*Cucumis melo* L.) germplasm and inheritance of resistance from plant introductions 157076, 420145, and 323498. *Hortscience*, 42, 215-221.
- Wolff, D. W. & Miller, M. E. (1998). Tolerance to monosporascus root rot and vine decline in melon (*Cucumis melo* L.) germplasm. *Hortscience*, 33, 287-290.
- Yi, S. S., Akashi, Y., Tanaka, K., Cho, T. T., Khaing, M. T., Yoshino, H., Nishida, H., Yamamoto, T., Win, K. & Kato, K. (2009). Molecular analysis of genetic diversity in melon landraces (*Cucumis melo* L.) from Myanmar and their relationship with melon germplasm from East and South Asia. *Genet Resour Crop Ev*, 56, 1149-1161.

- Yildiz, M., Akgul, N. & Sensoy, S. (2014). Morphological and Molecular Characterization of Turkish Landraces of *Cucumis melo* L. *Notulae Botanicae Horti Agrobotanici Cluj-Napoca*, 42, 51-58.
- Yildiz, M., Ekbic, E., Keles, D., Sensoy, S. & Abak, K. (2011). Use of ISSR, SRAP, and RAPD markers to assess genetic diversity in Turkish melons. *Sci Hortic-Amsterdam*, 130, 349-353.
- Yuste-Lisbona, F. J., Capel, C., Sarria, E., Torreblanca, R., Gómez-Guillamón, M. L., Capel, J., López-Sesé, A. I. (2011). Genetic linkage map of melon (*Cucumis melo* L.) and localization of a major QTL for powdery mildew resistance. *Mol breeding*, 27, 181-192.
- Zietkiewicz, E., Rafalski, A. & Labuda, D. (1994). Genome fingerprinting by simple sequence repeat (SSR)-anchored polymerase chain reaction amplification. *Genomics*, 20, 176-183.

Chapter 39

**PRESERVATION AND CHARACTERIZATION OF
WOODLAND GRAPE (*VITIS VINIFERA* SSP. *SYLVESTRIS*
GMEL.) GENOTYPES OF THE SZIGETKÖZ, HUNGARY**

***Gizella Jahnke^{1,*}, Zóra Annamária Nagy¹, Gábor Koltai²,
Edit Hajdu³ and János Májer¹***

¹National Agricultural Research and Innovation Centre, Research Institute for Viticulture and Enology, Research Station of Badacsony, Badacsonytomaj, Hungary

²University of West Hungary Faculty of Agricultural and Food Sciences,
Mosonmagyaróvár, Hungary

³National Agricultural Research and Innovation Centre, Research Institute for Viticulture and Enology, Research Station of Kecskemét, Kecskemét, Hungary

ABSTRACT

The evolution of cultivated plants played important role in the ascent of humanity. A large number of theories exist about the evolution of the European grapevine (*Vitis vinifera* ssp. *sativa* L.), it is supposed, that woodland grape itself, or crossing with other species could be the progenitor. The woodland grape (*Vitis vinifera* ssp. *sylvestris* GMEL.) in Hungary is a protected species. The quest and preservation of its populations are significant in terms of nature conservation and reserve of biodiversity as well.

In the years of 2010-2015 32 woodland grape genotypes were collected in the Szigetköz, Hungary and ex-situ preserved in the genebank National Agricultural Research and Innovation Centre, Research Institute for Viticulture and Enology, in Badacsonytomaj, Hungary. In 2015-2016 these genotypes were characterised by SSR analysis and were compared with 20 grape rootstocks and 16 *Vitis vinifera* ssp. *sativa* cultivars to ensure the true-to-typeness.

Based on the results dendrogram was constructed. In the dendrogram the *Vitis vinifera* ssp. *sylvestris* accessions form a distinct group, but are closer to the *Vitis vinifera* ssp. *sativa* cultivars, than to the rootstocks. This raises the probability, that these accessions are true-to-type woodland grapes.

* Corresponding Author's Email: gjahnke@mail.iif.hu (H-8261 Badacsonytomaj, Római út 181).

Keywords: woodland grape, *Vitis sylvestris*, SSR, biodiversity

INTRODUCTION

The evolution of cultivated plants played important role in the ascent of humanity. Research of their origin and evolution started at the beginning of the 20th century, but till nowadays a lot of questions remain open.

A large number of theories exist about the evolution of the European grapevine (*Vitis vinifera* ssp. *sativa*). According to De Candolle (1894) the grapes originate from the Trans-Caucasus. Based on the geographical principle in evolution by Darwin (1883) the regions of the primary origin of cultivated plants was created by Vavilov (1932) based on the diversity of the wild relatives of the given species. In this system the European grapevine (*Vitis vinifera* ssp. *sativa*) was classified (together with pistacio and almond) to the Central Asiatic Center.

In the after-glacial Eurasia, the existed woodland grape (*Vitis vinifera* ssp. *sylvestris* GMEL.) spread in whole Europe, and existed even in the southern part of Scandinavia. Man liked its fruits in its natural territory, collected and consumed them. The first *Vitis vinifera* type seeds (long seeds with well-developed “beak”) were found between the excavation finds originating from the 2nd millennium BC. Keeping to west and south, the *Vitis vinifera*-like seed findings turned up gradually from later and later ages. This proves that once the *Vitis vinifera* ssp. *sylvestris* was taken into cultivation from the Trans-Caucasus by the peoples of the ancient Asia. Later the already *Vitis vinifera* was received by the peoples of the antique West-Asia and the people, who lived in the islands of the Aegean-see, spread it on the northern and southern bank of the Mediterranean-see (Kozma, 1991).

According to Terpó (1986) the *Vitis vinifera* ssp. *sativa* is not uniform, but the progeny of more original grape species, the main fundaments between the *Vitis vinifera* ssp. *sylvestris* GMEL. could be the hermaphrodite flowered *V. hissarica* and the *V. nuristanica*. In 1988 he developed a new intraspecific system of *Vitis vinifera* ssp. *sylvestris* GMEL. The substance of his taxonomy was that he sorted the woodland grapes into subspecies based on the hairs of the leaves, and into varietas based on the shape of the leaves. He deduced the eco-geographical groups (convarietas) of *Vitis vinifera* ssp. *sativa* directly or indirectly from these varieties (Jahnke et al., 2014).

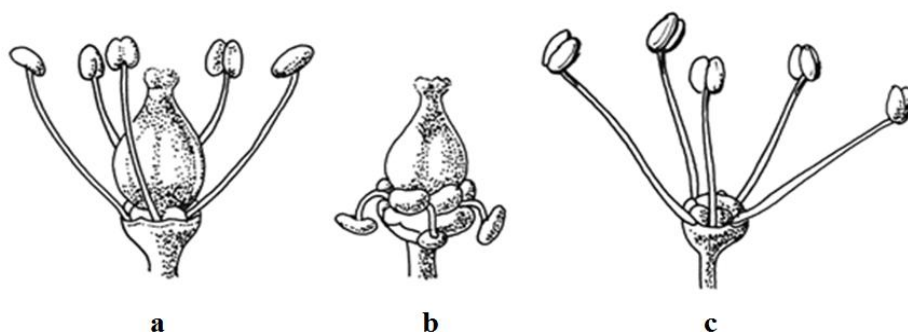


Figure 1. Grapevine flower types: hermaphrodite (a), female (b) and male (c) flower (Bényei et al., 1999).

The ancient cultivars of *Vitis vinifera* ssp. *sativa* were classified in three con-varieties pontica, orientalis and occidentalis by Negrul (1969). In his work he pointed, that the crosses between cultivars of different con-varieties lead to valuable types: those between pontica and orientalis gives some promising wine types; and those of orientalis x occidentalis gives early-ripening types; those of occidentalis x pontica segregates and gives some high-quality wine types.

Accordingly this geographical cultivar groups (con-varieties) of *Vitis vinifera* ssp. *sativa* were not likely to have simultaneously developed, but they formed from the different woodland grape types side by side, or crossing one another respectively, as follows: First the pontican cultivar group (convar. pontica) developed in West- and East-Georgia and in Asia Minor. Its initial form could be the local woodland grape, the *Vitis vinifera* ssp. *sylvestris* GMEL. var. *balcanica*, *Vitis vinifera* ssp. *sylvestris* GMEL. var. *typica* (var. *sylvestris*). Inside the eastern cultivar group (convar. orientalis), the subconvar. *caspiaca* was born in the grape-growing countries of the antique Asia Minor, from the woodland grapes near the places bordering on the Caspian-lake (*Vitis vinifera* ssp. *sylvestris* GMEL. var. *abberans*). The origin of the convar. orientalis subconvar *antasiatica* date from the period later. The hybrids of the pontican cultivars and the local woodland grapes (*Vitis vinifera* ssp. *sylvestris* GMEL.) were the starter forms of the western (convar. occidentalis) cultivars (Kozma, 1991).

The *Vitis vinifera* ssp. *sylvestris* GMEL. in Hungary is a protected species (Farkas, 1999). The quest and reservation of its populations are significant in terms of nature conservation and reserve of biodiversity as well. As pointed before, it is supposed, that this species itself, or crossing with other species could be the progenitor of the European grapevine (*Vitis vinifera* ssp. *sativa*). The ex situ conservation of the quested individuals has a great importance in the practical point of view as well, as they can serve as a resistance source in the future breeding programs.

Seed proteins and enzymes (AcP, ADH, EST, G-6-PDH, MDH, PGM, POD) from several cultivars and wild ecotypes of *Vitis vinifera* ssp. *sativa* have been used to evaluate taxonomic differences between *V. vinifera* ssp. *sativa* and *sylvestris*. Only total proteins in the pH range of 4.0-5.5 and AcP, EST and G-6-PDH were useful for genotype differentiation. The cluster analysis (UPGMA), based on Jaccard genetic distance and determined on the presence/absence of electrophoretic profiles, reveals 2 distinct groups, supporting the hypothesis of the authors that *V. sativa* and *V. sylvestris* should be regarded as 2 separate taxa (Scienza et al., 1994. in Jahnke et al. 2012).

Isozyme polymorphism of mature leaves of woodland grape accessions from south-western Turkey was studied by polyacrylamide gel electrophoresis (PAGE) of Acid phosphatase (ACP), Catechol oxidase (CO), Glutamate oxalacetate transaminase (GOT), Malate dehydrogenase (MDH), and peroxidase (PER) enzymes. GOT was monomorphic with two isozyme bands. Polymorphism was detected in the case of PER, CO, ACP and MDH (Söylemezoglu et al., 2001).

The microsatellite analysis of the grape can be traced back to the early nineties. A lot of SSR loci were identified and characterised in grapes (Thomas and Scott, 1993; Scott et al., 2000; etc.) and were mapped (Adam-Blondon et al., 2004.; Constantini et al., 2007).

The characterisation of grapevine cultivars by microsatellite DNA markers in Europe in the framework of an international cooperation called GENRES 081 (European Network for Grapevine Genetic Resources Conservation and Characterisation) was carried out between 1997 and 2002. In this research programme 6 microsatellite primer pairs were determined and

suggested for the characterisation of cultivars. An European project The GRAPEGEN06 which can be considered as the continuation of the GENRES 081 project started in January 2007. The main objective of GrapeGen06 was to contribute to the successful long term preservation of the *Vitis* genetic resources for the use of future generations.

The microsatellite analyses of *Vitis vinifera* ssp. *sylvestris* accessions from Hungary had begun about 10 years before. The earliest publications suggest the preservation of the genotypes based on morphological and microsatellite (Bodor et al., 2010) and microsatellite analyses (Jahnke et al., 2014).

Based on the relevant publications a conservation of natural genetic diversity of *Vitis vinifera* ssp. *sylvestris* was suggested for the populations of Tunisia (Zoghalmi et al., 2013), Azerbaijan, Georgia (De Lorenzis et al., 2014) the populations of Zagros mountains (Iran) (Doulati-Baneh et al., 2014) and Italy (Biagini et al., 2012.; Biagini et al., 2014).

A new project started in September 2013 in Hungary aimed in the quest and ex situ conservation of *Vitis sylvestris* GMEL. individuals in the area of Szigetköz and Fertő-Hanság National Park as well, as the morphological description and analyses of them by molecular markers. The results of the planned analyses can go a long way into the clarification of the origin of *Vitis vinifera* ssp. *sativa*, and to the explanation of the development of the convarieties of the European grapevine (Jahnke et al, 2014).

MATERIALS AND METHODS

The woodland grape stocks were labeled with plastic stripes in-situ in the Szigetköz, in 2013. All of the stocks were marked on map, were located and the GPS coordinates were saved. Photos were taken in spring and in autumn of 2013 and 2014 about all of the individuals. In June, 2013 young shoots from the individuals were collected and grafted to rootstocks in Badacsony, Hungary.

Seeds were collected from 5 (genetically identical) female flowered stocks in the Szigetköz, in autumn, 2013. The seedlings were planted outdoor in early spring, 2014, and young shoots of 23 seedlings were successfully grafted.

The plant material (dormant canes) of 20 *Vitis* rootstocks, 16 *Vitis vinifera* ssp. *sativa* varieties, 4 wild wines from Gemenc (Hungary) and 32 *Vitis vinifera* ssp. *sylvestris* accessions originated from the collection of the NAIRC RIVE of Badacsony and Kecskemét (Hungary) and from the collection of the University of Pannonia in Cserszegtomaj (Hungary) for rootstocks.

DNA was extracted from the phloem of the dormant canes with DNA Plant Mini Kit (Quiagen), following the manufacturer's instructions. The amount and quality of DNA was determined spectrophotometrically. The DNA was diluted to a concentration of 10 ng/ml.

Microsatellite (SSR) analysis was performed in 8 loci. Three multiplex polymerase chain reactions were carried out in a total volume of 10 µl containing 5 µl of Hot Start Master Mix (Quiagen, Germany), 0.1-0.4 µM of each primer, and 1µl of template DNA. Forward primers were fluorescently labeled with 6-FAM, VIC, NED or PET as reported in Table 2. The following thermal profile was used: (1) 95°C for 30 min (hot start); (2) 95°C for 30 sec, 55°C for 30 sec (decreased by 1°C in each cycle), 72°C for 30 sec per 30 cycles; (4) 95°C for 30 sec, 55°C for 15 sec, 72°C for 30 sec per 30 cycles; (5) 72°C for 7 min.

Table 1. List of the analyzed Vitis accessions

Accession ID	Accession Name	Genetic Origin	Origin of the Accession
V._berl._R1	Resseguier N1	V. berlandieri	Cserszegtomaj, Hungary
V._rup._FW3	Fort Worth N3	V. rupestris	
V._rup._T	Taylor	V. rupestris	
V._cord.	8029 Mtp2	V. cordifolia	
V._rip._GdM	Gloire de Montpellier	V. riparia	
Aramon_rup_G1	Aramon Ganzin N1	V. vinifera x V. rupestris	
V._vip._Ggb	Riparia Grand glabre	V. riparia	
V._rup._FW1	Fort Worth N1	V. rupestris	
Jacquez	Jacquez	V. Bourquina (Vinifera x Aestivalis)	
Vialla	Vialla	V. labrusca x V. riparia	
V._cin._Arnold	Cinerea Arnold	V. cinerea	
V._aest._S.	Sauvage	V. aestivalis	
V._sol.	Solonis	V. solonis	
V._rup._FW2	Fort Worth N2	V. rupestris	
V._berl._R107	Resseguier N107	V. berlandieri	
Aramon_rup_G2	Aramon Ganzin N2	V. vinifera x V. rupestris	
N._Mex.	V. Novo Mexicana	V. riparia x V. candicans	
T5C	Teleki 5C E20	V. berlandieri x V. riparia	Badacsony, Hungary
SO4	Teleki-Fuhr SO4 (133)	V. berlandieri x V. riparia	Cserszegtomaj, Hungary
5BB	Teleki-Kober 5BB	V. berlandieri x V. riparia	
Gemenc_1	Gemenc 1	Vitis sp.	The forest of Gemenc (in-situ conserved)
Gemenc_2	Gemenc 2	Vitis sp.	
Gemenc_3	Gemenc 3	Vitis sp.	
Gemenc_4	Gemenc 4	Vitis sp.	
B.1.	Sylvestris B.1.	Vitis vinifera ssp. sylvestris	Szigetköz, Hungary (ex-situ conserved in Badacsony, Hungary)
B.10	Sylvestris B.10	Vitis vinifera ssp. sylvestris	
B.12	Sylvestris B.12	Vitis vinifera ssp. sylvestris	
B.13	Sylvestris B.13	Vitis vinifera ssp. sylvestris	
B.16	Sylvestris B.16	Vitis vinifera ssp. sylvestris	
B.19	Sylvestris B.19	Vitis vinifera ssp. sylvestris	
B.2	Sylvestris B.2	Vitis vinifera ssp. sylvestris	
B.21	Sylvestris B.21	Vitis vinifera ssp. sylvestris	
B.24	Sylvestris B.24	Vitis vinifera ssp. sylvestris	
B.26	Sylvestris B.26	Vitis vinifera ssp. sylvestris	
B.27	Sylvestris B.27	Vitis vinifera ssp. sylvestris	
B.30	Sylvestris B.30	Vitis vinifera ssp. sylvestris	
B.33	Sylvestris B.33	Vitis vinifera ssp. sylvestris	
B.34	Sylvestris B.34	Vitis vinifera ssp. sylvestris	
B.36	Sylvestris B.36	Vitis vinifera ssp. sylvestris	
B.37	Sylvestris B.37	Vitis vinifera ssp. sylvestris	
B.41	Sylvestris B.41	Vitis vinifera ssp. sylvestris	
B.47	Sylvestris B.47	Vitis vinifera ssp. sylvestris	
B.48	Sylvestris B.48	Vitis vinifera ssp. sylvestris	
B.49	Sylvestris B.49	Vitis vinifera ssp. sylvestris	
B.5	Sylvestris B.5	Vitis vinifera ssp. sylvestris	

Table 1. (Continued)

Accession ID	Accession Name	Genetic Origin	Origin of the Accession
B.50	Sylvestris B.50	Vitis vinifera ssp. sylvestris	Szigetköz, Hungary (ex-situ conserved in Badacsony, Hungary)
B.51	Sylvestris B.51	Vitis vinifera ssp. sylvestris	
B.31	Sylvestris B.31	Vitis vinifera ssp. sylvestris	
S1	Sylvestris S1	Vitis vinifera ssp. sylvestris	
S4_1	Sylvestris S4/1	Vitis vinifera ssp. sylvestris	
S4_2	Sylvestris S4/2	Vitis vinifera ssp. sylvestris	
S4_3	Sylvestris S4/3	Vitis vinifera ssp. sylvestris	
S6_1	Sylvestris S6/1	Vitis vinifera ssp. sylvestris	
S6_2	Sylvestris S6/2	Vitis vinifera ssp. sylvestris	
S6_4	Sylvestris S6/4	Vitis vinifera ssp. sylvestris	
S7	Sylvestris S7	Vitis vinifera ssp. sylvestris	
Szirén	Szirén	Vitis vinifera ssp. sativa	Kecskemét, Hungary
Trilla	Trilla	Vitis vinifera ssp. sativa	
Gesztus	Gesztus	Vitis vinifera ssp. sativa	
Heureka	Heuréka	Vitis vinifera ssp. sativa	
Generosa	Generosa	Vitis vinifera ssp. sativa	
Kecskemét_7	Kecskemét 7	Vitis vinifera ssp. sativa	
Cserszegi _fuszeres	Cserszegi füzeres	Vitis vinifera ssp. sativa	
Irsai_Oliver	Irsai Olivér	Vitis vinifera ssp. sativa	
Kovidinka	Kövidinka	Vitis vinifera ssp. sativa	
Pinot_gis	Pinot gis	Vitis vinifera ssp. sativa	
Ezerjó	Ezerjó	Vitis vinifera ssp. sativa	
Pozsonyi_feher	Pozsonyi fehér	Vitis vinifera ssp. sativa	
Kadarka	Kadarka	Vitis vinifera ssp. sativa	
Muscat_Lunel	Muscat Lunel	Vitis vinifera ssp. sativa	
Muscat_Ottonel	Muscat Ottonel	Vitis vinifera ssp. sativa	
Traminer	Piros tramini	Vitis vinifera ssp. sativa	



Figure 2. Grafting of woodland grape.

Table 2. Primer concentration, label used and multiplex construction for the amplification of 8 loci

	SSR locus	Chromosome*	Conc. (µl)	Labelled forward and reverse primer sequences (5'-3')	Allele size reference range*
Multiplex 1.	VMC6F1	2	4	F: PET – GAG TAA GAG AGA AGC AAG AAA A	121-169
				R: GAG TAA GAG AGA AGC AAG AAA A	
	VVMD27	5	2	F: 6-FAM – GTA CCA GAT CTG AAT ACA TCC GTA AGT	175-223
				R: ACG GGT ATA GAG CAA ACG GTG T	
	VVMD5	16	4	F: VIC – CTA GAG CTA CGC CAA TCC AA	220-268
				R: TAT ACC AAA AAT CAT ATT CCT AAA	
Multiplex 2.	VMC6E1	14	1	F: VIC – CAC TGG CCT GTT GGG AGA TAA T	122-175
				R: CCT TCA ACT GGA AAA GCC TGT C	
	VMC6G1	11	4	F: PET – TGC ATA GTG CTG TAG GCC ATT G	169-197
				R: TCT GTC ATT GCT GTC CCT TTC A	
	VVMD7	7	2	F: 6-FAM – AGA GTT GCG GAG AAC AGG AT	231-267
				R: CGA ACC TTC ACA CGC TTG AT	
Multiplex 3.	VMCNG4b9	6	1	F: NED – CTG GGG AGC ATA TAC ACA TAC CAG	138-188
				R: CTC TCT CTT CCC GAT AGC CAC C	
	VVMD28	3	2	F: PET – AAC AAT TCA ATG AAA AGA GAG AGA GAG A	216-285
				R: TCA TCA ATT TCG TAT CTC TAT TTG CTG	

* according to Migliaro et al. 2013.

One primer of each primer pairs were fluorescently labeled on the 5' end of the DNA chain according to Migliaro et al. (2013). PCR products were run on a PE-Applied Biosystem 3100 Automated Capillary DNA Sequencer, the length of the products were determined using GeneScan 2.0 software (Applied Biosystem).

Estimates of genetic similarity between pairs were calculated by the Jaccard index (Jaccard, 1908). For the generation of distance matrices and UPGMA dendograms a demo version of MolMarker was used (Györfyné Jahnke and Smidla, 2014).

RESULTS

20 stocks of woodland grapes (*Vitis sylvestris* GMEL.) were found in 3 sites, 2 sites in the Szigetköz, and one near Jánossomorja. In Badacsony 8 of the genotypes of the 20 stocks found in-situ were conserved by clonal propagation (grafting to Teleki 5C rootstocks). 24 seedlings were successfully germinated and grafted to Teleki 5C rootstocks.

Table 3. Summary statistics based on SSR results

Marker	Major Allele Frequency	Genotype No	Allele No	Gene Diversity	Heterozygosity	PIC	f
VMC6F1	0,3429	29	17	0,8261	0,5143	0,8112	0,3836
VVMD27	0,384	33	18	0,8139	0,6667	0,8015	0,1880
VVMD5	0,2676	36	19	0,8827	0,7042	0,8743	0,2090
VMC6E1	0,2676	31	19	0,8835	0,9437	0,8753	-0,0611
VMC6G1	0,5580	18	13	0,6400	0,2609	0,6088	0,5971
VVMD7	0,4167	34	18	0,8024	0,6667	0,7935	0,1759
VMCNG4b9	0,1449	45	19	0,9261	0,9275	0,9214	0,0057
VVMD28	0,1643	41	24	0,9119	0,7429	0,9057	0,1923
Mean	0,3182	33,375	18,375	0,8358	0,6783	0,8240	0,1953



Figure 3. Woodland grape seedlings in vivo.



Figure 4. Flowers of male flowered *Vitis sylvestris*.

Table 4. Results of microsatellite (SSR) analysis in 8 loci. – =no amplification/null allele

Accession ID	VMC6F1		VVMD27		VVMD5		VMC6E1		VMC6G1		VVMD7		VMCNG4b9		VVMD28	
V_berl_R1	175	175	193	213	228	228	124	132	-	-	231	233	176	176	232	250
V_rup_FW3	151	151	187	199	254	262	134	136	181	183	253	259	156	164	232	240
V_rup_T	149	149	199	211	252	254	134	136	181	183	251	259	156	164	232	240
V_cord.	151	151	189	195	238	238	122	128	185	187	237	241	168	182	238	238
V_rip_GdM	125	125	203	209	266	272	134	160	179	179	231	265	156	164	240	264
Aramon_rup_G1	133	139	179	203	254	254	134	168	181	181	243	249	162	162	236	252
V_vip_Ggb	149	165	195	195	266	266	124	134	181	181	239	253	154	164	236	252
V_rup_FW1	125	161	187	211	260	262	134	136	181	181	251	253	-	-	240	240
Jacquez	175	175	177	187	230	246	142	142	179	191	237	239	158	176	234	240
Vialla	149	149	185	207	266	272	130	134	181	181	235	251	154	156	230	240
V_cin_Arnold	149	149	183	207	232	240	128	130	183	189	231	231	150	158	264	286
V_aest_S.	161	161	181	195	244	246	132	134	191	191	237	247	156	158	250	250
V_sol.	125	165	193	207	254	268	128	136	181	181	253	259	154	164	252	254
V_rup_FW2	149	149	199	211	254	262	134	136	181	183	253	259	156	164	240	250
V_berl_R107	175	175	189	189	238	238	122	130	181	183	231	231	174	180	246	246
Aramon_rup_G2	133	153	203	203	254	254	134	168	181	191	239	249	162	162	254	264
N_Mex.	161	165	183	205	264	266	140	154	181	181	251	253	164	166	250	250
T5C	145	175	201	209	236	266	130	160	183	183	231	265	156	166	240	262
SO4	145	175	201	209	236	266	130	160	183	183	233	265	156	166	240	262
5BB	145	145	189	209	236	266	130	160	181	181	233	265	156	170	220	256
Gemenc_1	149	149	213	215	264	264	146	154	181	181	241	253	164	180	232	232
Gemenc_2	159	159	185	187	-	-	154	160	179	179	233	239	-	-	-	-
Gemenc_3	123	123	187	187	254	254	154	162	161	181	239	239	168	180	240	274
Gemenc_4	139	139	211	215	246	264	152	154	181	181	241	253	154	164	252	254
B.1.	129	133	187	189	234	234	152	154	173	173	239	239	180	186	236	236
B.10	133	137	187	189	232	234	152	154	173	173	239	263	180	186	218	230
B.12	129	133	187	201	232	234	150	152	173	173	239	239	180	186	218	236
B.13	133	137	187	187	230	234	162	164	173	173	239	261	168	172	238	266
B.16	129	133	189	199	234	266	150	152	173	173	239	263	184	186	230	236
B.19	129	133	187	201	232	234	154	160	173	173	239	263	164	182	218	236
B.2	129	131	187	187	234	266	154	160	173	173	239	265	182	186	218	236
B.21	133	133	187	187	230	234	152	154	173	173	239	265	164	180	230	230
B.24	121	129	189	189	230	234	154	154	173	173	239	239	170	186	236	266
B.26	121	133	187	187	234	266	154	160	173	173	239	263	180	186	218	236
B.27	129	133	187	187	234	266	152	154	173	173	239	239	184	186	230	266
B.30	129	133	187	189	232	234	152	154	173	173	239	239	164	168	218	230
B.33	129	133	187	189	234	234	152	154	173	173	239	263	182	186	230	236
B.34	133	133	189	201	234	234	152	160	173	173	239	239	184	186	218	236
B.36	133	133	189	201	234	234	154	160	173	173	239	263	184	186	230	236
B.37	129	133	187	189	238	238	150	152	173	173	239	239	164	180	230	230
B.41	133	165	187	187	234	234	152	160	173	173	239	239	164	180	230	230
B.47	129	133	187	189	232	234	140	154	173	173	239	239	178	180	236	236
B.48	129	133	187	189	234	266	140	154	173	173	239	239	182	182	230	266
B.49	129	133	187	199	234	234	150	152	173	173	239	239	164	180	230	236
B.5	129	133	187	189	230	234	152	154	173	173	239	239	168	180	236	266
B.50	133	161	189	199	228	230	152	154	173	173	239	239	164	180	230	230
B.51	129	133	185	187	232	234	154	160	173	173	239	239	184	186	230	230
B31	133	159	187	187	240	240	152	160	173	173	239	239	164	186	230	230
S1	121	123	201	203	236	238	134	136	179	179	257	261	158	162	222	244
S4/1	133	161	187	187	234	234	152	154	173	173	239	239	166	170	238	238
S4/2	133	133	187	187	234	266	154	162	173	173	241	263	170	182	238	238
S4/3	129	129	189	189	224	234	154	154	173	173	239	239	170	186	236	266

Table 4. (Continued)

Accession ID	VMC6F1		VVMD27		VVMD5		VMC6E1		VMC6G1		VVMD7		VMCNG4b9		VVMD28	
S6/1	133	133	187	187	232	234	162	164	173	173	239	263	168	172	238	266
S6/2	129	133	187	187	230	234	154	162	173	173	239	239	170	172	238	266
S6/4	129	133	187	187	230	234	154	154	173	173	241	263	166	168	238	266
S7	133	133	187	187	234	254	142	154	173	173	239	257	162	170	238	238
Szirén	129	129	177	185	236	238	142	154	173	173	243	249	140	160	220	268
Trilla	121	137	177	179	274	274	144	154	173	191	233	255	168	174	266	278
Gesztus	129	129	179	187	274	274	144	168	173	173	239	255	164	174	220	266
Heuréka	133	137	187	187	254	274	144	164	-	-	241	261	170	174	222	236
Generosa	133	137	183	187	234	236	154	164	173	197	241	257	160	160	238	278
Kecskemét_7	139	139	-	-	224	224	154	168	173	173	239	243	160	174	236	236
Cserszegi_fuszeres	133	133	177	187	228	238	166	168	181	199	251	251	154	160	220	230
Irsai_Olivér	-	-	177	179	224	240	154	164	173	173	247	249	-	-	-	-
Kovidinka	133	133	179	179	236	236	166	168	173	191	239	249	154	174	236	250
Pinot_gis	129	129	183	187	230	254	154	168	173	181	239	243	160	164	220	238
Ezerjo	133	133	177	183	228	234	154	168	173	195	239	239	150	156	222	248
Pozsonyi_feher	133	133	179	179	228	236	146	154	173	197	249	255	140	174	268	278
Kadarka	-	-	-	-	224	238	142	168	191	197	247	255	176	178	230	262
Muscat_Lunel	169	169	-	-	224	238	142	146	-	-	233	249	160	168	248	268
Muscat_Ottonel	129	129	177	187	224	238	142	146	173	181	239	243	160	164	260	268
Piros_tramini	133	133	187	187	234	236	-	-	177	177	223	223	160	170	222	238

The mean number of alleles per locus and mean genotype number in rootstocks were 12.875 and 13.875; in woodland grapes 7.750 and 9.625 and in cultivated grapevines 8.500 and 10.875 respectively. The H (Heterosigosity) and PIC (Polimorphic Information Content) values in rootstocks were 0.7266 and 0.8440; in woodland grapes 0.625 and 0.5682; while among cultivated grapevines it was 0.7473 and 0.7781 respectively.

The detailed results of SSR analyses are presented in Table 4.

Based on the results UPGMA dendrogram was constructed, which is presented in Figure 5.

DISCUSSION

The summary statistics of loci in rootstocks, woodland grapes and cultivated grapevines shows that the woodland grape accessions has the lowest genetic diversity, followed by the cultivated varieties. The highest genetic diversity was found in rootstocks.

Similarly low genetic diversity was observed with 7.56 alleles per locus, and 0.464 heterosigosity in the woodland grapes of the Geisenheim collection in Germany (Bitz et al., 2015). Low genetic diversity was observed in Austria as well (Regner et al. 2004).

Contradictorily quite high heterosigosity (0.67) was observed In Iranian wild woodland grape (*Vitis vinifera* ssp. *sylvestris*) populations. The study of the genetic diversity of woodland grapes and cultivated varieties of Azerbaijan and Georgia an east-to-west distribution from Caucasus to Europe confirming the suggestion of Myles et al. (2011) and Imazio et al. (2013).

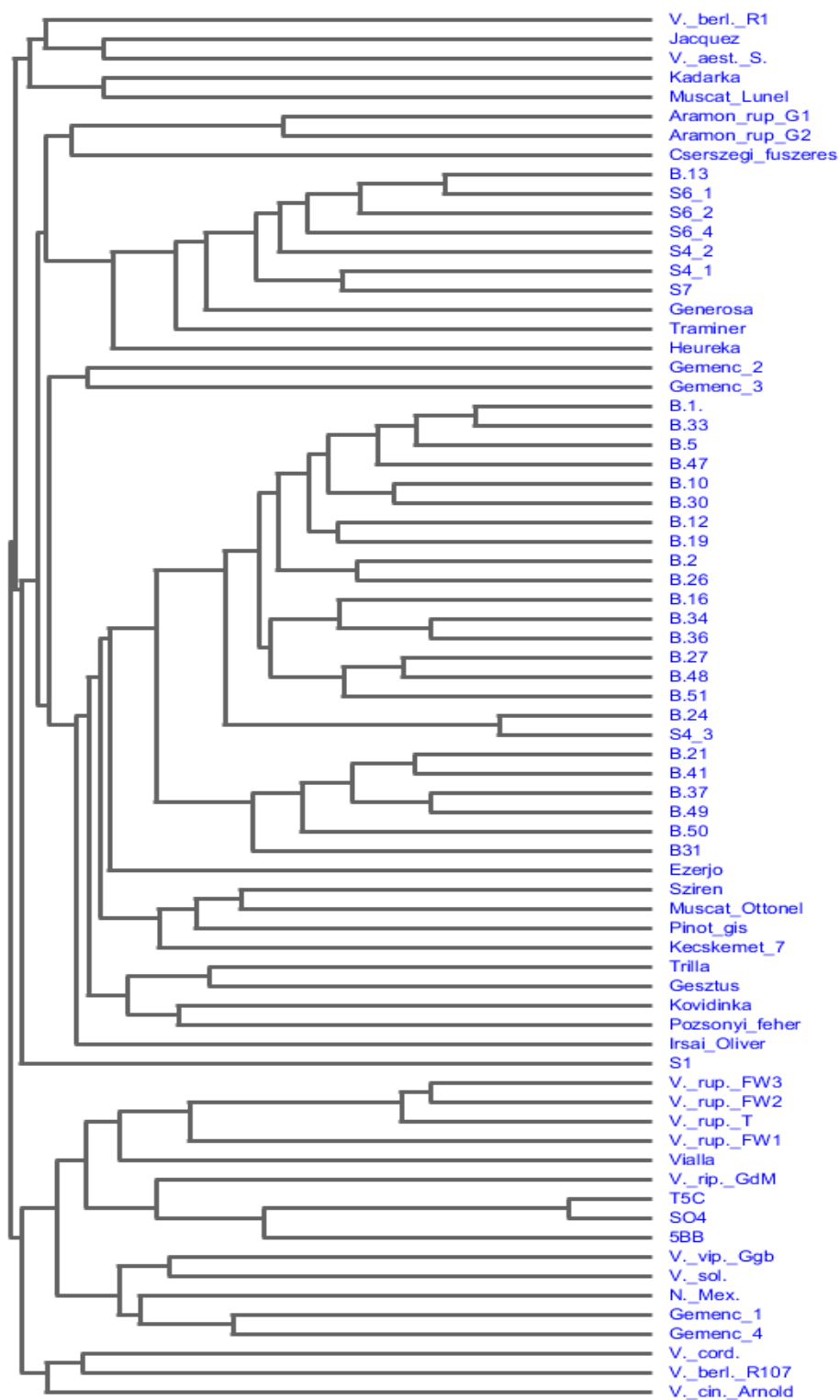


Figure 5. UPGMA dendrogram of the accessions based on SSR results.

In this study, the low genetic diversity of *Vitis vinifera* ssp. *sylvestris* shows that the population of the Szigetköz (Hungary) may originated from the Western-European populations.

Based on the SSR results presented in Table 4, UPGMA dendrogram was constructed. The main groups (*V. vinifera* ssp. *sylvestris*, *V. vinifera* ssp. *sativa*, and rootstocks) mainly form 5 distinct groups. *Vitis vinifera* ssp. *sylvestris* GMEL. accessions form two distinct group in the dendrogram. The rootstocks Aramon Ganzin N1 and (N2 (*V. vinifera* x *V. rupestris*), shows similarity to *Vitis vinifera* ssp. *sativa* variety ‘Cserszegi fűszeres’, which is not surprising taking into account the hybrid origin of this accessions.

Most of the *V. vinifera* ssp. *sativa* varieties forms a large group with most of the *V. vinifera* ssp. *sylvestris* genotypes, but the two subspecies (*sativa* and *sylvestris*) forms distinct subgroups. Most of the rootstock accessions form an distinct group.

The *Vitis vinifera* ssp. *sylvestris* genotypes form distinct groups and show similarity with *Vitis vinifera* ssp. *sativa* genotypes, which clearly shows their true-to-typeness.

The 4 in-situ conserved genotypes of the Gemenc forest (Hungary) fall into 2 groups. Gemenc 2 and Gemenc 3 shows similarity to both *sativa* and *sylvestris* groups, supposing their *Vitis vinifera* ssp. *sylvestris* or hybrid (*sativa* x *sylvestris*) origin.

The other group of Gemenc 1 and Gemenc 4 is in the rootstock group, which probably shows their hybrid (*V. sylvestris* x *V. ripara*) origin.

To clarify these questions further genetic, molecular and morphological analyses are planned. Our results support – as suggested by Bodor et al. (2010) - the further quest, ex-situ and in-situ preservation and analyses of the *Vitis vinifera* ssp. *sylvestris* germplasm in Hungary.

ACKNOWLEDGMENT

This research was funded by the Hungarian Scientific Research Fund (project no. PD-109386).

REFERENCES

- Adam-Blodon, A. F., Roux, C., Claux, D., Butterlin, G., Merdinoglu, D., This, P., 2004. Mapping 245 SSR markers on the *Vitis vinifera* genome: a tool for grape genetics. *Theor. Appl. Genet.* 109, 1017-1027.
- Bényei, F., Lőrincz, A., Sz. Nagy, L. (1999). Szőlőtermesztés. (Grape growing – in Hungarian) *Mezőgazda Kiadó. Budapest.*
- Biagini, B., De Lorenzis, G., Imazio, S., Failla, O., Scienza, A. 2014. Italian wild grapevine (*Vitis vinifera* L. subsp. *sylvestris*) population: insights into eco-geographical aspects and genetic structure. *Tree Genetics & Genomes* 10, 1369-1385.
- Biagini, B., De Lorenzis, G., Scienza, A., Failla, O., Imazio, S. Maghradze, D. 2012. Wild Grapevine (*Vitis vinifera* L. Subsp. *sylvestris* (Gmelin) Hegi) in Italy: Distribution and Preliminary Genetic Analysis. *Acta Hort.* 948, 211-216.
- Bitz, L., Zinelabidine, L. H., Rühl, E. H., & Bitz, O. (2015). Microsatellite analysis of traditional eastern grapevine varieties and wild accessions from Geisenheim collection in Germany. *Vitis* 54, 17-21.

- Bodor, P., Höhn, M., Pedryc, A., Deák, T., Dücső, I., Uzun, I., Cseke, K., Böhm É. I., Bisztray, G. D. 2010. conservation value of the native Hungarian wild grape (*Vitis sylvestris* Gmel.) evaluated by microsatellite markers. *Vitis* 49, 23-27.
- Costantini, L., Grando, M. S., Feingold, S., Ulanovsky, S., Mejia, N., Hinrichsen, P., Doligez, A., This, P., Cabezas, J. A., Martinez-Zapater, J. M. 2007. Generation of a Common Set of Mapping Markers to Assist Table Grape Breeding. *Am. J. Enol. Vitic.* 58, 102-111.
- Darwin, C. 1883. *The Variation of Animals and Plants Under Domestication* New York: D. Appleton & Co.
- De Candolle A. 1894. Termesztett növényeink eredete. (The origin of our cultivated plants-in Hungarian) *Királyi Magyar Természettudományi Társulat*, Budapest 201- 204.
- De Lorenzis, G., Simone, Di Lorenzo, G. B., Failla, O., Musayev, M. K., Salimov, V., Maghradze, D., Chipashvili, R. 2015. Study of Genetic Diversity in *V. vinifera* subsp. *sylvestris* in Azerbaijan and Georgia and Relationship with Species of the Cultivated Compartment. *Acta Hort.* 1074, 49-53.
- Doulati-Baneh, H., Mohammadi, S. A., Labra, M., De Mattia, F., Bruni, I., Mezzasalma, V., Abdollahi, R. 2015. Genetic characterization of some wild grape populations (*Vitis vinifera* subsp. *sylvestris*) of Zagros mountains (Iran) to indentify a conservation strategy. *Plant Genet. Res.*, 13, 27-35.
- Farkas, S., 1999. Magyarország védett növényei. (Protected plants of Hungary. – in Hungarian) Mezőgazda Kiadó, Budapest 166-167.
- Győrffyné Jahnke, G. Smidla, J. 2014. MolMarker- Molekuláris markerekkel kapott kutatási eredmények értékelését segítő felhasználóbarát szoftver. (MolMarker - User-friendly software to help evaluate the research results obtained by molecular markers) *Proc. 56th Georgikon Sci. Conf.* 135-143.
- Imazio, S., Maghradze, D., De Lorenzis, G., Bacilieri, R., Laucou, V., This, P., Scienza, A., Failla, O. 2013. From the cradle of grapevine domestication: molecular overview and description of Georgian grapevine (*Vitis vinifera* L.) germplasm. *Tree Gen. Gen.*, 9, 641-658.
- Jaccard, P. 1908. *Nouvelles recherches sur la distribution florale*. [New research on floral distribution].
- Jahnke, G., Májer, J, Remete, J. 2012. Gel electrophoresis of grapevine (*Vitis vinifera* ssp. *sativa*) isozymes - a review. In: Magdelin S (ed.) *Gel Electrophoresis - Advanced Techniques*. InTech Open Access Publisher, Rijeka 67-82.
- Jahnke, G., Nagy, Z. A., Koltai, G., Májer, J. 2014. SSR Analysis of some *Vitis vinifera* ssp. *sylvestris* (Gmel.) Accessions of the Szigetköz and Fertő-Hanság National Park, Hungary In: Bálo B., Majer P., Váradi Gy. (ed.): *Xth International Terroir Congress 2014, Proceedings*: 267-271.
- Kozma P. 1991. A szőlő és termesztése I. (The grapevine and its growing I. – in Hungarian) *Akadémiai Kiadó*, Budapest.
- Migliaro, D., Morreale, G., Gardiman, M., Landolfo, S., Crespan, M. 2013. Direct multiplex PCR for grapevine genotyping and varietal identification. *Plant Genetic Resources*, 11(02), 182-185.
- Myles, S., Boyko, A. R., Owens, C. L., Brown, P. J., Grassi, F., Aradhya, M. K., Prins, B., Reynolds, A., Chia, J.M., Ware, D., Bustamante, C.D., Buckler, E.S. 2011. Genetic structure and domestication history of the grape. *Proc. Nat. Acad. Sci.*, 108, 3530-3535.

- Negrul, A. M., 1969. Questions of the origin and breeding of vines on a genetical basis. *NI Vavilov and agricultural science.*, 323-339.
- Regner, F., Hack, R., Gangl, H., Leitner, G., Mandl, K., Tiefenbrunner, W. 2015. Genetic variability and incidence of systemic diseases in wild vines (*Vitis vinifera* ssp. *silvestris*) along the Danube. *Vitis*, 43, 123-130.
- Scienza, A., Villa, P., Tedesco, G., Parini, L., Ettort, C., Magenes, S., Gianazza, E. 1994. A chemotaxonomic investigation on *Vitis vinifera* L. II. Comparison among ssp. *sativa* traditional cultivars and wild biotypes of ssp. *silvestris* from various Italian regions. *Vitis* 33, 217-224.
- Scott, K. D., Lee, L. S., Dow, T., Henry, R. J. 2000. Isolation and characterisation of new grape microsatellites. *Acta Hort.* 528, 199-200.
- Söylemezoglu, G., Agaoglu, Y. S., Uzun H. I. 2001. Ampelographic Characteristics and Isozymic Analyses of *Vitis Vinifera* Spp. *Silvestris* Gmel. in Southwestern Turkey. *Biotechnol. and Biotechnol. Eq.* 15, 106-113.
- Terpó A. 1986. A kultúrfajok eredete. (The origin of cultivated specieses – in Hungarian) In: Terpó A (ed.): Növényrendszertan az ökonómbotanika alapjaival I. (Plant systematics with the basics of the econobotany I. – in Hungarian) *Mezőgazdasági Kiadó*, Budapest 108-109.
- Thomas, M. R., Scott, N. S. 1993. Microsatellite repeats in grapevine reveal DNA polymorphism when analysed as sequence tagged sites (STSs). *Theor. Appl. Genet.* 86, 895-990.
- Vavilov, N. I., 1932. The Process of Evolution in Cultivated Plants. *Proceedings of the Sixth International Congress of Genetics, Ithaca, New York, USA* 331-342.
- Zoghalmi, N., Riahi, L., Laucou, V., Mliki, A., Ghorbel, A., This, P. 2013. Genetic structure of endangered wild grapevine *Vitis vinifera* ssp. *silvestris* populations from Tunisia: Implications for conservation and management. *For. Ecol. Manage.* 310, 896-902.

Chapter 40

A MINI REVIEW ON MORPHOLOGICAL AND GENETIC DIVERSITY OF SWEET (*PRUNUS AVIUM* L.) AND SOUR CHERRY (*PRUNUS CERASUS* L.) CULTIVARS

***Ioannis Ganopoulos^{1,2}, Panagiotis Madesis¹,
Filippos A. Aravanopoulos², Athanasios Tsaftaris^{1,3},
Thomas Sotiropoulos⁴ and Konstantinos Kazantzis^{4,*}***

¹Institute of Applied Biosciences, CERTH, Themi, Thessaloniki, Greece

²Forest Genetics & Tree Breeding, Faculty of Forest &
Environmental Science, Aristotle University of Thessaloniki, Thessaloniki, Greece

³Department of Genetics and Plant Breeding, School of Agriculture, Aristotle University
of Thessaloniki, Thessaloniki, Greece

⁴Institute of Plant Breeding and Phytogenetic Resources,
Department of Deciduous Fruit Tree Growing in Naoussa
(Hellenic Agricultural Organisation – ‘DEMETER’), Naoussa, Greece

ABSTRACT

Two major cherry species are grown for their fruit, the diploid sweet cherry and the tetraploid sour cherry. Cherries are characterized by high genetic diversity mainly due to their self-incompatibility propagation system. Estimation of the species phenological and genetic diversity has been performed using a number of different traits and marker systems including morphological and anatomical characteristics, as well as isoenzyme and molecular markers. Different molecular markers have been used, spanning from restriction fragment length polymorphisms (RFLPs) to single nucleotide polymorphisms (SNPs), simple sequence repeat (SSR) markers being the most frequently. Moreover, molecular markers have also been used to mark and trace specific agronomic traits, such as self-(in)compatibility (S-alleles) or fruit weight thus developing functional markers. The markers reviewed herein will be useful not only for monitoring the genetic diversity in cherry breeding programs, but also for gene conservation, while these or other markers may permit marker-assisted selection for favorable agronomic traits.

Keywords: sweet cherry, sour cherry, genetic diversity, molecular markers, S-allele, morphological markers

INTRODUCTION

Sweet (*Prunus avium*) and sour cherry (*P. cerasus*), (Rosaceae), are highly valued for their excellent quality fruits and wood. Sour cherry is a tetraploid species ($2n=4x=32$) originated through natural hybridization between sweet cherry ($2n=2x=16$) and the wild tetraploid ground cherry (*P. fruticosa*) (Rakonjac et al., 2010).

The cherry world production faced an increase in terms of production in many new and traditional regions. Sweet cherries are one of the few remaining seasonal fruit crops as it cannot be stored for a long time, and in many markets no other item creates as much seasonal in-store activity as fresh cherries (Kappel et al., 2012).

A ‘genetic marker’ is a DNA sequence associated with a gene which has a clear effect in the phenotype, but a genetic marker can also be associated with a non-coding part of the genome (which does not contribute to the phenotype). Genetic diversity studies in cherries have been performing by various biochemical (e.g., isoenzyme) and DNA-based (molecular) markers. Different molecular markers have been applied in cherries, such as restriction fragment length polymorphism (RFLP) and several PCR-based techniques. Among these techniques, randomly amplified polymorphic DNA (RAPD), amplified fragment length polymorphism (AFLP), microsatellites or simple-sequence repeats (SSR), inter-simple sequence repeats (ISSR), single nucleotide polymorphisms (SNP) and functional markers (detecting the molecular changes behind a specific trait) have been used. The aim of this review is to present an overview of the different techniques used and results obtained for the assessment of cherry genetic diversity, and to assess their applicability in breeding. The implications of novel genomic technologies such as next generation sequencing (NGS) and Genotyping by Sequence (GBS) are also discussed.

MORPHOLOGICAL TRAITS

Morphological traits have been traditionally used for species description and taxonomical identification. Morphological traits are the result of genetic x environment interaction and usually are controlled by more than one gene, presenting a continuous variability that is correlated with the number of genes involved (Ayala, 1982, El-Esawi et al., 2012).

Many studies indicated that fruit and leaf traits are crucial factors in phenotyping and morphologically characterizing the diversity in sweet cherry breeding materials (Antonius et al., 2012, Ganopoulos et al., 2011b, Hjalmarsson and Ortiz, 2000, Lacis et al., 2009a, Rakonjac et al., 2010). Principal component analysis (PCA) has been employed to determine the relationships among cultivars, to study correlations among tree traits and to evaluate sweet cherry (Beyer et al., 2002, Christensen, 1974, Ganopoulos et al., 2011b, Hjalmarsson and Ortiz, 2000, Khadivi-Khub, 2014, Lacis et al., 2009a, Petruccioli et al., 2013, Rodrigues et al., 2008, Sánchez et al., 2008), sour cherry (Hillig and Iezzoni, 1988, Iezzoni and Pritts, 1991, Krahel et al., 1991, Rakonjac et al., 2010) and *Cerasus* (Khadivi-Khub et al., 2012, Shahi-Gharahlar et

al., 2010). Recently, Ganopoulos et al. (2016) estimated the diversity of morpho-physiological traits in a worldwide collection of sweet cherry cultivars (n=146) in Greece using multivariate analysis. They found that the characterized sweet cherry collection had a high potential for specific breeding goals and suggested that a multivariate statistics approach taking into account both the genetic and the phenotypic data could be of the utmost importance for the analysis of the diversity of GeneBank collections. Furthermore, the same group examined the morpho-physiological diversity in a diverse collection of sour cherry cultivars using multivariate analysis and revealed significant diversity regarding the morpho-physiological traits used.

ISOENZYME ANALYSIS

Isoenzymes are different forms of an enzyme, their difference being in their amino acid sequences, yet they catalyse the same biochemical reaction. Amino acid changes in the different forms of an enzyme cause differences in their polarity and structure, which result in changes of their electrophoretic mobility which forms the basis of their separation and visualization. Isoenzymes may be encoded by genes at different loci or different alleles at the same locus (known as alleloenzymes or allozymes). The detection of isoenzymes was the first molecular marker strategy to be used, because of their co-dominant gene expression and very good reproducibility (Santi et al., 1990, Arulsekhar and Parfitt, 1986, Cerezo and Arus, 1989).

(Granger et al., 1993) used 10 isozymes which allowed the identification of 76 sweet cherry varieties despite the fact that some of them which shared the same morphological characters, were previously considered as the same variety. The following enzyme systems have been applied to study polymorphism in 36 sweet, sour and ground cherry cultivars: isocitrate dehydrogenase (IDH), phosphoglucosomerase (PGI), phosphoglucosomutase (PGM), shikimate dehydrogenase (SKDH), 6-phosphoglucosomate dehydrogenase (6-PGD), leucine aminopeptidase (LAP) and malate dehydrogenase (MDH) (Beaver et al., 1995). Beaver et al. (1995), suggested that sour cherry is more polymorphic than sweet cherry, which might be attributed to its allotetraploid nature. Isoenzymes have also been used by (Santi et al., 1990) and (Frascaria et al., 1993) for the identification of *P. cerasus* and *P. avium*. Furthermore, (Corts et al., 2008) applied five isozymes in a collection of 17 sweet and sour cherry cultivars, however their study showed low isoenzymatic polymorphism.

GENETIC DIVERSITY STUDIES WITH MOLECULAR MARKERS

Sweet Cherry

Different molecular markers have been used in cherries genetic diversity studies. (Stockinger et al., 1996) and (Gerlach and Stösser, 1997) have applied RAPD markers and deemed them suitable for cultivar identification in sweet cherry. However, the method of choice in the last 15 years regarding studies on cherries genetic diversity is SSR markers (Cantini et al., 2001, Dirlwanger et al., 2002, Wünsch and Hormaza, 2002, Schueler et al., 2003, Bianchi et al., 2004, Vaughan and Russell, 2004, Aka Kaçar et al., 2005, Ohta et al., 2005, Hölten and Gregorius, 2006, Pedersen, 2006), due to their stability, co-dominance and multi-allelism.

Peach (*Prunus persica*) was the donor of the most commonly used SSR primers in *Prunus* sp. (Dirlewanger et al., 2002), sweet cherry (Sosinski et al., 2000, Dirlewanger et al., 2002) and sour cherry (Lacis et al., 2009). Moreover, several SSRs have been developed later in order to determine genetic relationships in all *Prunus* species (Lacis et al., 2009b, Antonius et al., 2012). Due to the fact that SSRs are transferable among *Prunus* species, the same SSR primers are used in order to detect intra-species variation in related species (Dirlewanger et al., 2002, Wünsch and Hormaza, 2002).

(Ganopoulos et al., 2011b) assessed the genetic diversity of 21 Greek sweet cherry cultivars with 15 microsatellite markers and detected a total of 92 alleles (mean number of alleles per locus=5.80). (Guarino et al., 2009) examined a collection of 60 sweet cherry accessions of an Italian germplasm to estimate genetic diversity and to identify genetic relationships among ancient accessions using a set of 28 microsatellite markers. The expected heterozygosity (H_e) ranged from 0.031 to 0.822 while the total probability of identity was 5.59×10^{-16} . (Sharma et al., 2015) determined the genetic diversity of 24 Czech sweet cherry cultivars using 16 microsatellite markers generating 70 alleles. The observed heterozygosity value of these cultivars ranged from 0.25 to 0.96 with an average of 0.72. (Farsad and Esna-Ashari, 2016) characterized 23 Iranian sweet cherry cultivars considered for breeding programs using 16 polymorphic microsatellite markers producing 177 alleles that varied from 4 to 16 alleles with a mean heterozygosity value of 0.82. (Stanys et al., 2012) characterized the genetic diversity of 31 Lithuanian sweet cherry cultivars with 14 microsatellite markers and found a H_o value of 0.65. These results are in accordance with the values obtained by other studies for sweet cherry: 0.59 (Schueler et al., 2003), 0.49 (Wünsch and Hormaza, 2002), 0.50 (Clarke and Tobutt, 2003), 0.61 (Vaughan and Russell, 2004), 0.71 (Ganopoulos et al., 2011b).

One of the recent advances in the field of PCR based methods is high resolution melting analysis (HRM), which is a powerful method for fingerprinting, detection of mutations, SNPs, polymorphisms and epigenetic differences in distinct DNA samples (Wittwer, 2009, Ganopoulos et al., 2011a). This technique is becoming the method of choice as it incorporates unique advantages over other genotyping methods: it is cost-effective, fast, simple, powerful and suitable for high-throughput and accurate analysis. (Fernandez i Marti et al., 2012) applied HRM analysis to SNP genotyping a collection of 104 sweet cherry cultivars. They found a mean value of six alleles per locus, the observed heterozygosity (H_o) ranged from 0.14 to 0.66 while the expected heterozygosity (H_e) was ranged from 0.38 to 0.73. Moreover, (Ganopoulos et al., 2013) described the application HRM analysis for the characterization of PCR products and the identification of nine gene-based SNPs for distinguishing the main Greek sweet cherry cultivars. They found a H_e average value of 0.518 based on nine SNPs. The combined power of discrimination for the SNP markers was 0.999. Authors suggested that the ability of HRM to accurately discern nucleotide changes in a DNA sequence made it a cost- and time-effective alternative to traditional sequencing for the detection of gene-based SNPs.

Recently, (Campoy et al., 2016) genotyped a total of 210 genotypes including modern cultivars and landraces from 16 countries by using the RosBREED cherry 6 K SNP array v1. They found higher diversity in landraces compared to bred cultivars. These results were in agreement with the loss of diversity associated to breeding.

Sour Cherry

Various molecular markers and associated techniques are available today for investigation of plant germplasm and SSR markers have been widely used for genetic diversity in sour cherry (Cantini et al., 2001, Canli, 2004, Pedersen, 2006, Lacis, 2010, Antonius et al., 2012, Najafzadeh et al., 2016).

(Antonius et al., 2012) collected 77 cultivars established into the Finnish national sour cherry germplasm collection and studied its genetic diversity with nine microsatellite markers which produced 72 alleles in total. Furthermore, (Clausen et al., 2013) found no intra-cultivar variation in Denmark sour cherry clones by using 10 microsatellite markers. Recently, (Najafzadeh et al., 2016) used 19 microsatellite markers in order to investigate the genetic diversity of 12 Iranian sour cherry genotypes generating 148 alleles. Previously, (Cantini et al., 2001) demonstrated that 10 microsatellite markers could be used to differentiate among all but two of the 59 tetraploid cherry accessions examined in the USDA/ARS germplasm collection. The primer pairs amplified in their study ranged from 4 to 16 alleles with a mean of 10.7 alleles/locus.

(Peace et al., 2012) used a cherry-peach comparative genomics strategy to develop a moderate-density cherry SNP array relevant for sweet and sour cherry breeding germplasm based on SNPs discovered using next generation sequencing platforms. The array was evaluated using panels of sweet ($n = 269$) and sour ($n = 330$) cherry breeding germplasm.

S-Allele Genotyping

Many *Rosaceae* species have a distinct reproduction system that is characterized by self-incompatibility. This mechanism which restricts self-fertilization and promotes outbreeding controlled by alleles located at a particular locus, the *S*-locus. For incompatibility to occur an *S*-allele in the pollen tube needs to be in common with either of the two alleles in the stylar parent; in such as case, the pollen tube will stop developing and will not fertilize the egg (Chasan, 1991). The end result is that pollen from the same plant is rejected (self-incompatibility) and this is extended to any fertile pollen from any other cultivar that shares an *S*-allele with the stylar (De Nettancourt, 2001). Thus, all commercial cultivars which share *S*-alleles need the presence of trees, which will serve as efficient pollinators. Relevant research has led to the identification of self-incompatible and self-compatible genotypes (Lansari and Iezzoni, 1990, Yamane et al., 2001, Hauck et al., 2002). Due to the fact that sour cherry is a tetraploid hybrid of diploid sweet cherry and tetraploid ground cherry, the self-incompatibility mechanism appears to be conserved only in certain genotypes.

Sweet and sour cherry share many of the naturally occurring *S*-haplotypes. Yet there are other *S*-haplotypes like the S_1 , S_6 , and S_{13} , which have lost either pollen or stylar function and are present in sour cherry (Hauck et al., 2006, Tsukamoto et al., 2006). In terms of genetic information, the loss of function was due to DNA sequence alterations upstream of the *S-RNase*, *SFB* or *S-RNase* sequences.

Due to the *S*-incompatibility system most, if not all sweet cherry cultivars are self-incompatible, thus they require the presence of pollinators from specific compatible varieties. The standard practice in commercial orchards is to plant at every third position in every third

row for standard size trees, a pollinizer tree which belongs to a different incompatibility group. Yet this practice although it uses the minimum number of pollinator trees while ensuring at the same time the effective pollination and thus production has a major draw-back which is the loss of yield due to the presence of trees that do not fruit.

The S-incompatibility system allowed after the cloning and sequencing of the *S-RNase* locus from sweet cherry (Tao et al., 1999) the development of PCR methods suitable to distinguish the S-allele types. A large number of PCR primers has become available for S-RNase genotyping (Sonneveld et al., 2005, Wiersma et al., 2001), including the most important S-allele, the S_4' allele which confers self-compatibility (Ikeda et al., 2004). To date more than 20 S-RNases have been recognized leading to the development of 23 incompatibility classes and two SC classes containing the S_4' allele. The 23 incompatibility groups include only 10 S-alleles ($S_1 - S_7, S_9, S_{12}, S_{13}$) (Tobutt et al.).

The S-locus has also been used to investigate local variability of sweet cherries in many countries, for instance in Croatia (Ercisli et al., 2012), Germany (Schuster, 2012, Schuster et al., 2007), Greece (Ganopoulos et al., 2010), Latvia (Lacis et al., 2008), Sicily (Marchese et al., 2007), Spain (Cachi and Wünsch, 2014, Wünsch and Hormaza, 2004, Gisbert et al., 2008), Turkey (Ipek et al., 2011, Szikriszt et al., 2013) and wild cherry populations from Belgium (De Cuyper et al., 2005), Turkey (Szikriszt et al., 2013) and the UK (Vaughan et al., 2008). Certain S-haplotypes showed increased frequency in certain countries like haplotype S_{16} which was found in high frequency in the Sicilian germplasm, or the S_1 haplotype in cultivars from Germany (Schuster et al., 2007), and S_2 in Turkey (Ipek et al., 2011, Szikriszt et al., 2013). However, in many cases, the same S-haplotype was found in local material from distant European regions. In nineteen Greek cultivars, four alleles, S_1, S_3, S_4 , and S_9 were widespread and together were responsible for 85% of the S-haplotypes (Ganopoulos et al., 2010). Haplotype S_6 was highly frequent in genetic resources from Latvia and Scandinavia (Lacis et al., 2008), but is also very frequent in sweet cherries from the 'Jerte Valley' in Spain (Wünsch and Hormaza, 2004). (Cachi and Wünsch, 2014) performed the S-genotyping of 73 Spanish sweet cherry varieties and found that the S-haplotypes S_3, S_6 and S_{22} were the most frequent while the haplotype S_{16} was only found in the Balearic Islands. Furthermore, the S_{12} allele which was identified in wild cherry from Belgium and the UK (De Cuyper et al., 2005, Vaughan et al., 2008) was also present in high frequency in local material from Mediterranean countries like Croatia, Greece, Turkey and Spain (Gisbert et al., 2008, Ipek et al., 2011, Ercisli et al., 2012). On the contrary, allele S_{12} is absent in the native Greek populations (Ganopoulos et al., 2012) and in the collection of Greek sweet cherry cultivars (Ganopoulos et al., 2010).

Recently, S-alleles in 47 sweet cherry cultivars were recognized from local material of Central and Eastern Europe, and specifically mostly from Ukraine and the Czech Republic, 43 S-genotypes were found for the first time and the analyzed sweet cherry cultivars were assigned to 20 of the existing incompatibility groups (Lisek et al., 2015).

As S-incompatibility requires the presence of pollinators which reduces potential yield, sweet cherry breeding has focused (among others) in the production of commercial self-compatible varieties. Marker assisted selection for early detection of SC progeny is one possibility in sweet cherry breeding programs as the markers can identify the pollen-part mutant S_4' from its wild-type allele. This mutant allele has a four bp deletion in the *SFB4'* region (Sonneveld et al., 2005, Ushijima et al., 2004) and thus can be identified and distinguished from

its wild-type allele by a dCAPS marker system (Ikeda et al., 2004). However, unlike sweet cherry, individual sour cherry seedlings would need to be screened for multiple alleles.

CONCLUSION

The value of developing markers and the sequencing of whole genomes is the provision of a wealth of knowledge and data which can be used to understand basic molecular functions, and regulation of gene expression, especially for genes responsible for traits of agronomic interest in breeding programs. Thus, this knowledge could be translated into targeted and successful breeding program approaches. The development of saturated genetic maps enhanced by novel genetic markers has allowed the identification of chromosomal positions where disease resistance and fruit quality genes are located. Marker assisted selection can now be used in breeding programs.

The technological progress in sequencing and especially in next generation sequencing (NGS) has extended so far that genotyping by sequencing (GBS) is now possible (Davey et al., 2011). This leads to new pathways for genetic diversity management, particularly for conservation and survey of large populations. It is not far the day where new technologies such as NGS and GBS would let breeders and scientists to determine population characteristics establishing genome or nucleotide diversity. GBS would probably unlock the potential for global and quantitative management of genetic diversity, and let us predict an *a priori* genetic resource management.

REFERENCES

- Aka Kaçar, Y., Lezzoni, A. F. & Çetiner, S. 2005. Sweet cherry cultivar identification by using SSR markers. *Journal of Biological Sciences*, 5, 616-619.
- Antonius, K., Aaltonen, M., Uosukainen, M. & Hurme, T. 2012. Genotypic and phenotypic diversity in Finnish cultivated sour cherry (*Prunus cerasus* L.). *Genetic resources and crop evolution*, 59, 375-388.
- Arulsekhar, S. & Parfitt, D. E. 1986. Isozyme analysis procedures for stone fruits, almond, grape, walnut, pistachio, and fig. *HortScience (USA)*.
- Ayala, F. J. 1982. Genetic variation in natural populations: problem of electrophoretically cryptic alleles. *Proceedings of the National Academy of Sciences*, 79, 550-554.
- Beaver, J. A., Lezzoni, A. F. & Ramm, C. W. 1995. Isozyme diversity in sour, sweet, and ground cherry. *Theoretical and Applied Genetics*, 90, 847-852.
- Beyer, M., Hahn, R., Peschel, S., Harz, M. & Knoche, M. 2002. Analysing fruit shape in sweet cherry (*Prunus avium* L.). *Scientia Horticulturae*, 96, 139-150.
- Bianchi, V. J., Sansavini, S. & Fachinello, J. C. 2004. Microsatellite markers for identification of *Prunus* spp. rootstocks. *Scientia Agricola*, 61, 303-306.
- Cachi, A. M. & Wünsch, A. 2014. S-genotyping of sweet cherry varieties from Spain and S-locus diversity in Europe. *Euphytica*, 197, 229-236.
- Campoy, J. A., Lerigoleur-Balsemin, E., Christmann, H., Beauvieux, R., Girollet, N., Quero-García, J., Dirlwanger, E. & Barreneche, T. 2016. Genetic diversity, linkage

- disequilibrium, population structure and construction of a core collection of *Prunus avium* L. landraces and bred cultivars. *BMC plant biology*, 16, 1.
- Canli, F. 2004. Development of a second generation genetic linkage map for sour cherry using SSR markers. *Pak J Biol Sci*, 7, 1676-1683.
- Cantini, C., Iezzoni, A. F., Lamboy, W. F., Boritzki, M. & Struss, D. 2001. DNA fingerprinting of tetraploid cherry germplasm using simple sequence repeats. *Journal of the American Society for Horticultural Science*, 126, 205-209.
- Cerezo, M. & Arus, P. 1989. Identification of almond cultivars by pollen isoenzymes. *Journal of the American Society for Horticultural Science (USA)*.
- Chasan, R. 1991. Plant molecular biology blossoms in the desert. *The Plant Cell*, 3, 1255.
- Christensen, J. V. 1974. Numerical studies of qualitative and morphological characteristics of 41 sweet cherry cultivars. 2. *Tidsskrift for Planteavl*.
- Clarke, J. B. & Tobutt, K. R. 2003. Development and characterization of polymorphic microsatellites from *Prunus avium* 'Napoleon.' *Molecular Ecology Notes*, 3, 578-580.
- Clausen, S. K., Andersen, S. B., Henriksen, K., Toldam-Andersen, T. B. & Grout, B. W. W. 2013. Assessment of genetic diversity within sour cherry clones. *Scientia Horticulturae*, 164, 556-562.
- Corts, R. M., Rodrigues, L. C., Marcide, J. M. O. & Sánchez, R. P. 2008. Characterization of sour (*Prunus cerasus* L.) and sweet cherry (*Prunus avium* L.) varieties with five isozyme systems. *Revista Brasileira de Fruticultura*, 30, 154-158.
- Davey, J. W., Hohenlohe, P. A., Etter, P. D., Boone, J. Q., Catchen, J. M. & Blaxter, M. L. 2011. Genome-wide genetic marker discovery and genotyping using next-generation sequencing. *Nature Reviews Genetics*, 12, 499-510.
- De Cuyper, B., Sonneveld, T. & Tobutt, K. R. 2005. Determining self-incompatibility genotypes in Belgian wild cherries. *Molecular Ecology*, 14, 945-955.
- De Nettancourt, D. 2001. *Incompatibility and incongruity in wild and cultivated plants*, Springer Science & Business Media.
- Dirlwanger, E., Cosson, P., Tavaud, M., Aranzana, M., Poizat, C., Zanetto, A., Arús, P. & Laigret, F. 2002. Development of microsatellite markers in peach [*Prunus persica* (L.) Batsch] and their use in genetic diversity analysis in peach and sweet cherry (*Prunus avium* L.). *Theoretical and Applied Genetics*, 105, 127-138.
- El-Esawi, M., Bourke, P., Germaine, K. & Malone, R. 2012. *Assessment of morphological variation in Irish Brassica oleracea species*.
- Ercisli, S., Radunic, M., Gadze, J., Ipek, A., Skaljic, M. & Cmelik, Z. 2012. S-RNase based S-genotyping of Croatian sweet cherry (*Prunus avium* L.) genotypes. *Scientia horticulturae*, 139, 21-24.
- Farsad, A. & Esna-Ashari, M. 2016. Genetic diversity of some Iranian sweet cherry (*Prunus avium*) cultivars using microsatellite markers and morphological traits. *Cytology and Genetics*, 50, 8-19.
- Fernandez I Marti, A., Athanson, B., Koepke, T., Font I Forcada, C., Dhingra, A. & Oraguzie, N. 2012. Genetic diversity and relatedness of sweet cherry (*Prunus avium* L.) cultivars based on single nucleotide polymorphic markers. *Frontiers in plant science*, 3, 116.
- Frascaria, N., Santi, F. & Gouyon, P. H. 1993. Genetic differentiation within and among populations of chestnut (*Castanea sativa*. *Heredity*, 70, 634-641.

- Ganopoulos, I., Aravanopoulos, F., Argiriou, A. & Tsiftaris, A. 2012. Genome and population dynamics under selection and neutrality: an example of S-allele diversity in wild cherry (*Prunus avium* L.). *Tree genetics & genomes*, 8, 1181-1190.
- Ganopoulos, I., Argiriou, A. & Tsiftaris, A. 2011a. Microsatellite high resolution melting (SSR-HRM) analysis for authenticity testing of protected designation of origin (PDO) sweet cherry products. *Food Control*, 22, 532-541.
- Ganopoulos, I., Tsalaballa, A., Xanthopoulou, A., Madesis, P. & Tsiftaris, A. 2013. Sweet cherry cultivar identification by high-resolution-melting (HRM) analysis using gene-based SNP markers. *Plant Molecular Biology Reporter*, 31, 763-768.
- Ganopoulos, I. V., Argiriou, A. & Tsiftaris, A. S. 2010. Determination of self-incompatible genotypes in 21 cultivated sweet cherry cultivars in Greece and implications for orchard cultivation. *The Journal of Horticultural Science and Biotechnology*, 85, 444-448.
- Ganopoulos, I. V., Kazantzis, K., Chatzicharisis, I., Karayiannis, I. & Tsiftaris, A. S. 2011b. Genetic diversity, structure and fruit trait associations in Greek sweet cherry cultivars using microsatellite based (SSR/ISSR) and morpho-physiological markers. *Euphytica*, 181, 237-251.
- Gerlach, H. K. & Stösser, R. 1997. Patterns of random amplified polymorphic DNAs for sweet cherry (*Prunus avium* L.) cultivar identification. *Angewandte Botanik*, 71, 212-218.
- Gisbert, A. D., Badenes, M. L., Tobutt, K. R., Llacer, G. & Romero, C. 2008. Determination of the S-allele composition of sweet cherry (*Prunus avium* L.) cultivars grown in the southeast of Spain by PCR analysis. *The Journal of Horticultural Science and Biotechnology*, 83, 246-252.
- Granger, A. R., Clarke, G. R. & Jackson, J. F. 1993. Sweet cherry cultivar identification by leaf isozyme polymorphism. *Theoretical and Applied Genetics*, 86, 458-464.
- Guarino, C., Santoro, S., De Simone, L. & Cipriani, G. 2009. *Prunus avium*: nuclear DNA study in wild populations and sweet cherry cultivars. *Genome*, 52, 320-337.
- Hauck, N. R., Ikeda, K., Tao, R. & Iezzoni, A. F. 2006. The mutated S1-haplotype in sour cherry has an altered S-haplotype-specific F-box protein gene. *Journal of Heredity*, 97, 514-520.
- Hauck, N. R., Yamane, H., Tao, R. & Iezzoni, A. F. 2002. Self-compatibility and incompatibility in tetraploid sour cherry (*Prunus cerasus* L.). *Sexual Plant Reproduction*, 15, 39-46.
- Hillig, K. W. & Iezzoni, A. F. 1988. Multivariate analysis of a sour cherry germplasm collection. *Journal of the American Society for Horticultural Science (USA)*.
- Hjalmarsson, I. & Ortiz, R. 2000. In situ and ex situ assessment of morphological and fruit variation in Scandinavian sweet cherry. *Scientia horticulturae*, 85, 37-49.
- Höltken, A. M. & Gregorius, H.-R. 2006. Detecting local establishment strategies of wild cherry (*Prunus avium* L.). *BMC ecology*, 6, 1.
- Iezzoni, A. F. & Pritts, M. P. 1991. Applications of principal component analysis to horticultural research. *HortScience*, 26, 334-338.
- Ikeda, K., Igic, B., Ushijima, K., Yamane, H., Hauck, N. R., Nakano, R., Sassa, H., Iezzoni, A. F., Kohn, J. R. & Tao, R. 2004. Primary structural features of the S haplotype-specific F-box protein, SFB, in *Prunus*. *Sexual Plant Reproduction*, 16, 235-243.
- Ipek, A., Gulen, H., Akcay, M. E., Ipek, M., Ergin, S. & Eris, A. 2011. Determination of self-incompatibility groups of sweet cherry genotypes from Turkey. *Genet Mol Res*, 10, 253-260.

- Kappel, F., Granger, A., Hrotkó, K. & Schuster, M. 2012. Cherry. *Fruit Breeding*. Springer.
- Khadivi-Khub, A. 2014. Assessment of cultivated cherry germplasm in Iran by multivariate analysis. *Trees*, 28, 669-685.
- Khadivi-Khub, A., Zamani, Z. & Fatahi, M. R. 2012. Multivariate analysis of *Prunus* subgen. *Cerasus* germplasm in Iran using morphological variables. *Genetic resources and crop evolution*, 59, 909-926.
- Krahl, K. H., Lansari, A. & Iezzoni, A. F. 1991. Morphological variation within a sour cherry collection. *Euphytica*, 52, 47-55.
- Lacis, G. 2010. *Characterisation of the Latvian and Swedish sweet and sour cherry genetic resources*.
- Lacis, G., Kaufmane, E., Rashal, I., Trajkovski, V. & Iezzoni, A. F. 2008. Identification of self-incompatibility (S) alleles in Latvian and Swedish sweet cherry genetic resources collections by PCR based typing. *Euphytica*, 160, 155-163.
- Lacis, G., Kaufmane, E., Trajkovski, V. & Rashal, I. 2009a. Morphological variability and genetic diversity within Latvian and Swedish sweet cherry collections. *Acta Univ. Latv.*, 753, 19-32.
- Lacis, G., Rashal, I., Ruisa, S., Trajkovski, V. & Iezzoni, A. F. 2009b. Assessment of genetic diversity of Latvian and Swedish sweet cherry (*Prunus avium* L.) genetic resources collections by using SSR (microsatellite) markers. *Scientia Horticulturae*, 121, 451-457.
- Lansari, A. & Iezzoni, A. 1990. A preliminary analysis of self-incompatibility in sour cherry. *HortScience*, 25, 1636-1638.
- Lisek, A., Rozpara, E., Głowacka, A., Kucharska, D. & Zawadzka, M. 2015. Identification of S-genotypes of sweet cherry cultivars from Central and Eastern Europe. *Hort. Sci.(Prague)*, 42, 13-21.
- Marchese, A., Tobutt, K. R., Raimondo, A., Motisi, A., Bošković, R. I., Clarke, J. & Caruso, T. 2007. Morphological characteristics, microsatellite fingerprinting and determination of incompatibility genotypes of Sicilian sweet cherry cultivars. *The Journal of Horticultural Science and Biotechnology*, 82, 41-48.
- Najafzadeh, R., Arzani, K., Bouzari, N. & Saei, A. 2016. Genetic variation and identification of promising sour cherries inferred from microsatellite markers. *Russian Journal of Genetics*, 52, 64-73.
- Ohta, S., Katsuki, T., Tanaka, T., Hayashi, T., Sato, Y.-I. & Yamamoto, T. 2005. Genetic variation in flowering cherries (*Prunus* subgenus *Cerasus*) characterized by SSR markers. *Breeding science*, 55, 415-424.
- Peace, C., Bassil, N., Main, D., Ficklin, S., Rosyara, U. R., Stegmeir, T., Sebolt, A., Gilmore, B., Lawley, C. & Mockler, T. C. 2012. Development and evaluation of a genome-wide 6K SNP array for diploid sweet cherry and tetraploid sour cherry. *PLoS One*, 7, e48305.
- Pedersen, B. H. 2006. DNA fingerprints of 51 sweet and sour *Prunus* accessions using simple sequence repeats. *The Journal of Horticultural Science and Biotechnology*, 81, 118-124.
- Petrucelli, R., Ganino, T., Ciaccheri, L., Maselli, F. & Mariotti, P. 2013. Phenotypic diversity of traditional cherry accessions present in the Tuscan region. *Scientia Horticulturae*, 150, 334-347.
- Rakonjac, V., Akšić, M. F., Nikolić, D., Milatović, D. & Čolić, S. 2010. Morphological characterization of 'Oblačinska' sour cherry by multivariate analysis. *Scientia Horticulturae*, 125, 679-684.

- Rodrigues, L. C., Morales, M. R., Fernandes, A. J. B. & Ortiz, J. M. 2008. Morphological characterization of sweet and sour cherry cultivars in a germplasm bank at Portugal. *Genetic Resources and Crop Evolution*, 55, 593-601.
- Sánchez, R. P., Sánchez, M. A. G. & Corts, R. M. 2008. Agromorphological characterization of traditional Spanish sweet cherry (*Prunus avium* L.), sour cherry (*Prunus cerasus* L.) and duke cherry (*Prunus x gondouinii* Rehd.) cultivars. *Spanish Journal of Agricultural Research*, 6, 42-55.
- Santi, F., Lemoine, M. & Bruant, B. Genetic markers for *Prunus avium* L: inheritance and linkage of isozyme loci. 1990 1990. *EDP Sciences*, 131-139.
- Schueler, S., Tusch, A., Schuster, M. & Ziegenhagen, B. 2003. Characterization of microsatellites in wild and sweet cherry (*Prunus avium* L.) markers for individual identification and reproductive processes. *Genome*, 46, 95-102.
- Schuster, M. 2012. Incompatible (S-) genotypes of sweet cherry cultivars (*Prunus avium* L.). *Scientia Horticulturae*, 148, 59-73.
- Schuster, M., Flachowsky, H. & Köhler, D. 2007. Determination of self-incompatible genotypes in sweet cherry (*Prunus avium* L.) accessions and cultivars of the German Fruit Gene Bank and from private collections. *Plant Breeding*, 126, 533-540.
- Shahi-Gharahlar, A., Zamani, Z., Fatahi, M. R. & Bouzari, N. 2010. Assessment of morphological variation between some Iranian wild *Cerasus* sub-genus genotypes. *Hort Environ Biotechnol*, 51, 308-318.
- Sharma, K., Xuan, H. & Sedlák, P. 2015. Assessment of genetic diversity of Czech sweet cherry cultivars using microsatellite markers. *Biochemical Systematics and Ecology*, 63, 6-12.
- Sonneveld, T., Tobutt, K. R., Vaughan, S. P. & Robbins, T. P. 2005. Loss of pollen-S function in two self-compatible selections of *Prunus avium* is associated with deletion/mutation of an S haplotype-specific F-Box gene. *The Plant Cell*, 17, 37-51.
- Sosinski, B., Gannavarapu, M., Hager, L. D., Beck, L. E., King, G. J., Ryder, C. D., Rajapakse, S., Baird, W. V., Ballard, R. E. & Abbott, A. G. 2000. Characterization of microsatellite markers in peach [*Prunus persica* (L.) Batsch]. *Theoretical and Applied Genetics*, 101, 421-428.
- Stanys, V., Baniulis, D., Morkunaite-Haimi, S., Siksnianiene, J. B., Frercks, B., Gelvonauskiene, D., Stepulaitiene, I., Staniene, G. & Siksnianas, T. 2012. Characterising the genetic diversity of Lithuanian sweet cherry (*Prunus avium* L.) cultivars using SSR markers. *Scientia Horticulturae*, 142, 136-142.
- Stockinger, E. J., Mulinix, C. A., Long, C. M. & Brettin, T. S. 1996. A linkage map of sweet cherry based on RAPD analysis of a microspore-derived callus culture population. *Journal of Heredity*, 87, 214-218.
- Szikriszt, B., Doğan, A., Ercisli, S., Akcay, M. E., Hegedűs, A. & Halász, J. 2013. Molecular typing of the self-incompatibility locus of Turkish sweet cherry genotypes reflects phylogenetic relationships among cherries and other *Prunus* species. *Tree genetics & genomes*, 9, 155-165.
- Tao, R., Yamane, H. & Akira, H. 1999. Cloning of genomic DNA sequences encoding encoding S1-, S3-, S4-and S6-RNases (accession nos. AB031815, AB031816, AB031817, and AB0311818) from sweet cherry (*Prunus avium* L.). *Plant Physiol*, 121, 1057.
- Tobutt, K. R., Sonneveld, T. & Bošković, R. Cherry (in) *compatibility genotypes-harmonization of recent results from UK, Canada, Germany, Japan and USA*. 2001. 41-46.

- Tsukamoto, T., Hauck, N. R., Tao, R., Jiang, N. & Iezzoni, A. F. 2006. Molecular characterization of three non-functional S-haplotypes in sour cherry (*Prunus cerasus*). *Plant molecular biology*, 62, 371-383.
- Ushijima, K., Yamane, H., Watari, A., Kakehi, E., Ikeda, K., Hauck, N. R., Iezzoni, A. F. & Tao, R. 2004. The S haplotype-specific F-box protein gene, SFB, is defective in self-compatible haplotypes of *Prunus avium* and *P. mume*. *The Plant Journal*, 39, 573-586.
- Vaughan, S. P., Bošković, R. I., Gisbert-Climent, A., Russell, K. & Tobutt, K. R. 2008. Characterisation of novel S-alleles from cherry (*Prunus avium* L.). *Tree Genetics & Genomes*, 4, 531-541.
- Vaughan, S. P. & Russell, K. 2004. Characterization of novel microsatellites and development of multiplex PCR for large-scale population studies in wild cherry, *Prunus avium*. *Molecular Ecology Notes*, 4, 429-431.
- Wiersma, P. A., Wu, Z., Zhou, L., Hampson, C. & Kappel, F. 2001. Identification of new self-incompatibility alleles in sweet cherry (*Prunus avium* L.) and clarification of incompatibility groups by PCR and sequencing analysis. *Theoretical and Applied Genetics*, 102, 700-708.
- Wittwer, C. T. 2009. High-resolution DNA melting analysis: advancements and limitations. *Human mutation*, 30, 857-859.
- Wünsch, A. & Hormaza, J. I. 2002. Molecular characterisation of sweet cherry (*Prunus avium* L.) genotypes using peach [*Prunus persica* (L.) Batsch] SSR sequences. *Heredity*, 89, 56-63.
- Wünsch, A. & Hormaza, J. I. 2004. Molecular evaluation of genetic diversity and S-allele composition of local Spanish sweet cherry (*Prunus avium* L.) cultivars. *Genetic Resources and Crop Evolution*, 51, 635-641.
- Yamane, H., Tao, R., Sugiura, A., Hauck, N. R. & Iezzoni, A. F. 2001. Identification and characterization of S-RNases in tetraploid sour cherry (*Prunus cerasus*). *Journal of the American Society for Horticultural Science*, 126, 661-667.

Chapter 41

RESEARCH ON PURPLE SEED STAIN OF SOYBEAN: GERMPLASM SCREENING AND GENETIC RESISTANCE

**Shuxian Li^{1,*}, Pengyin Chen², Bo Zhang³
and Grover Shannon⁴**

¹United States Department of Agriculture-Agricultural Research Service,
Crop Genetics Research Unit, Stoneville, MS, US

²Department of Crop, Soil, and Environment Sciences, University of Arkansas,
Fayetteville, AR, US

³Department of Crop and Soil Environmental Sciences, Virginia Tech,
Blacksburg, VA, US

⁴Division of Plant Sciences, University of Missouri, Portageville, MO, US

ABSTRACT

Soybean purple seed stain (PSS) causes seed decay and purple seed discoloration, resulting in overall poor seed quality and reduced market grade and value. It is a prevalent soybean disease that also affects seed vigor and stem establishment. PSS is caused by the fungus *Cercospora kikuchii* and other *Cercospora* spp. The most common symptom of this disease occurs on the seed. Infected seeds may appear healthy or have discoloration in seed coat varying from pink to light or dark purple spots with range in sizes from a small speck to the entire seed coat. Warm and humid environments favor pathogen growth and disease development. Management strategies for this disease include crop rotation with non-legume or non-host crops, fungicides applications, and tilling the soil to disrupt spore dissemination. Along with these strategies, the use of resistant cultivars may provide more reliable and economical control of PSS, especially when environmental conditions are conducive for disease development. In this chapter, general information about the PSS and an overview of research on germplasm screening and genetic resistance are presented and discussed.

Keywords: soybean, purple seed stain, *Cercospora kikuchii*, seed quality

* Corresponding Author's Email: Shuxian.li @ars.usda.gov (Tel: +1 (662)686-3061; Fax: +1 (662)686-5218).

INTRODUCTION

Purple seed stain (PSS) of soybean (*Glycine max* (L) Merr.) causes seed decay and purple seed discoloration, resulting in overall poor seed quality and reduced market grade and value (Wilcox and Abney 1973; Roy and Abney 1976; Walters 1980; Yeh and Sinclair 1982; Ward-Gauthier, et al. 2015). It is a prevalent disease that also affects seed germination, seed vigor and stand establishment (Pathan et al. 1989; Schuh 1999).

PSS was first reported in Korea in 1921 and was described as a purple discoloration of the seeds (Suzuki, 1921). In 1925, Matsumoto and Tomoyasu found a fungus “which seemed to belong to the genus *Cercospora*” growing in the purple seed coats of soybeans (Matsumoto and Tomoyasu, 1925). In the United States, PSS was first documented in Indiana in 1924 (Gardener, 1926) and then in North Carolina in 1927 (Lehman, 1928). At the present time, PSS occurs in most soybean production areas worldwide (Roy and Abney, 1976; Grau et al., 2004). Soybean yield loss in the United States due to PSS has been reported at 85.2 thousand metric tons (Wrather et al. 2010).

DISEASE SYMPTOMS

The most common and easily recognized symptom of PSS occurs on the seed (Schuh, 1999). Infected seeds may appear healthy or have discoloration varying from pink to light or dark purple spot on the seed coat. The size of discoloration ranges from small pinpoint spot to covering the entire seed coat (Figure 1). Many small cracks are usually found in the discolored areas on the seed coat making seeds appear rough and dull.

In addition to symptoms occurring on the seed, the disease develops on pods, stems, and leaves. Cerospora leaf blight (CLB) in soybean is an important disease associated with PSS (Ward-Gauthier et al., 2015). In general, leaf symptoms do not show until late in the season when plants reach R5 to R6 growth stages. In the field, the first visible symptom of CLB occurs on the top surface of leaves exposed to the sun with light purple discoloration. The leaves appear leathery and 'sunburned'. The discolored areas may be small, discrete, and form irregularly shaped spots at the beginning. As the disease progresses, the infected leaf area expands and deepens in color to reddish-purple or bronze. In some cases, both upper and lower surfaces of entire leaves can be completely discolored. Severe infection causes necrosis of leaf tissue resulting in defoliation.

THE CAUSAL AGENTS

It has been reported that both PSS and CLB are caused by the fungal pathogen *Cercospora kikuchii* (Tak. Matsumoto & Tomoy.) M. W. Gardner (Walters, 1980). The scientific classification of *C. kikuchii* is described by Wikipedia (https://en.wikipedia.org/wiki/Cercospora_kikuchii) as follows: “Kingdom: Fungi; Phylum: Ascomycota; Class: Dothideomycetes; Subclass: Dothideomycetidae; Order: Capnodiales; Family: Mycosphaerellaceae; Genus: *Cercoseptoria*; Species: *C. kikuchii*.” The binomial name is

Cercospora kikuchii (Matsumoto & Tomoy, 1925)". The vegetative compatibility groups and the population structure of *C. kikuchii* have been investigated (Cai and Schneider, 2005; 2008).



Figure 1. Symptoms of purple seed stain in soybean.

It was first documented in 1957 that *C. kikuchii* produces a non-host specific phytotoxin (Kuyama and Tamura, 1957). The toxin is named cercosporin, a red perylene quinone pigment. It has a molecular weight of 534 and is photoactivated (Upchurch et al., 1991). Later, cercosporin was found in different *Cercospora* spp. (Assante, et al., 1977; Lynch and Geoghegan, 1977). Since cercosporin has been isolated from necrotic lesions of infected plants, studies on the role of cercosporin in PSS and CLB disease development were begun. Upchurch et al. (1991) used *C. kikuchii* mutants that blocked cercosporin synthesis to demonstrate that cercosporin was an important factor in the pathogenicity for infection of soybean (Upchurch et al., 1991).

In past decades, *C. kikuchii* was thought to be the only causal agent of PSS and CLB. Recently, more *Cercospora* species were reported to infect soybean across the Americas (Soares et al., 2015). Analysis of eight nuclear genes and a mitochondrial gene; surveys of amino acid substitutions; assessment of cercosporin production; and multiple *Cercospora* spp. including *C. cf. flagellaris* were reported to be associated with PSS and CLB (Soares et al., 2015). Some *Cercospora* isolates that caused PSS but not CLB have been found (S. Li, unpublished). Identification of all *Cercospora* species that cause PSS and/or CLB is underway. Investigation of the genetics basis of pathogenicity related genes using genomics approaches will aid in the development of new control strategies for the *Cercospora* pathogens.

DISEASE DEVELOPMENT AND EPIDEMIOLOGY

Development of PSS and CLB is influenced by environmental factors, such as relative humidity and composition and pH values of the substrates, as well as by light/photoperiod. It was reported that a minimum dew period of 18 hours is required for leaf and pod infection by *C. kikuchii* (Schuh, 1991). It also was reported that germination of *C. kikuchii* conidia was significantly lower at 24 h of light (35.6%) compared to photo periods of 24 h of light/24 h of dark (87.7%), 12 h of light/12 h of dark (88.8%), and 12 h of dark/12 h of light (89.5%), (Schuh, 1993). A photo period of 12 h of dark and 12 h of light was most conducive to infection. High-relative humidity (>95%) resulted in higher disease severity (Schuh, 1993).

In addition, seed infection and disease development are favored by high temperature and humidity during the early reproductive stages of soybean (Schuh, 1993). Leaf and pod infection occur at temperatures ranging from 15 to 32°C (Schuh, 1992), although optimum infection occurs at 25°C (Schuh, 1991).

DISEASE MANAGEMENT

Several strategies have been used to control the disease, such as tillage, crop rotation (Almeida, et al., 2001), fungicide applications at pod-filling stages (TeKrony, ET AL., 1985), and the use of genetic resistance (Jackson, et al., 2006; 2008; Ploper, et al., 1992; Srisombun and Supapornhemin, 1993; Wilcox, et al., 1975). The use of *C. kikuchii*-resistant genotypes to control the disease is economical and environmentally friendly.

Soybean Germplasm Screening for Resistance to Purple Seed Stain

To identify sources of resistance to PSS and CLB, a total of 123 plant introductions (PI) from 28 different countries, representing maturity groups (MG) III, IV, and V, were screened (Alloatti, et al., 2015b). Incidence of *Cercospora* leaf blight (% CLB), visual PSS (% PSS), and seed infected by *C. kikuchii* (% *C. kikuchii*) in harvested seed were determined. It was reported that in 2007, % *C. kikuchii* ranged from 2 to 51% for MG III, 2 to 35% for MG IV, and 0 to 33% for MG V. In 2008, % *C. kikuchii* ranged from 0 to 45% for MG III, 1 to 71% for MG IV, and 0 to 15% for MG V (Table 1). A total of 4 and 10 PIs from MG III and IV, respectively were identified as resistant to PSS in both years. There were significant positive correlations between inoculated vs. non-inoculated treatments and between % PSS and % *C. kikuchii* infection (Alloatti, et al., 2015b). The PSS resistant plant introductions identified in this study are valuable to breeders in developing resistant cultivars.

In other studies, 42 soybean lines from MG III, IV, and V with different reactions to other diseases in previous trials were tested for their reaction to PSS in field trials in Mississippi, USA (Li and Sciumbato, 2011a; b; 2012). Significant differences ($P \leq 0.01$) in seed infection by *C. kikuchii* were observed among soybean lines ranging from 2.0% to 16.0%. Results from tests of MG III soybean lines indicated that PI 578486 and PI 437482 were among the lines with lowest percentages of seed infection, while cultivar AG4403 had the highest percentage of seed infection. Of 14 lines tested, four lines had less than 5% seed infection (Li and

Sciumbato, 2012). Among MG IV soybeans, percentage of seed infection by *C. kikuchii* ranged from 3.0% to 16.5%. Soybean lines PI 346307 and PI 80479 were among the lines with lowest percentages of seed infection, while PI 58765 had the highest percentage of seed infection. Of 14 lines tested, five lines had less than 5% seed infection (Li and Sciumbato, 2011a). Results from the test of MG V soybean indicated that PI 407749 and PI 381659 were among the lines with lowest percentages of seed infection, while PI 417098 had the highest percentage of seed infection. Of 14 lines tested, four lines had less than 5% of seed infection (Li and Sciumbato, 2011b).

Table 1. Incidence of purple seed stain, *Cercospora kikuchii*, and *Cercospora* leaf blight in genotypes grouped by maturity and year (Alloatti et al., 2015b)

	% PSS ^b				% <i>C. kikuchii</i>				% CLB ^c	
	2007		2008		2007		2008		2007	
	mean	range	mean	range	mean	range	mean	range	mean	range
MG ^a and Treatment										
MG III										
Inoculated	22.4	0 - 51	5.7	0 - 20	24.6	2 - 51	11.5	0 - 45	4.2	0 - 15
Non-inoculated	15.6	0.1 - 47	5.4	0.3 - 20	22.8	1.3- 49	10.5	0 - 40	3.1	0 - 12
MG IV										
Inoculated	11.8	0 - 33	7.3	0 - 42	17.9	2 - 35	13	1.3 - 71	12.9	0 - 38
Non-inoculated	9.7	0 - 41	4.3	0 - 21	15.4	0 - 42	8.4	0.7 - 34	10.2	1.7- 30
MG V										
Inoculated	2.3	0 - 16.3	1.9	0 - 13	5.0	0 - 33	4.9	0 - 15	13.3	0- 36.7
Non-inoculated	1.3	0 - 15	1.0	0 - 5	3.8	0 - 29	4.7	0.7- 11.3	8.2	0 - 22

^a Maturity group.

^b % Purple seed stain.

^c % *Cercospora* leaf blight.

Evaluation of Soybean Cultivars for Reaction to *Cercospora* Leaf Blight

Cercospora leaf blight (CLB) has become a prevalent disease in many soybean production areas, especially in Louisiana, U.S.A. Cai et al. (2008) found that the Louisiana population of *C. kikuchii* was dominated by a new lineage that differed from those collected in other locations at earlier times. Then, they tested whether the dominance of the new lineage was caused by higher aggressiveness, and screened soybean cultivars for resistance to CLB. Representative isolates from both lineages were used individually to inoculate six soybean cultivars in the greenhouse. Unexpectedly, the new lineage was less aggressive than the old lineages. Three virulence groups were defined in this pathogen based on correlation of the aggressiveness of individual isolates on soybean cultivars. Results from the test of eleven soybean cultivars for disease reaction at two locations over 3 years in the field identified two cultivars, AG5701 (Asgrow) and TV59R85 (Terral), that were among the more resistant cultivars to CLB both in the greenhouse and in the field (Cai et al., 2009).

GENETIC RESISTANCE

Using genetic resistance is one of the most efficient and least expensive ways to control PSS and CLB. There are reports about host resistance to PSS. Srisombun and Supapornhemin (1993) found that resistance to PSS in soybean cultivar SJ.2 was controlled by a single dominant gene. Ploper et al. (1992) identified that PI 417274, PI 417460, and Gnome were resistant to both *Phomopsis* spp. and *C. kikuchii*, while PI 80837 was reported to possess resistance to PSS (Wilcox et al. 1975; Ploper et al. 1992). Heritability estimates of 0.91 in the F₂ and 0.51 in the F₃ generation for resistance to *C. kikuchii* were found for the progeny of the cross ‘Amsoy’ x PI 80837, which indicated a strong genetic component for resistance (Wilcox et al. 1975). Jackson et al. (2006) reported that PSS resistance was conditioned by a single dominant gene in PI 80837. The major PSS gene (*Rpss1*) was mapped on Chr. 18 between markers Sat_308 (6.6 cM) and Satt594 (11.6 cM) (Jackson 2004; Jackson et al. 2008).

Table 2. Analysis of variance for % PSS and % *C. kikuchii* infection in two soybean F_{2:5} populations derived from AP 350 x PI 80837 and PI 80837 x MO/PSD-0259 (Alloatti et al., 2015a)

	% PSS ^a					% <i>C. kikuchii</i> infection ^b			
Population	DF	MSE	F	P		DF	MSE	F	P
AP 350 x PI 80837									
Block	2	1943	25.0	< 0.0001		2	3877	3.3	< 0.0001
Genotype	80	9680	3.1	< 0.0001		80	35660	3.1	< 0.0001
PI 80837 x MO/PSD-0259									
Block	2	7.8	0.3	0.76		2	87.5	0.5	0.59
Genotype	86	4624	3.7	< 0.0001		86	21559	3.1	< 0.0001

^a % seed with purple seed stain based on visual assessment; DF = Degree of freedom;

^b % seed with *C. kikuchii* infection based on PDA plating; DF = Degree of freedom.

Recently, Alloatti et al., (2015a) investigated the inheritance of PSS resistance and identified microsatellite markers tightly linked to the major resistance gene *Rpss1* in PI 80837. In this study, two populations were developed by crossing a PSS resistant line, PI 80837, to PSS susceptible lines AP 350 and MO/PSD-0259. A total of 168 of F_{2:5} lines were grown at Kibler, AR in a randomized complete block design with three replications. Each plot was harvested and seed samples were taken to evaluate the percentages of visual PSS (% PSS) and *C. kikuchii* infection (% *C. kikuchii*). Ranges, LS means, and confidence intervals of the parents were used to classify resistant and susceptible reactions in the F_{2:5} lines for the two variables evaluated.

Significant differences in both % purple seed stain (PSS) and % *C. kikuchii* infection were observed between the resistant and susceptible parents and among the F_{2:5} lines, derived from AP 350 x PI 80837 and PI 80837 x MO/PSD-0259 (Table 2). The % *C. kikuchii* was consistently higher than % PSS among parents and F_{2:5} lines, indicating latent infection in seed. The resistant parent PI 80837 had significantly lower % PSS (3%) and % *C. kikuchii* (6.7%) than the susceptible parents AP 350 (34% PSS, 63.3% *C. kikuchii*) and MO/PSD-0259 (14.7% PSS, 39.0% *C. kikuchii*). LS means for % PSS and % *C. kikuchii* varied from 0.3 to 27.3% and 1 to 50%, respectively for the F_{2:5} lines from AP 350 x PI 80837. The F_{2:5} lines from PI 80837

x MO/PSD-0259 had 0.3 to 26% PSS and 1 to 47% *C. kikuchii*. For the AP 350 x PI 80837 population, the heritabilities for resistance to PSS and *C. kikuchii* were 0.62 and 0.60, respectively. For the PI 80837 x MO/PSD-0259 population, the heritabilities for % PSS and % *C. kikuchii* were 0.72 and 0.67, respectively (Alloatti et al., 2015a).

Alloatti et al. (2015a) also used sixteen SSR markers in a 17.1 cM region on Chromosome (Chr.) 18 to screen the parents and both F_{2:5} populations. Significant differences in % PSS and % *C. kikuchii* were observed for the parents and both F_{2:5} populations. For the two variables evaluated, both populations showed a good fit to a ratio of 15:1 (resistant to susceptible), indicating two dominant genes were involved for resistance. One chromosomal region in the vicinity of Satt115 and Satt340 on Chr. 18 was determined to be associated with the resistance gene *Rpss1* in both populations. These results confirm the presence of a major resistance gene, *Rpss1*, in PI80837 and also indicate an additional putative gene for PSS resistance (Alloatti et al., 2015a).

In future studies, screening the soybean genome with more molecular markers covering other soybean genomic regions to find the location for other gene should be considered. The PSS resistant PI, confirmation of the resistance gene and identification of an additional resistant gene will be valuable for breeders to develop resistant cultivars. Molecular markers linked to the resistance gene will facilitate selection for PSS resistance in future breeding efforts.

SUMMARY

Soybean purple seed stain (PSS) causes seed decay and purple seed discoloration, resulting in overall poor seed quality and reduced market grade and value. In this book chapter, we introduce the general aspects of PSS, its symptoms, causal agents, associated disease CLB, disease development and epidemiology. An overview of research on germplasm screening and genetic resistance are also presented and discussed. In future studies, all *Cercospora* species or other fungi that cause PSS and/or CLB need to be identified. Investigation of the genetics basis of pathogenicity and discovery of other useful genes using genomic approaches will aid in the development of new control strategies for the pathogens. Moreover, screening the soybean genome to identify molecular markers linked to PSS-resistance genes will be valuable for breeders in developing resistant cultivars, and will facilitate selection for PSS resistance in future breeding efforts.

ACKNOWLEDGMENTS

This work was partially supported by USDA-ARS (6402-21220-012-00D). Mention of trade names or commercial products in this publication is solely for the purpose of providing specific information and does not imply recommendation or endorsement by the United States Department of Agriculture. USDA is an equal opportunity provider and employer.

REFERENCES

- Alloatti, J., Chen, P. Zeng, A., Li, S., Rupe, J., Florez-Palacios, L., Orazaly, M. (2015a). Inheritance of and molecular markers for purple seed stain resistance in soybean. *Euphytica*. 206:701-709. 2015. DOI: 10.1007/s10681-015-1492-2.
- Alloatti, J., Li, S., Chen, P., Jauregui, L., Smith, S. F., Florez-Palacios, L., Orazaly, M. and Rupe, J. (2015b). Screening a diverse soybean germplasm collection for reaction to purple seed stain caused by *Cercospora kikuchii*. *Plant Dis.* 99: 1140-1146.
- Almeida, A. M. R., Saraiva, O. F., Farias, J. R. B., Gaudencio, C. A., and Torres, E. (2001). Survival of pathogens on soybean debris under no-tillage and conventional tillage systems. *Pesqui. Agropecu. Bras.* 36:1231-1238.
- Assante, G., Locci, R., Camarda, L., Merlini, L., and Nasini, G. (1977). Screening of genus *Cercospora* for secondary metabolites. *Phytochemistry*. 16: 243-347. 1977.
- Cai, G. and Schneider, R. W. (2005). Vegetative compatibility groups in *Cercospora kikuchii*, the causal agent of cercospora leaf blight and purple seed stain in soybean. *Phytopathology* 95: 257-261.
- Cai, G. and Schneider, R. W. (2008) Population structure of *Cercospora kikuchii*, the causal agent of cercospora leaf blight and purple seed stain in soybean. *Phytopathology* 98: 823-829.
- Cai, G., Schneider, R. W., and Padgett, G. B. (2009) Assessment of lineages of *Cercospora kikuchii* in Louisiana for aggressiveness and screening Soybean Cultivars for resistance to Cercospora leaf blight. *Plant Dis.* 93: 868-874.
- Gardener, M. W. (1926). Indiana plant diseases, 1924. *Proc. Indiana Acad. Sci.* 35: 237-257.
- Grau, C. R., Dorrance A. E., Bond, J., Russin, J. S. (2004). Fungal diseases. In: Boerma HR and Specht JE (eds) *Soybeans: Improvement, Production, and Uses*, 3rd ed. Agronomy Monograph 16. American Society of Agronomy, Crop Science Society of America, and Soil Science Society of America, Madison, WI, USA, pp. 679-764.
- Jackson E. W. (2004) *Resistance to Phomopsis seed decay and purple seed stain in soybean and virulence differences among Phomopsis spp. causing seed decay*. Dissertation, University of Arkansas, Fayetteville, AR.
- Jackson, E. W., Fenn, P., and Chen, P. (2006). Inheritance of resistance to purple seed stain caused by *Cercospora kikuchii* in PI 80837 soybean. *Crop Sci.* 46:1462-1466.
- Jackson, E. W., Feng, C., Fenn P., and Chen P. (2008). Genetic mapping of resistance to purple seed stain in PI 80837 soybean. *J. Hered.* 99:319-322.
- Kuyama, S., and Tamura, T. (1957). A pigment of *Cercospora kikuchii* (Matsumoto et Tomoyasu). Cultivation of the fungus, isolation and purification of pigment. *J. Am. Chem. Soc.* 79: 5725-5726.
- Lehman, S. G. (1928). Soybean diseases. Fiftieth Annual Report of the North Carolina. *Agric. Expt. Sta.*, pp. 95.
- Li. S. and Sciumbato, G. (2011a). Evaluation of maturity group IV soybean lines for resistance to purple seed stain in Mississippi, 2010. Plant Disease Management Report. doi:10.1094/PDMR05: FC123. *American Phytopathological Society Press*, St. Paul, MN.
- Li. S. and Sciumbato, G. (2011b). Reaction of maturity group V soybean lines to purple seed stain in Mississippi, 2010. Plant Disease Management Report. doi:10.1094/PDMR05: FC122. *American Phytopathological Society Press*, St. Paul, MN.

- Li, S. and Sciumbato, G. (2012). Evaluation of maturity group III soybean lines for resistance to purple seed stain in Mississippi, 2010. Plant Disease Management Reports. doi:10.1094/PDMR06: ST004. American Phytopathological Society Press, St. Paul, MN.
- Lynch, F. J., and Geoghegan, M. J. (1977). Production of cercosporin by *Cercospora* species. *Tran. Br. Mycol. Soc.* 69: 496-498.
- Matsumoto, T. and Tomoyasu, R. (1925). Studies on the purple speck of soybean seed. *Ann. Phytopath. Soc. Japan* 1 (6): 1-14.
- Pathan, M. A., Sinclair, J. B., McClary, R. D. (1989). Effects of *Cercospora kikuchii* on soybean seed germination and quality. *Plant Dis.* 73:720-723.
- Ploper, L. D., Abney, T. S., and Roy, K. W. (1992). Influence of soybean genotype on rate of seed maturation and its impact on seedborne fungi. *Plant Dis.* 76:287-292.
- Roy KW, Abney TS (1976). Purple seed stain of soybeans. *Phytopathology* 66:1045–1049.
- Schuh, W. (1991). Influence of temperature and leaf wetness period on conidial germination in vitro and infection of *Cercospora kikuchii* on soybean. *Phytopathology* 81: 1315-1318.
- Schuh, W. (1992). Effect of pod development stage, temperature, and pod wetness duration on the incidence of purple seed stain of soybeans. *Phytopathology* 82:446-451.
- Schuh, W. (1993). Influence of interrupted dew periods, relative humidity, and light on disease severity and latent infections caused by *Cercospora kikuchii* on soybean. *Phytopathology* 83: 109-113.
- Schuh, W. (1999). Cercospora blight, leaf spot, and purple seed stain. Pages 17-18 in: G. L. Hartman et al. (ed.). *Compendium of Soybean Diseases*. 4th ed. APS Press, St. Paul, MN.
- Soares A. P. G., Guillin E. A., Borges, L.L., Silva, A. C. T. d, Almeida, Á. M. R. d., Grijalba, P. E., Gottlieb, A. M., Bluhm, B. H., Oliveira, L. O. d. (2015). More Cercospora Species Infect Soybeans across the Americas than Meets the Eye. *PLoS ONE* 10(8): e0133495. doi:10.1371/ journal.pone.0133495.
- Srisombun, S., and Supapornhem, P. (1993). Inheritance of soybean resistance to purple seed stain. *Soybean Genet. Newsl.* 20:92-93.
- Suzuki, K. (1921). *Studies on the cause of "shihan" of soybean*. Chosen Nakaiho 16 (2): 24-28. (In Japanese).
- TeKrony, D. M., Egli, D. B., Stucky, R. E., and Loeffler, T. M. (1985). Effect of benomyl applications on soybean seedborne fungi, seed germination, and yield. *Plant Dis.* 69:763-765.
- Upchurch, R. G., Walker, D. C., Rollins, J. A., Ehrenshaft, M., and Daub, M. E. (1991). Mutant of *Cercospora kikuchii* altered in cercosporin synthesis and pathogenicity. *Applied and Environmental Microbiology*. 57 (10): 2940-2945.
- Ward-Gauthier, N. A., Schneider, R. W., Chanda, A., Silva, E. C., Price, P. P., Cai, G. (2015). Cercospora leaf blight and purple seed stain. pp 37-41. In Hartman, G. L., Rupe, J. C., Sikora, E. J., Domier, L. L., Davis, J. A., and Steffey, K. L. (eds.) *Compendium of Soybean Diseases and Pets*, Fifth Edition. APS Press. Minnesota, USA.
- Walters, H. J. (1980). Soybean leaf blight caused by *Cercospora kikuchii*. *Plant Dis.* 64: 961-962.
- Wilcox, J. R. and Abney, T. S. (1973). Effects of *Cercospora kikuchii* on soybeans. *Phytopathology* 63:796-797.
- Wilcox, J. R., Laviolette, F. A., and Martin, R. J. (1975). Heritability of purple seed stain resistance in soybeans. *Crop. Sci.* 15:525-526.

- Wrather A, Shannon G, Balardin R, Carregal L, Escobar R, Gupta G K, Ma Z, Morel W, Ploper D, Tenuta, A (2010). Effect of diseases on soybean yield in the top eight producing countries in 2006. *Plant Health Progress* doi:10.1094/PHP-2010-0125-01-RS. Accessed November 1, 2014.
- Yeh, C. C., and J. B. Sinclair. 1982. Effect of *Cercospora kikuchii* on soybean seed germination and its interaction with *Phomopsis* sp. *Phytopathology* 105:265-270.

Chapter 42

**SOMATIC AND GONADAL TISSUE
CRYOPRESERVATION: AN ALTERNATIVE TOOL
FOR THE GERMPLASM CONSERVATION
IN WILD MAMMALS**

***Alexsandra Fernandes Pereira*, Alexandre Rodrigues Silva,
Gabriela Liberalino Lima and Andréia Maria da Silva***

Laboratory of Animal Germplasm Conservation, Federal Rural University of
Semi-Arid, Mossoró, RN, Brazil

ABSTRACT

The tissue cryopreservation represents an interesting tool for the conservation of animal biodiversity. The establishment of tissue banks has been indicated as a practical approach to the preservation of species and, associated with other biotechniques, it could provide the rescue or multiplication of endangered species. In general, a large number of wild species have been having their gonadal and somatic tissue cryopreserved and for this purpose, the vitrification is the method routinely used. There is a diversity in cryobiological properties and requirements among cell types within tissues, presenting a challenge for its procedure. Nevertheless, even with those obstacles, studies have shown satisfactory results in many wild mammalian species. The gonadal tissue use involves the possibility for the reestablishment of endocrine functions of the testes and ovary allowing the preservation and posterior use of spermatozoa and/or spermatogonial stem cell and oocytes for other assisted techniques. Recent developments in the autografting and xenografting of testes and ovary clearly demonstrated the potential value of cryopreserving gonadal tissue. Already on the somatic tissue, skin samples have been widely utilized because of the possibility of sampling a large group of animals, without a dependency of limitations regarding gender or age. Moreover, this tissue can be obtained quite easily at using a simple methodology with a reduced cost. In this sense, this chapter highlights the importance of

* Corresponding Author's Email: alexsandra.pereira@ufersa.edu.br (UFERSA. BR 110, Km 47, Costa e Silva, Mossoró, RN, Brazil. Postal Code: 59640-900).

applying tissue cryopreservation to wild mammals conservation at showing the most recent studies in this area and the perspectives for its use in conservative programs.

Keywords: assisted reproduction, biodiversity, conservation, resource banks

1. INTRODUCTION

In the global overview, where the species extinction is an increasing problem, the worldwide conservation is a major concern due the biodiversity role in supporting life on Earth (Lopes et al., 2016). In this context, the establishment of germplasm banks is of a great importance, ensuring a genetic source of gametes that could be used in Assisted Reproduction Techniques or ART programs (Norton, 2014). Moreover, the germplasm cryopreservation is an effective approach for conservation and recovery of genetic diversity of wild mammals because it is not limited by the reproductive life span or the location of the breeders (Liu et al., 2012).

Currently, the conventional method of the effective preservation of endangered animal species is the cell cryopreservation; nevertheless, the cell cryopreservation cannot save tissue cells at all periods. Thun, the tissue cryopreservation is an alternative methodology for conservation of genetic material by establishment of biological resource banks or cryobanking (León-Quinto et al., 2009), aiming still its application in other reproductive biotechnologies, as somatic cell nuclear transfer or SCNT (Guan et al., 2010) and intracytoplasmic sperm injection or ICSI (Campos-Junior et al., 2014).

These resource banks can be used so both for *in situ* and *ex situ* conservation. Some studies have reported success in establishment of the banks derived from ovarian (León-Quinto et al., 2009), testicular (Campos-Junior et al., 2014) and somatic (León-Quinto et al., 2009) tissues. In this sense, progress has been made with different methodologies in same species and for each tissue and this chapter aimed to show the theoretical and technical aspects involved and the peculiarities observed in each tissue and/or species.

2. THE TESTICULAR TISSUE

In mammals, the spermatogenesis is a continuous, strictly regulated process, where spermatogonia multiply and differentiate into spermatozoa (Russell et al., 1990). This regulation involves all the microenvironment around these cells, called niche, which gives support and produces signals that regulate both cell renewal and spermatogonial differentiation. Thus, protocols for the testicular tissue cryopreservation should promote the preservation not only of spermatogonia, but of the whole niche around them, the Sertoli cells, myoid cells, Leydig cells and other interstitial cells, preserving the communication among them (Ning et al., 2012).

The cryopreservation of testicular samples is a technique applicable to different wild mammals, allowing the recovery and storage of genetic potential from valuable males. It consists on a practical method to be used when other techniques, as semen cryopreservation, are not available or not applicable (Abrishami et al., 2010). As an example, when animals in

zoological collections die or are castrated, the reproductive potential contained in their testicular tissue is currently discarded. Thus, the possibility to salvage this material based in the cryopreservation research could represent a valuable addition to genome resource banking (Pukazhenthil et al., 2006). Moreover, the technique could be applied to individuals in any reproductive stage (León-Quinto et al., 2009), and not necessarily after testicular excision, since there is the possibility of testicular tissue recovery through appropriate biopsy techniques.

In most cases, the testicular tissue cryopreservation has been associated to the xenografting procedures (Kaneko et al., 2013; Pothana et al., 2015). During the xenotransplantation of fragments into immunodeficient mice, a functional communication between recipient brain and donor testis has been established. This interaction allows the progression of spermatogenesis and the further recovery of fertile sperm (Arregui et al., 2014). Nevertheless, the obtaining of sperm derived from fresh xenografted testis tissue is only reported in few wild species, as bison (*Bison bison*, Abbasi et al., 2011), white-tailed deer (*Odocoileus virginianus*, Abbasi et al., 2012), and collared peccary (*Pecari tajacu*, Campos-Junior et al., 2014). Already, the results of sperm obtaining after testis tissue cryopreservation is even more rare in wild species, being only reported for rhesus monkey (*Macaca mulatta*, Poels et al., 2012) and Indian spotted mouse deer (*Moschiola indica*, Pothana et al., 2015).

Due to the importance and great perspectives for its application in wild mammals, our goal was to list the current relevant research conducted on the cryopreservation of testicular tissue, addressing aspects as the tissue collection, preparation, cryopreservation and evaluations.

2.1. Tissue Source

The collection of testicular tissue can be conducted from both pre and post pubescent individuals, as well as from live or recently dead animals. The death of animals is one of the main reasons for the loss of genetic diversity of endangered species. Moreover, sperm cannot be collected from immature males and cryobanking of testicular tissue combined with testis xenografting is a potential option for their conservation (Pothana et al., 2015).

The testis development is unusual because several cell types as Sertoli, Leydig, and spermatogonial cells arise from bipotential precursors present in the precursor tissue, the genital ridge (Svingen & Koopman, 2013). In this sense, the pre pubertal period is characterized by diverse population changes in the seminiferous epithelium and formation of the tubular lumen (Assis Neto et al., 2003; Aponte et al., 2005), which drastically differs from the morphological and physiological features present in post pubertal individuals. In the testicle of pubertal animals, the seminiferous tubules development is complete, as well as the maturation of Sertoli cells, hematotesticular barrier and spermatogenesis (Svingen & Koopman, 2013).

In the rare studies conducted in wild animals, the testicle has been completely collected and processed for preservation after the animal death, or after castration procedures (Thuwanut et al., 2013; Pothana et al., 2015). To our knowledge, there is no report regarding the application of biopsy techniques for testicular tissue recovery and preservation in wild mammals; however, we emphasize that this is an interesting alternative to be used with conservative purposes.

2.2. Tissue Preparations

In general, a nutritive medium is necessary for recovering and washing the testicular tissue immediately after its removal of the animal. This medium largely varies among research groups. In a study conducted in various wild species (*Jungle cat*, *Felis chaus*, *Panthera leo*, *Panthera pardus*, *Rusa deer*, *Rusa timorensis*, *Muntiacus feae*, *Sumatran serows* and *Capricornis sumatraensis*), the testicular tissue was immediately collected after the death. Further, samples were washed in standard sterile saline (NSS) supplemented with 1% penicillin-streptomycin and this media was used for the transport to the laboratory (Thuwanut et al., 2013). In a similar trial conducted in marmoset monkey (*Callithrix jacchus*), testicular tissue was washed and transported in ice-cold using sterile Dulbecco's modified Eagle's medium (DMEM) (Schlatt et al., 2002), while the phosphate buffering saline (PBS) was used for Indian spotted mouse deer (Pothana et al., 2015).

Another factor that usually varies among researches consists on the fragment size that will be subjected to cryopreservation. Thus, each fragment of testicular tissue can be cut in 0.5–1.0 mm³ for marmoset monkeys (Schlatt et al., 2002), 3.0 mm³ in collared peccary (Borges et al., 2015c), 1.0–2.0 mm³ in the Indian spotted mouse deer (Pothana et al., 2015) and 2.7 mm³ for others wild animals (Thuwanut et al., 2013).

2.3. Cryopreservation Techniques and Evaluations

Freezing Protocols

The testicular tissue cryopreservation is a challenge, since many cell types are involved in its compartments (Hovatta, 2000). The spermatogonia, Sertoli cells and Leydig cells are rich in cytoplasm and therefore, they are highly vulnerable to freezing and thawing processes. These cells have a large amount of intracellular water, which increases the risk for the formation of intracytoplasmic ice crystals, leading to the destruction of organelles. Moreover, various other modifications may occur, as excessive cellular dehydration caused by osmotic shock after the formation of extracellular ice and cytoskeletal breakdown upon exposure to low temperatures (Woelders et al., 2004).

Thus, for initiate the cryopreservation, testicular tissue should be completely detached from the epididymides. After, fragments (0.5–0.3 cm³) can be cryopreserved by slow freezing, rapid freezing or vitrification, using media composed by buffering substances, cryoprotectant agents (CPAs), sugars and protein sources.

In marmoset monkeys, testicular tissue fragments were incubated in 1.5 mL Leibovitz-L15 medium supplemented with 1.5 M dimethyl sulfoxide (DMSO), 0.1 M sucrose and 1% human serum albumin (HSA) for 20 min. Then, it was stored in cryovials and subjected to a slow freezing protocol in which the cooling was conducted using a programmable freezer at –140°C. Samples were then plunged into liquid nitrogen (LN₂) and transferred to storage. Subsequently, the warming of the cryovials was at room temperature for 1 min plus 1 min in water bath at 37°C and CPAs were removed by descending gradient of DMSO (1.0, 0.5 and 0.0 M) (Schlatt et al., 2002). In addition, fragments of testicular tissue derived from Indian spotted mouse deer were exposed during 30 min to DMEM/F12 HEPES supplemented with 10% DMSO and 80% fetal bovine serum (FBS). Cryovials were stored in an isopropyl alcohol container and placed in –80°C freezer during 24 h, and then transferred to LN₂ for storage. After warming, fragments

were cultured in DMEM/F12 HEPES medium supplemented with 10% FBS, 100 IU/mL penicillin–streptomycin e 40 µg/mL gentamycin (Pothana et al., 2015).

Thuwanut et al. (2013), who preserved the testicular tissue from various species, described a fast freezing protocol. Initially, fragments were incubated in 0.5 mL freezing medium I composed of mTCM199 containing 25 mM HEPES, 10% FBS and 7.5% DMSO and 7.5% ethylene glycol, EG, at room temperature for 10 min. Then, fragments were incubated in a 0.5 mL freezing medium II containing mTCM199 supplemented with 0.5 M sucrose, 20% FBS, 15% DMSO and 15% EG, during 30 min at 4°C. Tissue was stored in pre-cooled (4°C) wall-side cryovials that were horizontally placed on a rack 4 cm above LN₂ vapor for 10 min and plunged into LN₂. Warming of the cryovials was conducted at room temperature for 10 min in solution of mTCM199, 20% FBS and 1.0 M sucrose.

A vitrification technique was described for freezing the Rhesus monkey testicular tissue (Poels et al., 2012). Samples were equilibrated during 10 min at 4°C in Leibovitz-L15 medium supplemented with 7.5% EG, 7.5% DMSO, 0.25 M sucrose and 25 mg/mL HAS. Then, a further incubation for 5 min at 4°C in Leibovitz-L15 medium with 15% EG, 15% DMSO, 0.5 M sucrose and 25 mg/mL HSA was conducted. Samples were finally plunged into LN₂, the straws were inserted into precooled cryotubes and stored for 24 h in LN₂. Fragments were further warmed at 35°C in a solution containing 1.0 M sucrose in Leibovitz L-15 with 25 mg/mL HAS, and finally transferred to baths of warming solutions with decreasing sucrose concentrations (0.5, 0.25, and 0 M).

Based on these statements, we can realize that different protocols have been described for the cryopreservation of testicular tissue of wild animals, and they seems to report variate results. However, this fact indicates that there is no ideal protocol for testicular tissue conservation, justifying the development of future studies in order to improve the protocols.

Addition of Cryoprotectant Agents

The addition of CPAs is an essential and crucial factor for the success of cryopreservation techniques (Mycock et al., 1995; Szein et al., 2001). The penetrant CPAs are small molecules that readily penetrate cell membranes, producing bonds among hydrogen ions with intracellular water molecules, reducing thereby the freezing temperature of the mixture, and preventing the formation of ice crystals (Szein et al., 2001). Several studies have examined the cryopreservation of testis cell suspensions or tissue fragments using penetrant CPAs, as glycerol, EG, DMSO and propanediol (Abrishami et al., 2010). The use of at 3.0 and 6.0 M concentrations was efficient for the testicular tissue vitrification in collared peccaries, promoting nuclear and epithelium conservation (Borges et al., 2015c). On the other hand, the DMSO has shown the most promising results for the testicular tissue preservation in marmoset monkeys (Schlatt et al., 2002) and Indian spotted mouse deer (Pothana et al., 2015). Furthermore, the DMSO with EG, has been successfully used in the cryopreservation of testicular tissue in several wild mammals (Thuwanut et al., 2013).

The non-penetrating CPAs are high molecular weight agents, which act by increasing the viscosity of the solution (e.g., polyvinyl alcohol) or by binding to the heads of the phospholipid groups (e.g., sucrose and trehalose), protecting cell membranes from the cold injuries (Santos et al., 2008). They remain in the extracellular medium, promoting cell dehydration and are used in conjunction with the penetrating CPAs. Among the non-penetrating CPAs, sucrose has been widely used for testicular tissue cryopreservation (Schlatt et al., 2002; Yamini et al., 2016).

Nevertheless, the metabolites from the degradation of the cell cryoprotectants can be toxic, which is a limiting factor for the successful use (Fahy, 2010).

Tissue Analysis

The immunochemistry for DNA analysis and the xenotransplantation for the evaluation of the proliferative activity are the main assessments currently used for cryopreserved testicular tissue (Pothana et al., 2015). Through histological analysis, slides stained in hematoxylin-eosin have been used in order to analyze the number of seminiferous tubules, epithelium quality and nucleus aspects (Milazzo et al., 2008; Borges et al., 2015c). In collared peccaries, a nuclear preservation with intact nucleolus and only few cell membrane detached from the epithelium of the seminiferous tubules was observed after the testicular tissue vitrification using EG (3.0 M or 6.0 M), besides a low number of cells presenting vacuoles (Figure 1) (Borges et al., 2015c).

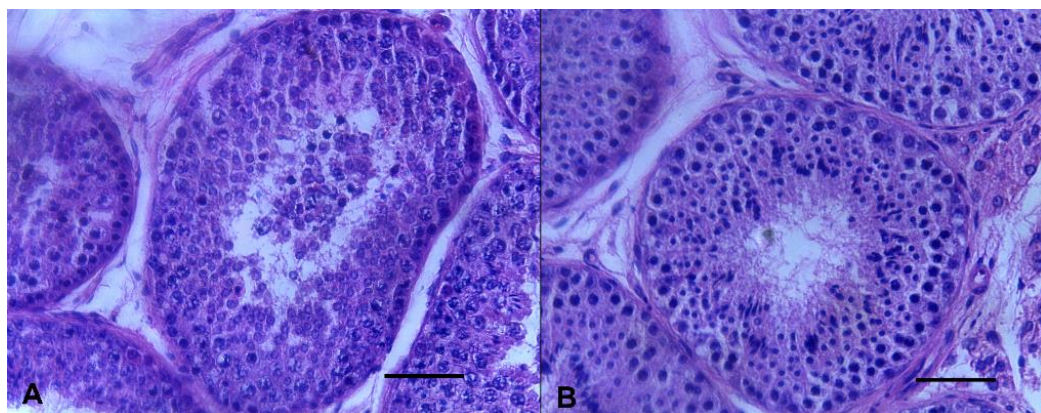


Figure 1. Testicular tissue vitrification in collared peccaries using 3.0 M EG as CPAs. (A) Fresh testicular tissue. (B) vitrified testicular tissue. Scale bar: 200 μ m.

In vitro growth of cells derived from male gonads has been reported as an efficient tool for the evaluation and development of cryopreserved tissue (Lee et al., 2016; Yokonishi et al., 2014). To optimize the culture medium, the use of supplements as specific growth factors for different cell types, hormones, vitamins and lipids are needed (Valk et al., 2010). In domestic animals, *in vitro* culture as classical organ culture method and 3D printing has been used (Sato et al., 2011); nevertheless, in wild mammals, only the use of xenotransplantation is reported (Campos-Junior et al., 2014).

The xenotransplantation, considered as an *in vivo* culture, is currently used for the development of spermatogenesis using cryopreserved testicular tissue (Abrishami et al., 2010; Mota et al., 2012; Lee et al., 2016). Briefly, the xenotransplantation involves the grafting of small pieces (1–2 mm³) of testis parenchyma from one species under the skin or testicular capsule of an orchidectomised immunodeficient mouse. Unlike attempts to reproduce spermatogenesis under *in vitro* conditions, this technique has the advantage of maintaining the complex architecture of the testis (Pukazhenthil et al., 2006). In wild mammals, xenotransplantation was performed in ferret (Gourdon & Travis, 2012) bison (Abbasi et al., 2011), white-tailed deer (Abbasi et al., 2012) and collared peccary (Campos-Junior et al., 2014).

while for cryopreserved testis tissues, it was only reported for Rhesus monkey (Poels et al., 2012) and Indian spotted mouse deer (Pothana et al., 2015).

In collared peccary, fresh testis fragments from pre-pubertal individuals were successfully preserved under the skin of SCID mice with complete spermatogenesis being observed at 6 months after grafting. Additionally, when testis cells isolated from the same peccaries were xenografted, they were able to interact and *de novo* testis morphogenesis occurred, with complete spermatogenesis being observed only at 8 months after xenografting (Campos-Junior et al., 2014).

In the Indian spotted mouse deer, the testicular tissue fragment was submitted to xenotransplantation and xenografts were recovered after 24 weeks, showing advanced germ cells. The nuclear staining confirmed the proliferative status of spermatocytes, the increases in the tubular lumen diameters and indicated testicular maturation (Pothana et al., 2015).

2.4. Use of Testicular Tissue in the Conservation Programs

The testicular tissue cryopreservation is an interesting tool to be used in programs for the conservation of genetic variability. This technique has been improved in domestic animals, in which the production of viable sperm undergoing ICSI was possible, generating normal offspring (Kaneko et al., 2013). In wild mammals, however, there are only few studies describing the application of this technique. In general, the testicular tissue comes derived from dead animals; nevertheless, the biopsy collection can be an alternative to tissue collection.

Thus, the prospects for the technical improvement and its use in conservation programs are promising. In 2002, the Biological Resource Bank for Spain endangered wildlife was raised. This bank has stored cryopreserved testicular tissue derived from Iberian lynx (*Lynx pardinus*) using the CPAs as DMSO and glycerol (León-Quinto et al., 2009). Following this example, other biobanks can apply this technique for valuable genetic material conservation.

3. THE OVARIAN TISSUE

The success of ARTs applied for the conservation of genetic material derived from endangered females is limited. The main causes involve the low number of matured oocytes that can be obtained through hormonal stimulation or during *post-mortem* gamete rescue (Jewgenow et al., 2011; Tanpradit et al., 2015). To circumvent these obstacles, the storage of intragonadal gametes by ovarian tissue cryopreservation has been suggest, representing a potential strategy for the preservation of germ cells of young individuals that die before reaching sexual maturity or without offspring (León-Quinto et al., 2009). In this sense, genome storage centers around the world are stimulating the ovarian tissue collection and cryopreservation from many threatened and critically endangered animal species as a means of preserving genetic diversity (Paris et al., 2004).

Mammals have a large ovarian reserve of preantral follicles (PAFs) stored in the ovarian cortex, from which, a great pool of oocytes are enclosed in primordial follicles that represents about 90% of all ovarian follicle population (Santos et al., 2010). These follicles are more resistant to low temperatures once the oocyte at this stage of development has low metabolic

rate, absence of meiotic spindle, pellucid zone and cortical granules, and smaller amount of lipid droplets than matured oocytes (Hovatta, 2005). Therefore, female fertility preservation by ovarian tissue cryopreservation is a powerful strategy for species conservation (Tanpradit et al., 2015).

Some oocyte characteristics represent challenges to define a cryopreservation protocol for these cells. Immature germinal vesicle stage oocytes have not yet formed spindle, have no cortical granules and have high membrane permeability. Due to this, it has been suggested that germinal vesicle stage oocytes might be more resistant to chilling injury than metaphase II oocytes (Tucker et al., 1998). However, some authors reported spindle abnormalities after *in vitro* maturation of frozen oocytes at germinal vesicle (Boiso et al., 2002; Stachecki et al., 2004).

Since that most oocytes are located within primordial follicles in the ovarian cortex, obtaining a small fragment of cortical tissue potentially enables cryopreservation of large numbers of oocytes at germinal vesicle stage (Practice Committee of American Society for Reproductive Medicine, 2014). Consequently, cryopreserving the ovarian tissue avoids many limitations found in matured oocytes preservation, as the low number of matured oocytes available in the ovaries and the possible deleterious effects of its conservation under low temperatures (Santos et al., 2010). Besides, many studies are focusing on ovarian tissue preservation due its large application on ARTs in association with transplant, promising tools for ovarian function resumption (Faustino et al., 2011). Because of the characteristics of being relatively resistant and tolerant to ischemia due the relatively low metabolic rate, it is assumed that PAFs are the first to benefit from neo-vascularization after transplantation (Aerts et al., 2010).

3.1. Tissue Source and Preparation

The ovaries can be obtained from animals of any age, including fetuses, as well as immediately after death (Cleary et al., 2001). The major limitation of its use is the difficulty of preserving ovary tissue, considering the diversity of cell types and tissue components (Hovatta, 2005), that vary in size, distribution, and quantity and by many follicular development stages cells. It represents an obstacle to find an optimal permeability for CPAs (Lunardi et al., 2016).

Despite the obstacles, the cryopreservation process can be applied using ovarian fragments obtained from live animals by laparoscopy (baboons, *Papio anubis* – Lu et al., 2014; Rhesus monkeys – Ting et al., 2013), ovariectomy (lions, *Panthera leo* – Wiedemann et al., 2012) and after animal death (common wombats, *Vombatus ursinus* – Wolvekamp et al., 2001).

The size of ovarian tissue fragments also plays an important role in the success of the cryopreservation procedures. Commonly, small fragments (<1.0 cm) are used, especially when vitrification techniques are performed (Amorim et al., 2011), allowing the penetration of the vitrification solution into tissue. Moreover, the cryopreservation of the whole ovary has been reported successfully in some species, as wallabies (*Macropus rufogriseus*, Mattis et al., 2002), yellow toothed caviar (*Galea spixii spixii*, Praxedes et al., 2014) and marmoset monkeys (*Callithrix jacchus*, Motohashi & Ishibashi, 2016). For all these cases, after removal of ovarian samples, they are washed in a variety of media supplemented with protein source and antibiotics, before being exposed to cryopreservation solution.

3.2. Cryopreservation Techniques and Evaluations

The cryopreservation process involves the steps of exposure to CPA, chilling, storage, thawing or warming and cryoprotectant removal (Santos et al., 2008). The step of the CPA addition is crucial and indispensable for the success of the preservation and its presence is necessary for the cell survival (Mycock et al., 1995). Nevertheless, the metabolites produced from the degradation of these substances inside the cell might be toxic and it is a limiting point for the success of technique (Fahy, 2010).

Regarding the methods for ovarian tissue cryopreservation, the slow freezing and vitrification are described for this purpose. The slow freezing or conventional freezing protocol involves ovarian tissue exposure to low concentration of CPAs, followed by a step-by-step temperature lowering, according to the program's settings of programmed freezers (Filatov et al., 2016). Some studies revealed that this procedure led to some damages on ovarian stromal, granulosa and theca cells (Amorim et al., 2012), resulted from intracellular ice formation during freeze-thawing processes, which damages cells and your interactions (Tanpradit et al., 2015).

The chilling injury usually occurs between +15 and -5°C and causes changes in the lipid droplets, membranes, and microtubule of the mitotic or meiotic spindle of the oocytes (Aman & Parks, 1994; Martino et al., 1996). On the other hand, ice crystal formation, which occurs between -5 and -80°C, is considered the major source of cryoinjuries (Paynter, 2005). The ice crystal formation occurs due a high instability state caused by super cooling of intracellular water, and it is responsible by the mechanical rupture of cell plasmatic membrane (Zhmakin, 2008). The thawing process might cause it too, forming microcrystals as the result of ice recrystallization (Salle et al., 2002). Both injuries are very common in slow freezing protocols; however, it has been suggested that rapid cooling might prevent it (Wang et al., 2012).

In order to avoid damages caused by conventional freezing protocols, vitrification procedures have been used. It is considered a cheap method that can be performed under field conditions with no need for special equipment, making it a good alternative for use in various settings often encountered with wildlife mammals (Saragusty & Arav, 2011), including after animal death (Amorim et al., 2011). It avoids the development of ice crystals (Mukaida & Oka, 2012), promoting the formation of an amorphous glassy solid state (-20.000 to -40.000°C/min, Lin et al., 2008), using high cooling rates and CPA concentrations (Chung et al., 2013).

The efficiency of ovarian tissue vitrification has made possible the current establishment of female germplasm banking in donor programmes (Nagy et al., 2009). It is proposed that this procedure would be the most significant advance in cryobiology (Amorim et al., 2011) and, in the future, it is expected that vitrification becomes the most adequate method for the cryopreservation of any cell and tissue (Mukaida & Oka, 2012).

When vitrification is adopted, the Leidenfrost effect is a factor that can affect heat transfer, thus, resulting in low cooling rate (Sansinena et al., 2010). It occurs when the object is plunged into LN₂, as soon as it happens, the heat flow from the object to nitrogen resulting in strong nitrogen vaporization around the sample surface, that acts as an insulator, delaying heat transfer. To avoid this phenomenon, it has been proposed the slush nitrogen method (Santos et al., 2012), that consists of applying negative pressure, obtaining a solid/liquid nitrogen mixture (slush), that increases the cooling rate (Talevi et al., 2016). Nevertheless, this method has not been tested in wild animals yet.

Due the lack of basic information about the reproductive physiology of the majority of wild species, and the ignorance on oocyte osmotic tolerance or toxicity tests (Comizzoli et al. 2012), female gamete cryopreservation is not a well-established technique for these species (Andrabi & Maxwell, 2007). Once established, knowledge on cryopreservation of follicles *in situ* could be applied to endangered species allowing the maintenance and exchange of rare female genome. In this regard, the development of *in vitro* culture systems capable to promote complete growth of immaturated oocytes, present in ovarian tissue, has to be accomplished (Lopes et al., 2016). The use of oocytes enclosed in PAFs would provide the rescue of over than 95% of all ovarian follicle population (Silva et al., 2004). The possibility of *in vitro* culture and maturation of such oocytes would provide new knowledge about the mechanisms involved in the folliculogenesis of a given species, and also offers opportunities for its use in many others reproductive techniques (Comizzoli et al., 2010). Still, the required conditions for the complete *in vitro* development of PAFs are not established, in wild mammals, once data regarding follicular and oocyte growth are scarce (Figueiredo et al., 2008).

As an alternative, ovarian tissue transplantation has been reported as a promising way for obtaining viable oocytes for further exploration in ARTs. As the use of *in vivo* strategies to mature frozen-thawed ovarian tissue for animal conservation have been limited due immunological barriers from grafting ovarian tissue between individuals, xenotransplantation is a way to overtake these limitations (Paris et al., 2004). Many reports pointed ovarian tissue can survive, grow and mature after xenotransplantation into different species and sexes, presenting a practical tool for the study of *in vivo* follicular development (Meirow et al., 2016). It not only provides the possibility of access a source of gametes from ovarian tissue, but also acts as a tool to understand the mechanism of follicular development (Tahaei et al., 2015; Kikuchi et al., 2011).

Promising results were reported in many species, including the offspring obtained from Rhesus monkeys freshly autografted ovarian tissue (Lee et al., 2004), and many children derived from cryopreserved ovarian tissue autograft (Donnez et al., 2015; Jensen et al., 2015). Therefore, it is hypothesized that oocytes derived from such grafted ovarian tissue have the potential to give rise to live young (Paris et al., 2004), and be applied for several other wild ones.

3.3. Current Applications of Ovarian Tissue in Conservation Programs

Ovarian tissue cryobanks allow the storage of a large source of follicles from which many matured oocytes may be subsequently used in ARTs. Several research groups and centers are working to collect and to preserve a variety of biological material from threatened or endangered mammalian species (Andrabi & Maxwell, 2007). In this sense, some reports (Table 1) present encouraging results for this use as a tool for endangered species conservation.

With the purpose of forming a biological resource bank, Leon-Quinto et al. (2009) cryopreserved the ovaries from six Iberian lynx females, 24 to 72 h after animal death. Thus, the ovary was divided into three parts and cryopreserved by using slow freezing protocols adapted from domestic mammals. This work highlights the importance and role of tissue cryopreservation for germplasm storage from endangered animals.

Table 1. Cryopreservation of ovarian tissue in some wild mammals

Technique	Species	Main outcomes	Authors
Vitrification	<i>Papio anubis</i>	Autografting: follicle survival, growth and ovulation	Lu et al. (2014); Nyachio et al. (2013)
	<i>Callithrix Jacchus</i>	Preservation of follicle morphology	Motohashi and Ishibashi (2016)
	<i>Macaca mulatta</i>	Follicle survival and growth; hormonal activity	Ting et al. (2013, 2012, 2011); Yeoman et al. (2005)
	<i>Macaca fascicularis</i>	Autotransplant, hormonal cycle restoration, ICSI	Suzuki et al. (2012); Hashimoto et al. (2010)
	<i>Peccary tajacu</i>	Preservation of follicle morphology	Lima et al. (2012)
	<i>Galea spixii spixii</i>	Preservation of follicle morphology	Praxedes et al. (2014)
Slow freezing	<i>Panthera leo</i>	Xenografting, normal morphology; follicle survival, growth and maturation	Wiedemann et al. (2012)
	<i>Iberian lynx</i>	Biological resource banks	León-Quinto et al. (2009)
	<i>Loxodonta africana</i>	Xenografting, tissue survival and function	Gunasena et al. (1998)
	<i>Vombatus ursinus</i>	Xenografting, tissue survival and function; follicle growth and development	Wolvekamp et al. (2001); Cleary et al. (2003, 2004)
	<i>Callithrix Jacchus</i>	Xenografting, preservation of follicle morphology; proliferative activity	Schönfeldt et al. (2011, 2012); Candy et al. (1995)
	<i>Macaca mulatta</i>	Follicle survival, growth, antrum formation and production of hormones	Yeoman et al. (2005); Ting et al. (2011)
	<i>Macaca fascicularis</i>	Autotransplant, hormone cycle restoration; ICSI	Schnorr et al. (2002); Yeoman et al. (2005)
	Wild carnivores*	Preservation of follicle morphology	Wiedemann et al. (2013)
	<i>Dasyprocta aguti</i>	Preservation of follicle morphology	Wanderley et al. (2012)
	<i>Macropus eugenii</i>	Xenografting, follicle development	Mattiske et al. (2002)

* African lions (*Panthera leo*), Amur leopard (*Panthera pardus orientalis*), Black-footed cat (*Felis nigripes*), Oncilla (*Leopardus tigrinus*), Geoffroy's cat (*Leopardus geoffroyi*), Northern Chinese leopard (*Panthera pardus japonensis*), Rusty-spotted cat (*Prionailurus rubiginosus*), Serval (*Leptailurus serval*), Sumatran tiger (*Panthera tigris sumatrae*).

Abbreviation: ICSI: intracytoplasmic sperm injection.

4. THE SOMATIC TISSUE

The cryopreservation of somatic tissue is an alternative source for the preservation of genetic material, aiming the conservation of animal biodiversity. The establishment of somatic sample banks has been indicated as a practical approach to the preservation of species and

associated with other reproductive biotechniques, as nuclear transfer (cloning), provides the restoration of endangered species (Loi et al., 2001; Folch et al., 2009).

In general, the cryopreservation of somatic samples has different applications beyond of the SCNT, including the preservation of the genetic material of the species. Thus, it can avoid the total loss of biodiversity in individuals who died prior to the reproductive stage, as well as allowing the study of the species by *in vitro* cell culture and basic research related to genetic, toxicology, among others (León-Quinto et al., 2009; 2011). Additionally, the cryo-preservation allows proper storage of samples, both during the transport to the laboratory as for a long time aiming the cryobank formation (Wong et al., 2012). In all cases, the use of conservation techniques as cooling, slow freezing and vitrification is required.

Different protocols have been employed and methodological adjustments, especially on the type and the concentration of the CPAs and technique type, are constantly being developed for wild species. Moreover, the methodologies applied for wild mammals are established in accordance with procedures used in domestic mammals.

Both somatic tissues as cells can be cryopreserved (León-Quinto et al., 2009; Machado et al., 2016); nevertheless, in situations where the habitat of animals of interest is difficult to access or far from *in vitro* culture laboratory, methods can be proposed for the conservation of somatic tissue before cell recovery (Silvestre et al., 2003). Additionally, use of tissue samples has also the advantages that can be cryopreserved without the need of a specialized laboratory, in situations where the procedures of *in vitro* culture are not yet established. Thus, an interesting step would be the establishment of the period and tissue storage conditions that do not influence the viability of cells recovered.

Therefore, we aimed here to approach the technical and theoretical aspects with emphasis in the source of somatic tissue and the recovery procedures, besides of the laboratory processing and the different freezing techniques and methods for its evaluation.

4.1. Tissue Sources

Somatic tissue samples have been widely used due the possibility of sampling a large group of animals. Moreover, it does not depend on the limitations of gender or age, allowing an unlimited number of samples and extrapolation of the cell characteristics of these animals for the entire population (Andrabi & Maxwell, 2007; León-Quinto et al., 2009). This is an interesting aspect, especially in the case of species from which is difficult to obtain gametes.

Nevertheless, the establishment of cryobanking requires the development of appropriate protocols and the suitable choice of the region or tissues to be collected from the animals (Cetinkaya & Arat, 2011). Moreover, this tissue should be obtained quite easily at using a simple methodology at low costs (Singh & Ma, 2014).

Skin Biopsy

In general, skin tissue fragments can be obtained by biopsies from live animals (Borges et al., 2015a,b) and/or during necropsies from dead individuals (Oh et al., 2008). In the latter situation, information on the cause and the *post-mortem* time are important for further procedures. Moreover, different solutions are described for the tissue collection as the phosphate buffer used for *P. tajacu* (Borges et al., 2015a,b) or a culture media used for *Ursus arctos* (Caamaño et al., 2008) and *Panthera tigris tigris* (Guan et al., 2010), both with

supplementations to ensure tissue viability during transport and/or procedures. In the latter work, somatic tissue derived from the skin were collected in DMEM supplemented with 100 U/mL ampicillin and 100 µg/mL streptomycin. Already in wild endangered Chilean native species (Tovar et al., 2008), biopsied skin, including dermal, epidermal and fat tissues were recovered in Earl Balanced Salt Solution (EBSS) plus antibiotic-antimycotic mixture.

In all situations (biopsies or necropsies), the region to be manipulated in the animal will be shaved and sanitized prior to collection. In general, skin fragments can be recovered of different regions, as auricular (*Bos gaurus*: Mahesh et al., 2012), abdominal (*Canis lupus*: Oh et al., 2008), scalp (*Cervus elaphus*; Berg et al., 2007) and fetal origin (*L. pardinus*; León-Quinto et al., 2014). In some species of difficult access, as *Hippocamelus biculcus* (Tovar et al., 2008) the darts can be used for to collect skin samples. For biopsies derived from the darts, a special system is used and it is fired to the area of the large muscle masses utilizing a projector powered by cartridge. Thus, a small sample is aseptically recovered from inside the dart, and destined to the following procedures.

Others Sources of Somatic Tissues

Due to the ease of harvesting for skin samples, few studies relate the conservation of other somatic tissues sources. An interesting example can be cited for *L. pardinus* (León-Quinto et al., 2009). In this study, the authors attempted to form a bank of somatic samples and recovered different somatic types, as muscle, oral mucosa, bone marrow, spine marrow and intestines, under the similar conditions used for skin samples. As a result, they obtained a total of 69 individuals (35 males and 34 females) from the somatic material collected. Thus, considering that in 2006 (period of experiment), the population of *L. pardinus* was about 170 individuals, a somatic bank reflected a very important fraction of the population biodiversity existing for the species.

4.2. Tissue Preparations

In the laboratory, all fragments can be washed through an aseptic procedure in order to ensure the viability of the samples. For that, various washing media can be employed. In general, biopsies should be washed in nutrient medium supplemented with antibiotics, buffers and/or protein sources. For *L. pardinus*, the authors described a medium consisting of DMEM supplemented with 25 mM HEPES, 100 U/mL penicillin, 0.1 mg/mL streptomycin and 1% fungizone (León-Quinto et al., 2014).

After washing, the determination of fragment size depends on the technique employed (Xia et al., 2013), but they usually measure around 1.0 mm³ (Song et al., 2007). Nevertheless, there are variations depending on the tissue availability (Tovar et al., 2008) and the cryopreservation type and materials used (Borges et al., 2015a,b). In *P. tajacu* and *Dasyprocta leporina* (Costa et al., 2015), for examples, samples were fragmented in 9.0 mm³ (3 x 3 x 1 mm), according to the vitrification protocol used.

4.3. Cryopreservation Techniques and Evaluations

After the obtainment of the tissue samples, these can be subjected to cooling (Tovar et al., 2008) and/or cryopreservation techniques (Table 2), as slow freezing (Caamaño et al., 2008; León-Quinto et al., 2009) and vitrification (Borges et al., 2015a,b; Costa et al., 2015). In general, the choice of method for tissue preservation depends on the required period and handling conditions.

The tissue samples can be firstly cooled in the presence or absence of nutrient medium over a relatively short-term, for example, for the transport of samples between geographically distant regions or intricate conditions. In the context, Tovar et al. (2008) verified the cooling conditions of tissue samples derived from different species for short-term at 4°C using chilled domestic cooler and EBSS as medium for conservation. These authors observed that skin samples derived from wild species of Chilean could be stored at 4°C for one week and still yield primary cultures with fibroblasts resulting in a 92% efficiency for the fragments conservation.

As another alternative, fragments could be stored for long periods in order to form biobanks. Both the methods (slow freezing or vitrification) consist on the exposure of tissue to CPAs, cooling storage, warming and removal of CPAs from tissue fragments (Silva et al., 2015). Although the methodologies currently applied for cell cryopreservation be routinely used and extrapolated for tissues, adaptations and establishment of protocols are necessary in order to adjust the specific requirements of the tissue (Zieger et al., 1996). This is more even complex because they contain different cell types that vary in volume and water permeability (Gandolfi et al., 2006).

For all processes, the use of penetrating and non-penetrating CPAs is necessary and different responses can be observed after their use. In *Sus scrofa*, for example, frozen skin fragments without the presence of CPAs showed culturable cells (Zhang et al., 2012). Already in *P. tajacu*, Borges et al. (2015a,b) obtained viable cells derived from tissues vitrified with DMSO, EG and sucrose. In *U. arctos*, viable cells were obtained from tissues frozen with DMSO and vitrified with the association of DMSO and EG (Caamaño et al., 2008). In this sense, the presence of CPAs and its choice are important factors for the efficiency of cryopreservation (Ehrlich et al., 2015). Additionally, during the tissue exposure to CPA followed of cooling, the main differences between the types of cryopreservation, slow freezing and vitrification are highlighted.

Briefly, the slow freezing protocol used for somatic tissue (Caamaño et al., 2008) consists in the following steps. First, the incubation of small fragments in 1.0 mL of freezing medium (DMEM supplemented with 10% DMSO and 10% FBS) using cryovials. Then, the cooling using the system Mr. Frosty (box containing isopropanol) at -80°C and overnight. Followed by the storage in LN₂. Finally, the warming of the cryovials at 38°C for 2 min, removal of freezing medium and mix with 1.0 mL of DMEM plus 10% FBS.

For the vitrification method, as occurs for other tissues, there are some variations (Carvalho et al., 2011, 2012). Two different systems have already been employed for wild animals. The first consists on a conventional method that uses cryovials and that was applied for *U. arctos* (Caamaño et al., 2008), *P. tajacu* (Borges et al., 2015a,b) and *D. leporina* (Costa et al., 2015). The other method consists on the solid-surface vitrification (SSV) procedure described for *P. tajacu* (Borges et al., 2015a,b) and *D. leporina* (Costa et al., 2015).

Table 2. Cryopreservation of somatic tissue in some wild mammals

Species	Tissue source	Cryopreservation				Authors
		Technique	Cryoprotectors	Evaluation ^a	Destination	
<i>U. arctos</i>	Skin biopsy	Freezing	20% DMSO, 20% EG	<i>In vitro</i> culture	Cryobanking	Caamaño et al. (2008)
<i>L. pardinus</i>	Skin biopsy, muscle, oral mucosa, bone marrow, spine marrow, intestines	Freezing	NI	NI	Cryobanking	Léon-Quinto et al. (2009)
<i>S. scrofa</i>	Skin biopsy	Freezing	None	<i>In vitro</i> culture and production of clone embryos	SCNT	Zhang et al. (2012)
<i>U. arctos</i>	Skin biopsy	Vitrification	10% DMSO, 10% FBS	<i>In vitro</i> culture	Cryobanking	Caamaño et al. (2008)
<i>P. tajacu</i>	Skin biopsy	Vitrification	20% DMSO, 20% EG, 0.25 M sucrose, 10% FBS	Histological analysis and <i>in vitro</i> culture	Cryobanking	Borges et al. (2015a,b)
<i>D. leporina</i>	Skin biopsy	Vitrification	20% DMSO, 20%, 0.25 M sucrose, 10% FBS	Histological analysis	Cryobanking	Costa et al. (2015)

Abbreviations: DMSO: dimethyl sulfoxide, EG: ethylene glycol, FBS: fetal bovine serum, SCNT: somatic cell nuclear transfer; NI: not informed.

^aMore information about the evaluation see reference.

In the protocol used for conventional vitrification, fragments are initially incubated in 1.0 mL DMEM as vitrification medium supplemented with 20% DMSO, 20% EG and 20% FBS and placed in cryovials. Samples were equilibrated for 10 min at room temperature, and finally plunged and stored in LN₂. The posterior warming of the cryovials was conducted at 38°C for 2 min, followed by the removal of vitrification medium and mix with 1.0 mL DMEM plus 10% FBS.

Regarding the SSV technique (Borges et al., 2015a,b), fragments are exposure to DMEM as vitrification medium supplemented with 20% DMSO, 20% EG, 0.25 M sucrose and 10% FBS) in Petri dishes for 5 min. The excess of CPAs is removed, and the samples are deposited on a metal cubic surface partially placed in LN₂. Then, samples are transferred to cryovials and immersed directly in LN₂. Finally, warming of the cryovials is conducted at 37°C for 1 min and the vitrification medium is removed through washings in DMEM supplemented sucrose at 0.50, 0.25 M and no sucrose for 5 min.

In general, the somatic tissue cryopreservation is a strategy for genetic conservation of wild animals (Benkeddache et al., 2012). For this purpose, the vitrification strategies are currently considered more efficient than conventional cryopreservation in domestic species (Brockbank et al., 2010). Although the development of vitrification has greatly simplified cryopreservation procedures, the protocols already established for other types of tissue may not be ideal to be applied for the tissue of the target species (Da-Croce et al., 2013). In the vitrification, the solution is rapidly cooled and transformed into a glassy (Amorim et al., 2011). The choice for the vitrification is due to the time consumed to perform the technique, the more economical and practicality to be performed in any laboratory (Ting et al., 2013). Moreover, for optimum conditions, a small volume of the vitrification solution is required to be in contact to the tissue cryopreserved. Thus, the SSV provides the use of a small amount of a cryoprotectant consisting of direct exposure of the tissue to a solid surface pre cooled (Carvalho et al., 2011).

In a previous study, we demonstrated that the SSV method was more able to preserve somatic tissue of *P. tajacu* (Borges et al., 2015a,b) and *D. leporina* (Costa et al., 2015) than the conventional vitrification. In these works, we showed that the SSV was the most appropriate method for the vitrification of somatic tissue and the recovery of viable somatic cells that could be destined to cloning.

On the other hand, a comparative study of two systems, freezing and conventional vitrification, was conducted in *U. arctos* skin biopsies (Caamaño et al., 2008). The authors observed that skin fragments could be preserved for long term, allowing fibroblasts to proliferate in culture. The vitrification led to poorer cell proliferation and more days to reach 70-80% confluence than freezing (19.2 versus 1.25 days). Moreover, the authors affirmed that skin vitrification needs further studies and refinement prior to its use on the field. Additionally, the vitrification has facilities as use only cryovials, vitrification solution and a small nitrogen tank.

As a criteria for the successful preservation of somatic tissue samples, the use of *in vitro* culture derived from samples is an essential tool for the identifying the cryopreservation damages. Thus, tissues fragments, after primary culture, can promote the detachment of the cells, as fibroblasts (Caamaño et al., 2008; León-Quinto et al., 2009; Borges et al., 2015b). All these cells can be assessed for quality, viability, proliferation and functional activity of the cultured cells. Additionally, the isolation of proliferating cells derived from cryopreserved tissues and the culture are essential analyzes for the establishment of the cryobanking (Léon-

Quinto et al., 2009), particularly for cloning purposes, which they can be used as karyoplasts that is nucleus donor.

Finally, some of cryopreservation artifacts can be observed by histological techniques. In general, the histological evaluation using hematoxylin-eosin provides a basic requirement for knowledge of the structures that were affected or maintained morphologically normal during cryopreservation. It enables the visualization of the tissue as both a whole as the analysis of their constituents, epidermis and dermis (Borges et al., 2015a). Other more specific techniques, as protein argyrophil related to nucleolar organizer regions allow for the recognition of cell proliferative activity (Cresta & Alves, 2007).

4.4. Use of Somatic Tissue in the Conservation Programs

Most somatic tissue banks are established from the skin of wild animals. In general, it is known that the ease of collection of the material and advances in cloning area allowed increasingly proper storage of somatic tissues. In this sense, several cases illustrate the importance of somatic cryobanking in wild mammals. Moreover, the same examples can be cited with the use of somatic cells.

In 2009, León-Quinto et al. established various biologic resource banks as a supporting tool for reproduction and conservation of *L. pardinus*. This species is considered the most endangered felid in the world, being confined in southern Spain and Sierra Morena (Simon et al., 2012). Thus, León-Quinto et al. showed that the somatic samples savage reflected on a very important fraction of the population biodiversity which will allow the development of a wide variety of works that can be easily extrapolated to the majority of the population. A total of 7 males, 6 females and 69 different individuals were collected for testicular, ovarian and somatic tissues, respectively (León-Quinto et al., 2009).

In 2010, Guan et al. established a fibroblast cell line derived from bengal tiger (*P. tigris tigris*). The tiger is warranted the highest level of protection by the Convention on International Trade in Endangered Species of Wild Fauna and Flora and, in 1989, the Chinese government brought *P. tigris tigris* into first category of National Protected Animal breed (Guan et al., 2010). Thun, in this study, the authors showed that also somatic samples can be cryopreserved and maintained in culture.

An interesting example of importance of somatic banks and cloning use can be evidenced with the study of Folch et al. (2009). The authors produced the first clone of *Capra pyrenaica pyrenaica*, an animal derived from as extinct caprine subspecies. Thus, the karyoplast used consisted on fibroblast derived from skin biopsies, obtained and cryopreserved in 1999 from the last living female, that died in 2000; the cytoplasm were matured oocytes derived from domestic goats. Thus, in 2009, one morphologically normal bucardo was produced. Unfortunately, newborn died some after birth. This study was the first using an extinct animal. Other previous works using cytoplasm of domestic animals with wild karyoplast resulting on live offspring were demonstrated for *B. gaurus* (Lanza et al., 2000) and *Ovis orientalis musimon* (Loi et al., 2001).

Finally, the somatic banks also can be employed for the production of induced cells to pluripotency (iPS or induced pluripotent stem). These cells can be obtained from fibroblasts using a culture system of defined factors (Takahashi & Yamanaka, 2006). In *Panthera uncial* (Verma et al., 2012) and *Mandrillus leucophaeus* (Ben-Nun et al., 2011), somatic cells already

were induced to pluripotency. The use of this tool would be interesting to allow an alternative way to produce gametes in endangered wild mammals.

5. GENERAL CONSIDERATIONS

The biodiversity conservation by tissue cryopreservation is a promising tool with different purposes, and can be applied both for the multiplication of a species as to a basic investigative study. In general, with regards to the conservation programs, there is a variation according to the species and the tissue; nevertheless, clearly there was progress in wild felids, especially for the Iberian lynx. Finally, improvements are still needed as the practical establishment of biological recourse banks in various species and their application in reproductive techniques.

REFERENCES

- Abbasi, S., & Honaramooz, A. (2011). Xenografting of testis tissue from bison calf donors into recipient mice as a strategy for salvaging genetic material. *Theriogenology*, 76, 607-14.
- Abbasi, S., & Honaramooz, A. (2012). Feasibility of salvaging genetic potential of post-mortem fawns: production of sperm in testis tissue xenografts from immature donor white-tailed deer (*Odocoileus virginianus*) in recipient mice. *Anim Reprod Sci*, 135, 47-52.
- Abrishami, M., Anzar, M., Yang, Y., & Honaramooz, A. (2010). Cryopreservation of immature porcine testis tissue to maintain its developmental potential after xenografting into recipient mice. *Theriogenology*, 73, 86-96.
- Aerts, J.M., & Bols, P.E. (2010). Ovarian follicular dynamics: a review with emphasis on the bovine species. Part I: folliculogenesis and pre-antral follicle development. *Reprod Domest Anim*, 45, 171-179.
- Aman, R.R., & Parks, J.E. (1994). Effects of cooling and rewarming on the meiotic spindle and chromosomes of *in vitro*-matured bovine oocytes. *Biol Reprod*, 50, 103-110.
- Amorim, C.A., Curaba, M., Van Langendonckt, A., Dolmans, M.M., & Donnez, J. (2011). Vitricification as an alternative means of cryopreserving ovarian tissue. *Reprod Biomed Online*, 23, 160-186.
- Amorim, C.A., Dolmans, M.M., David, A., Jaeger, J., Vanacker, J., Camboni, A., Donnez, J., & Van Langendonckt, A. (2012). Vitricification and xenografting of human ovarian tissue. *Fertil Steril*, 98, 1291-1298.
- Andrabi, S.M.H., & Maxwell, W.M.C. (2007). A review on reproductive biotechnologies for conservation of endangered mammalian species. *Anim Reprod Sci*, 99, 223-243.
- Aponte, P.M., Rooij, D.G., & Bastidas, P. (2005). Testicular development in Brahman bulls. *Theriogenology*, 64, 1440-1455.
- Arregui, L., & Dobrinski, I. (2014). Xenografting of testicular tissue pieces: 12 years of an *in vivo* spermatogenesis system. *Reproduction*, 148, 71-84.
- Assis Neto, A.C., Carvalho, M.A.M., Melo, M.I.V., Miglino, M.A., Oliveira, M.F., & Mariana, A.N.B. (2003). Fases do desenvolvimento e diferenciação testicular em cutias (*Dasyprocta aguti*) criadas em cativeiros. [Phases of the development and testicular differentiation in agoutis (*Dasyprocta aguti*) raised in captivity]. *Braz J Vet Res Ani Sci*, 40, 71-79.

- Benkeddache, D., Bodinier, P., Joly, T., Berchiche, M., & Vignon, X. (2012). Recovery of viable cells from rabbit skin biopsies after storage at -20°C for up to 10 days. *Cell Tissue Bank*, 13, 479-486.
- Ben-Nun, I.F., Montague, S.C., Houck, M.L., Tran, H.T., Garitaonandia, I., Leonardo, Y.C., Wang, Y.C., Charter, S.J., Laurent, L.C., Ryder, O.A., & Loring, J.F. (2011). Induced pluripotent stem cells from highly endangered species. *Nat Methods*, 8, 829-836.
- Berg, D.K., Li, C., Asher, G., Wells, D.N., & Oback, B. (2007). Red deer cloned from antler stem cells and their differentiated progeny. *Biol Reprod*, 77, 384-394.
- Boiso, I., Martí, M., Santaló, J., Ponsá, M., Barri, P.N., & Veiga, A. (2002). A confocal microscopy analysis of the spindle and chromosome configurations of human oocytes cryopreserved at the germinal vesicle and metaphase II stage. *Hum Reprod*, 17, 1885-1891.
- Borges, A.A., Lima, G.L., Santos, M.L.T., Queiroz Neta, L.B., Praxedes, E.C.G., Santos, M.V.O., Oliveira, M.F., Silva, A.R., & Pereira, A.F. (2015a). Histological evaluation of ear tissue of collared peccary (*Pecari tajacu*) after different vitrification techniques. *Anim Reprod*, 12, 230.
- Borges, A.A., Queiroz Neta, L.Q., Santos, M.V.O., Santos, M.L.T., Lima, G.L., Oliveira, M.F., Silva, A.R., & Pereira, A.F. (2015b). Isolation of somatic cell derived from ear tissue of collared peccary (*Pecari tajacu* Linnaeus, 1758) submitted to different vitrification techniques. *Anim Reprod*, 12, 842.
- Borges, P.A.C., Lima, G.L., Bezerra, J.A.B., Souza, A.L.P., Peixoto, G.C.X., Praxedes, E.C.G., & Silva, A.R. (2015c). Cryopreservation of testicular tissue: a potential tool for the conservation of male genetic material from collared peccaries (*Pecari tajacu*). *Anim Reprod*, 12, 272.
- Brockbank, K.G., Chen, Z.Z., & Song, Y.C. (2010). Vitrification of porcine articular cartilage. *Cryobiology*, 60, 217-221.
- Caamaño, J.N., Rodriguez, A., Muñoz, M., De Frutos, C., Diez, C., & Gómez, E. (2008). Cryopreservation of brown bear skin biopsies. *Cell Preserv Technol*, 6, 83-86.
- Campos-Junior, P.H.A., Costa, G.M.J., Avelar, G.F., Lacerda, S.M.S.N., Costa, N.N., Orashi, O.T., Miranda, M.S., Barcelos, L.S., Jorge, E.C., Guimarães, D.A., & França, L.R. (2014). Derivation of sperm from xenografted testis cells and tissues of the peccary (*Tayassu tajacu*). *Reproduction*, 147, 291-299.
- Candy, C.J., Wood, M.J., & Whittingham, D.G., (1995). Follicular development in cryopreserved marmoset ovarian tissue after transplantation. *Hum Reprod*, 10, 2334-2338.
- Carvalho, A.A., Faustino, L.R., Figueiredo, J.R., Rodrigues, A.P.R., & Costa, A.P.R. (2012). Vitricificação: uma alternativa para a preservação de embriões e material genético de fêmeas mamíferas em criobancos [Vitrification: an alternative for preserving embryos and genetic material mammalian females in cryobanking]. *Acta Vet Brasilica*, 5, 236-248.
- Carvalho, A.A., Faustino, L.R., Silva, C.M.G., Castro, S.V., Luz, H.K.M., Rossetto, R., Lopes, C.A.P., Figueiredo, J.R., Rodrigues, A.P.R., & Costa, A.P.R. (2011). Influence of vitrification techniques and solutions on themorphology and survival of preantral follicles after in vitro culture of caprine ovarian tissue. *Theriogenology*, 76, 933-941.
- Cetinkaya, G., & Arat, S. (2011). Cryopreservation of cartilage cell and tissue for biobanking. *Cryobiology*, 63, 292-297.
- Chung, K., Donnez, J., Ginsburg, E., & Meirow, D. (2013). Emergency IVF versus ovarian tissue cryopreservation: decision making in fertility preservation for female cancer patients. *Fertil Steril*, 99, 1534-1542.

- Cleary, M., Paris, M.C., Shaw, J., Jenkin, G., & Trounson, A.O. (2003). Effect of ovariectomy and graft position on cryopreserved common wombat (*Vombatus ursinus*) ovarian tissue following xenografting to nude mice. *Reprod Fertil Dev*, 15, 333-342.
- Cleary, M., Shaw, J.M., Jenkin, G., & Trounson, A.O. (2004). Influence of hormone environment and donor age on cryopreserved common wombat (*Vombatus ursinus*) ovarian tissue xenografted into nude mice. *Reprod Fertil Dev*, 16, 699-707.
- Cleary, M., Snow, M., Paris, M., Shaw, J., Cox, S.L., & Jenkin, G. (2001). Cryopreservation of mouse ovarian tissue following prolonged exposure to an ischemic environment. *Cryobiology*, 42, 121-133.
- Comizzoli, P., Songsasen, N., & Wildt, D.E. (2010). Protecting and Extending Fertility for Females of Wild and Endangered Mammals. *Cancer Treat Rev*, 156, 87-100.
- Comizzoli, P., & Wildt, D.E., (2012). On the horizon for fertility preservation in domestic and wild carnivores. *Reprod Domest Anim*, 47, 261-265.
- Costa, C.A.S., Borges, A.A., Santos, M.L.T., Queiroz Neta, L.B., Santos, M.V.O., França, P.H.F., Oliveira, M.F., Silva, A.R., & Pereira, A.F. (2015). Aplicação da vitrificação para fins de conservação de tecido somático de cutias (*Dasyprocta leporina*). [Application of vitrification for conservation purposes of somatic tissue of agoutis (*Dasyprocta leporina*)]. *Proceedings of the XXI Scientific Initiation Meeting, SEMIC, Mossoró/RN*.
- Cresta, F.B., & Alves, M.R. (2007). Evaluation of the corneal epithelium kinetics using cell proliferation markers. *Arq Bras Oftalmol*, 70, 953-960.
- Da-Croce, L., Gambarini-Paiva, G.H.R., Angelo, P.C., Bambirra, E.A., Cabral, A.C.V., & Godard, A.L.B. (2013). Comparison of vitrification and slow cooling for umbilical tissues. *Cell Tissue Bank*, 14, 65-76.
- Donnez, J., Dolmans, M.M., Diaz, C., & Pellicer, A. (2015). Ovarian cortex transplantation: time to move on from experimental studies to open clinical application. *Fertil Steril*, 104, 1097-1098.
- Ehrlich, L.E., Feig, J.S., Schiffres, S.N., Malen, J.A., & Rabin, Y. Large thermal conductivity differences between the crystalline and vitrified states of DMSO with applications to cryopreservation. *PloS One*, e0125862.
- Fahy, G.M. (2010). Cryoprotectant toxicity neutralization. *Cryobiology*, 60, 45-53.
- Faustino, L.R., Silva, C.M.G., Rossetto, R., Rodrigues, G.Q., Figueiredo, J.R., & Rodrigues, A.P.R. (2011). Estado atual e desafios da criopreservação de tecido ovariano em mamíferos. [Current status and challenges of cryopreservation of ovarian tissue in mammals]. *Rev Bras Reprod Anim*, 35, 3-15.
- Figueiredo, J.R., Rodrigues, A.P.R., Amorim, C.A., & Silva, J.R.V. (2008). Manipulação de oócitos inclusos em folículos ovarianos pré antrais. [Oocyte manipulation included in pre antral follicles]. In: *Biotécnicas Aplicadas à Reprodução Animal*, [Biotechniques Applied to Animal Reproduction] 303-327. Eds Golçalves, P.B.D. Figueiredo, J.R. & Freitas, V.J.F. São Paulo: Rocca.
- Filatov, M.A., Khranova, Y.V., Kiseleva, M.V., Malinova, I.V., Komarova, E.V., & Semenova, M.L. (2016). Female fertility preservation strategies: cryopreservation and ovarian tissue *in vitro* culture, current state of the art and future perspectives. *Zygote*, 1-19.
- Folch, J., Cocero, M.J., Chesné, P., Alabart, J.L., Domínguez, V. Cognié, Y., & Vignon, X. (2009). First birth of an animal from an extinct subspecies (*Capra pyrenaica pyrenaica*) by cloning. *Theriogenology*, 71, 1026-1034.

- Gandolfi, F., Paffoni, A., Brambilla, E.P., Bonetti, S., Brevini, T.A., & Ragni, G. (2006). Efficiency of equilibrium cooling and vitrification procedures for the cryopreservation of ovarian tissue: comparative analysis between human and animal models. *Fertil Steril*, 85, 1150-1156.
- Gourdon, J.C., & Travis, A.J. (2012). Spermatogenesis in ferret testis xenografts: a new model. *Comp Med*, 61, 145-149.
- Guan, W.J., Liu, C.Q., Li, C.Y., Liu, D., Zhang, W.X., & Ma, Y.H. (2010). Establishment and cryopreservation of a fibroblast cell line derived from Bengal tiger (*Panthera tigris tigris*). *CryoLetters*, 31, 130-138.
- Gunasena, K.T., Lakey, J.R.T., Villines, P.M., Bush, M., Raath, C., Critser, E.S., McGann, L.E., & Critser, J.K. (1998). Antral follicles develop in xenografted cryopreserved african elephant (*Loxodonta Africana*) ovarian tissue. *Anim Reprod Sci*, 53, 265-275.
- Hashimoto, S., Suzuki, N., Yamanaka, M., Hosoi, Y., Ishizuka, B., & Morimoto, Y. (2010). Effects of vitrification solutions and equilibration times on the morphology of cynomolgus ovarian tissues. *Reprod Bio Med Online*, 21, 501-509.
- Hovatta, O. (2005). Methods for cryopreservation of human ovarian tissue. *Reprod Biomed Online*, 10, 729-734.
- Hovatta, O. (2000). Cryopreservation of testicular tissue. *Mol Cell Endocrinol*, 169, 113-115.
- Jensen, A.K., Kristensen, S.G., Macklon, K.T., Jeppesen, J.V., Fedder, J., Ernst, E., & Andersen, C.Y. (2015). Outcomes of *transplantations* of cryopreserved ovarian tissue to 41 women in Denmark. *Hum Reprod*, 30, 2838-2845.
- Jewgenow, K., Wiedemann, C., Bertelsen, M.F., & Ringleb, J. (2011). Cryopreservation of mammalian ovaries and oocytes. *Int Zoo Yearb*, 45, 124-132.
- Kaneko, H., Kikuchi, K., Nakai, M., Somfai, T., Noguchi, J., Tanihara, F., Ito, J., & Kashiwazaki, N. (2013). Generation of live piglets for the first time using sperm retrieved from immature testicular tissue cryopreserved and grafted into nude mice. *PLoS One*, 8, e70989.
- Kikuchi, K., Nakai, M., Kashiwazaki, N., & Kaneko, H. (2011). Xenografting of gonadal tissues into mice as a possible method for conservation and utilization of porcine genetic resources. *Anim Sci J*, 82, 495-503.
- Lanza, R.P., Cibelli, J.B., Diaz, F., Moraes, C.T., Farin, P.W., Farin, C.E., Hammer, C.J., West, M.D., & Damiani, P. (2000). Cloning of an endangered species (*Bos gaurus*) using interspecies nuclear transfer. *Cloning*, 2, 79-90.
- Lee, D.M., Yeoman, R.R., Battaglia, D.E., Stouffer, R.L., Zelinski-Wooten, M.B., Fanton, J.M., & Wolf, D.P. (2004). Live birth after ovarian tissue transplant. *Nature*, 428, 137-138.
- Lee, K.H., Lee, W.Y., Kim, D.H., Lee, S.H., Do, J.T., Park, C., Kim, J.H., Choi, Y.S., & Song, H. (2016). Vitrified canine testicular cells allow the formation of spermatogonial stem cells and seminiferous tubules following their xenotransplantation into nude mice. *Sci Rep*, 24, 21919.
- León-Quinto, T., Simon, M.A., Cadenas, R., Jones, J., Martinez-Hernandez, F.J., Morenod, J.M., Vargas, A., Martinez, F., & Soria, B. (2009). Developing biological resource banks as a supporting tool for wildlife reproduction and conservation: The *Iberian lynx* bank as a model for other endangered species. *Anim Reprod Sci*, 112, 347-361.
- León-Quinto, T., Simón, A., Cadenas, R., Martínez, A., & Serna, A. (2014). Different cryopreservation requirements in foetal *versus* adult skin cells from an endangered mammal, the Iberian lynx (*Lynx pardinus*). *Cryobiology*, 68, 227-233.

- León-Quinto, T., Simón, M.A., Sánchez, Á., Martín, F., & Soria, B. (2011). Cryobanking the genetic diversity in the critically endangered Iberian lynx (*Lynx pardinus*) from skin biopsies. Investigating the cryopreservation and culture ability of highly valuable explants and cells. *Cryobiology*, 62, 145-151.
- Lima, G.L., Luz, V.B., Alves, A.M.C.V., & Silva, A.R. (2012). Vitrification of collared peccaries (*Tayassu tajacu*) ovarian tissue using various cryoprotectants. *Anim Reprod*, 9, 261.
- Lin, T., Yen, J., Kuo, T., Gong, K., Hsu, K., & Hsu, T. (2008). Comparison of the developmental potential of 2-week-old preantral follicles derived from vitrified ovarian tissue slices, vitrified whole ovaries and vitrified/transplanted newborn mouse ovaries using the metal surface method. *BMC Biotechnol*, 8, 38-50.
- Liu, J., Cheng, K.M., Purdy, P.H., & Silversides, F.G. (2012). A simple vitrification method for cryobanking avian testicular tissue. *Poult Sci*, 91, 3209-13.
- Loi, P., Ptak, G., Barboni, B., Fulka, J. Jr., Cappai, P., & Clinton, M. (2001). Genetic rescue of an endangered mammal by cross-species nuclear transfer using *post-mortem* somatic cells. *Nat Biotechnol*, 19, 962-964.
- Lopes, C.A., Alves, A.M., Jewgenow, K., Bão, S.N., & de Figueiredo, J.R. (2016). Cryopreservation of canine ovarian cortex using DMSO or 1,3-propanediol. *Theriogenology*, Epub ahead of print.
- Lu, X.L., Yu, J., Zhang, G., Wei, Z.T., Li, J.T., & Zhang, J.M. (2014). Effects of varying tissue sizes on the efficiency of baboon ovarian tissue Vitrification. *Cryobiology*, 69, 79-83.
- Lunardi, F.O., Aguiar, F.L.N., Duarte, A.B.G., Araújo, V.R., Lima, L.F., Sá, N.A.R., Correia, H.H.V., Domingues, S.F.S., Campello, C.C., Smits, J., de Figueiredo, J.R., & Rodrigues, A.P.R. (2016). Ovine secondary follicles vitrified out the ovarian tissue grow and develop *in vitro* better than those vitrified into the ovarian fragments. *Theriogenology*, 85, 1203-1210.
- Machado, L.C., Oliveira, V.C., Paraventi, M.D., Cardoso, R.N.R., Martins, D.S., & Ambrósio, C.E. (2016). Maintenance of brazilian biodiversity by germplasm bank. *Pesq Vet Bras*, 36, 62-66.
- Martino, A., Pollard, J.W., & Leibom S. (1996). Effect of chilling bovine oocytes on their developmental competence. *Mol Reprod Dev*, 45, 503-512.
- Mahesh, Y.U., Rao, B.S., Katari, V.C., Komjeti, S., Christo, D., Lakshmikantan, U., Pawar, R.M., & Shivaji, S. (2012). Cell cycle synchronization of bison (*Bos gaurus*) fibroblasts derived from ear piece collected post-mortem. *Reprod Dom Animal*, 47, 799-805.
- Mattiske, D., Shaw, G., & Shaw, J.M. (2002). Influence of donor age on development of gonadal tissue from pouch young of the tammar wallaby, *Macropus eugenii*, after cryopreservation and xenografting into mice. *Reproduction*, 123, 143-153.
- Meirow, D., Ra'anani, H., Shapira, M., Brenghausen, M., Derech Chaim, S., Aviel-Ronen, S., Amariglio, N., Schiff, E., Orvieto, R., & Dor, J. (2016). Transplantations of frozen thawed ovarian tissue demonstrate high reproductive performance and the need to revise restrictive criteria. *Fertil Steril*, Epub ahead of print.
- Milazzo, J.P., Vaudreuil, L., Cauliez, B., Gruel, E., Masse, L., Mousset-Sime, N., Mace, B., & Rives, N. (2008). Comparison of conditions for cryopreservation of testicular tissue from immature mice. *Hum Reprod*, 23, 17-28.

- Mota, P.C., Ehmcke, J., Westernströer, B., Gasseic, K., Ramalho-Santosa, L., & Schlat, S., (2012). Effects of different storage protocols on cat testis tissue potential for xenografting and recovery of spermatogenesis. *Theriogenology*, 77, 299-310.
- Motohashi, H.H., & Ishibashi, H. (2016). Cryopreservation of ovaries from neonatal marmoset monkeys. *Exp Anim*, Epub ahead of print.
- Mukaida, T., & Oka, C., (2012). Vitrification of oocytes, embryos and blastocysts. *Best. Pract Res Clin Obstet Gynaecol*, 26, 789-803.
- Mycock, D.J., Wesley-Smith, J., & Berjak P. (1995). Cryopreservation of somatic embryos of four species with and without cryoprotectant pre-treatment. *Annals of Botany*, 75, 331-336.
- Nagy, Z.P., Chang, C.C., Shapiro, D.B., Bernal, D.P., Elsner, C.W., Mitchell-Leef, D., Toledo, A.A., & Kort, H.I. (2009). Clinical evaluation of the efficiency of an oocyte donation program using egg cryobanking. *Fertil Steril*, 92, 520-526.
- Ning, L., Goossens, E., Geens, M., Van Saen, D., & Tournaye, H. (2012). Spermatogonial stem cells as a source for regenerative medicine. *Middle East Fertil Soc J.*, 17, 1-7.
- Norton, B.G. (2014). *The preservation of species*. New Jersey: Princeton University Press.
- Nyachio, A., Spiessens, C., Chai, D.C., Kiulia, N.M., Willemsen, D., Mwenda, J.M., Bourgain, C., & D'Hooghe, T.M. (2013). Ovarian tissue cryopreservation by Vitrification in olive baboons (*Papio anubis*): a pilot study. *Gynecol Obstet Investig*, 75, 157-162.
- Oh, H.J., Kim, M.K., Jang, G., Kim, H.J., Hong, S.G., Park, J.E., Park, K., Park, C., Sohn, S.H., Kim, D.Y., Shin, N.S., & Lee, B.C. (2008). Cloning endangered gray wolves (*Canis lupus*) from somatic cells collected *postmortem*. *Theriogenology*, 70, 638-647.
- Paris, M.C.J., Snow, M., Cox, S.L., & Shaw, J.M. (2004). Xenotransplantation: a tool for reproductive biology and animal conservation? *Theriogenology*, 61, 277-291.
- Paynter, S. (2005). A rational approach to oocyte cryopreservation. *Reprod Biomed Online*, 10, 578-586.
- Poels, J., Van Langendonck, A., Dehoux, J.P., Donneze, J., & Wyns C. (2012). Vitrification of non-human primate immature testicular tissue allows maintenance of proliferating spermatogonial cells after xenografting to recipient mice. *Theriogenology*, 77, 1008-1013.
- Pothana, L., Makala, H., Devi, L., Varma, V.P., & Goel, S. (2015). Germ cell differentiation in cryopreserved, immature, Indian spotted mouse deer (*Moschiola indica*) testes xenografted onto mice. *Theriogenology*, 83, 625-633.
- Practice Committee of American Society for Reproductive Medicine. (2014). Ovarian tissue cryopreservation: a committee opinion. *Fertil Steril*, 101, 1237-1243.
- Praxedes, E.C.G., Lima, G.L., Bezerra, J.A.B., Souza, A.L.P., Oliveira, M.F., & Silva, A.R., (2014). Vitrification of cavies (*Galea spixii spixii*) ovarian tissue – preliminar results. *Acta Vet Brasilica*, 8, 1-4.
- Pukazhenth, B.S., Nagashima, J., Travis, A.J., Costa, G.M., Escobar, E.N., França, L.R., & Wildt, D.E. (2015). Slow freezing, but not vitrification supports complete spermatogenesis in cryopreserved, neonatal sheep testicular xenografts. *Plos One*, 10, e01239.
- Russell, L.D., Ettlin, R.A., Hikim, A.P.S., & Clegg, E.D. (1990). *Histological and histopathological evaluation of the testis*. Cache River Press, Clearwater, Florida.
- Salle, B., Demirci, B., Franck, M., Rudigoz, R.F., Guerin, J.F., & Lornage, J. (2002). Normal pregnancies and live births after autograft of frozen-thawed hemiovaries into ewes. *Fertil Steril*, 77, 403-408.

- Sansinena, M., Santos, M.V., Zaritzky, N., Baeza, R., & Chirife, J. (2010). Theoretical prediction of the effect of heat transfer parameters on cooling rates of liquid filled plastic straws used for cryopreservation of spermatozoa. *Cryoletters*, 31, 120-129.
- Santos, M.V., Sansinena, M., Zaritzky, N., & Chirife, J. (2012). Assessment of external heat transfer coefficient during oocyte vitrification in liquid and slush nitrogen using numerical simulations to determine cooling rates. *Cryoletters*, 33, 31-40.
- Santos, P.R.S., Oliveira, M.F., Silva, A.R., & Assis Neto, A.C. (2012). Development of spermatogenesis in captive-bred Spix's yellow-toothed cavy (*Galea spixii*). *Reprod Fert Develop*, 24, 877-878.
- Santos, R.R., Amorim, C., Cecconi, S., Fassbender, M., Imhof, M., Lornage, J., Paris, M., Schoenfeldt, V., & Martinez-Madrid, B. (2010). Cryopreservation of ovarian tissue: An emerging technology for female germline preservation of endangered species and breeds. *Anim Reprod Sci*, 122, 151-163.
- Santos, R.R., Celestino, J.J.H., Lopes, C.A.P., Melo, M.A.P., Rodrigues, A.P.R., & Figueiredo, J.R. (2008). Criopreservação de folículos ovarianos pré-antrais de animais domésticos. [Cryopreservation of ovarian preantral follicles from domestic animals]. *Rev Bras Reprod Anim*, 32, 9-15.
- Saragusty, J., & Arav, A. (2011). Current progress in oocyte and embryo cryopreservation by slow freezing and vitrification. *Reproduction*, 141, 1-19.
- Sato, T., Katagiri, K., Yokonishi, T., Kubota, Y., Inoue, K., Ogonuki, N., Matoba, O., Ogura, A., & Ogawa, T. (2011). *In vitro* production of functional sperm in cultured neonatal mouse testes. *Nature*, 471, 504-507.
- Schlatt, S., Samuel Kim, S., & Gosden, R. (2002). Spermatogenesis and steroidogenesis in mouse, hamster and monkey testicular tissue after cryopreservation and heterotopic grafting to castrated hosts. *Reproduction*, 124, 339-346.
- Schnorr, J., Oehninger, S., Toner, J., Hsiu, J., Lanzendorf, S., Williams, R., & Hodgen, G. (2002). Functional studies of subcutaneous ovarian transplants in non-human primates: steroidogenesis, endometrial development, ovulation, menstrual patterns and gamete morphology *Hum Reprod*, 17, 612-619.
- Silva, A.R., Lima, G.L., Peixoto, G.C.X., & Souza, A.L.P. (2015). Cryopreservation in mammalian conservation biology: current applications and potential utility. *Res Reports Biodiversity Studies*, 4, 1-8.
- Silva, A.R., Morato, R.G., Silva, & L.D.M. (2004). The potential for gamete recovery from non-domestic canids and felids. *Anim Reprod Sci*, 81, 159-175.
- Silvestre, M.A., Saeed, A.M., Cervera, R.P., Escribá, M.J., & García-Ximénez, F. (2003). Rabbit and pig ear skin sample cryobanking: effects of storage time and temperature of the whole ear extirpated immediately after death. *Theriogenology*, 59, 1469-1477.
- Simon, M.A., Gil-Sánchez, J.M., Ruiz, G., Garrote, G., McCain, E.B., Fernández, L., López-Parra, M., Rojas, E., Arenas-Rojas, R., Del Rey, T., García-Tardío, M., & López, G. (2012). Reverse of the decline of the endangered Iberian lynx. *Conserv Biol*, 26, 731-736.
- Singh, M., & Ma, X. (2014). *In vitro* culture of fibroblast-like cells from sheep ear skin stored at 25-26°C for 10 days after animal death. *International J Biol*, 6, 96-102.
- Song, J.S., Hua, S., & Zhang, Y. (2007). Culture, characteristics and chromosome complement of Siberian tiger fibroblasts for nuclear transfer. *In Vitro Cell Dev Biol Anim*, 43, 203-209.
- Stachecki, J.J., Munne, S. & Cohen, J. (2004). Spindle organization after cryopreservation of mouse, human, and bovine oocytes. *Reprod Biomed Online*, 8, 664-672.

- Suzuki, N., Hashimoto, S., Igarashi, S., Takae, S., Yamanaka, M., Yamochi, T., Takenoshita, M., Hosoi, Y., Morimoto, Y., & Ishizuka, B. (2012). Assessment of long-term function of heterotopic transplants of vitrified ovarian tissue in cynomolgus monkeys. *Hum Reprod*, 27, 2420-2429.
- Svingen, T., & Koopman, P. (2013). Building the mammalian testis: origins, differentiation, and assembly of the component cell populations. *Genes & Dev.*, 27, 2409-2426.
- Sztejn, J.M., Noble, K., Farley, J.S., & Mobraaten, L.E. (2001). Comparison of permeating and nonpermeating cryoprotectants for mouse sperm cryopreservation. *Cryobiology*, 482, 28-39.
- Tahaei, L.S., Eimani, H., Hajmusa, G., Fathi, R., Rezazadeh Valojerdi, M., Shahverdi, A., & Eftekhari-Yazdi, P. (2015). Development of xenotransplanted sheep ovarian tissue into male and female immunodeficient rats. *Int J Fertil Steril*, 9, 354-360.
- Takahashi, K., & Yamanaka, S. (2006). Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors. *Cell*, 126, 663-676.
- Talevi, R., Barbato, V., Fiorentino, I., Braun, S., De Stefano, C., Ferraro, R., Sudhakaran, S., & Gualtieri, R. (2016). Successful slush nitrogen vitrification of human ovarian tissue. *Fertil Steril*, 105, 1523-1531.
- Tanpradit, N., Comizzoli, P., Srisuwatanasagul, S., & Chatdarong, K. (2015). Positive impact of sucrose supplementation during slow freezing of cat ovarian tissues on cellular viability, follicle morphology, and DNA integrity. *Theriogenology*, 83, 1553-1561.
- Thuwanut, P., Srisuwatanasagul, S., Wongbandue, G., Tanpradit, N., Thongpakdee, A., Tongthainan, D., Manee-in, S., & Chatdarong, K. (2013). Sperm quality and the morphology of cryopreserved testicular tissues recovered post-mortem from diverse wild species. *Cryobiology*, 67, 244-247.
- Ting, A.Y., Yeoman, R.R., Campos, J.R., Lawson, M.S., Mullen, S.F., Fahy, G.M., & Zelinski, M.B. (2013). Morphological and functional preservation of pre-antral follicles after vitrification of macaque ovarian tissue in a closed system. *Hum Reprod*, 28, 1267-1279.
- Ting, A.Y., Yeoman, R.R., Lawson, M.S., & Zelinski, M.B. (2011). *In vitro* development of secondary follicles from cryopreserved rhesus macaque ovarian tissue after slow-rate freeze or vitrification. *Hum Reprod*, 26, 2461-2472.
- Ting, A.Y., Yeoman, R.R., Lawson, M.S., & Zelinski, M.B. (2012). Synthetic polymers improve vitrification outcomes of macaque ovarian tissue as assessed by histological integrity and the *in vitro* development of secondary follicles. *Cryobiology*, 65, 1-11.
- Tovar, H., Navarrete, F., Rodríguez, L., Skewes, O., & Castro, F.O. (2008). Cold storage of biopsies from wild endangered native Chilean species in field conditions and subsequent isolation of primary culture cell lines. *In Vitro Cell Dev Biol Anim*, 44, 309-320.
- Tucker, M.J., Morton, P.C., Wright, G., Sweitzer, C.L., & Massey, J.B. (1998). Clinical application of human egg cryopreservation. *Hum Reprod*, 13, 3156-3159.
- Valk, J.V., Brunner, D., Desmet, K., Svenningsen, A.F., Honegger, P., Knudsen, L.E., Lind, T., Noraberg, J., Price, A., Scarino, M.L., & Gstraunthaler, G. (2010). Optimization of chemically defined cell culture media—replacing fetal bovine serum in mammalian *in vitro* methods. *Toxicol. in Vitro*, 24, 1053-1063.
- Verma, R., Holland, M.K., Temple-Smith, P., & Verma, P.J. (2012). Inducing pluripotency on somatic cells from the snow leopard (*Panthera uncia*), an endangered felid. *Theriogenology*, 77, 220-228.

- Wanderley, L.S., Luz, H.K., Faustino, L.R., Lima, I.M., Lopes, C.A., Silva, A.R., B  o, S.N., Campello, C.C., Rodrigues, A.P., & de Figueiredo, J.R. (2012). Ultrastructural features of agouti (*Dasyprocta aguti*) preantral follicles cryopreserved using dimethyl sulfoxide, ethylene glycol and propanediol. *Theriogenology*, 15, 260-7.
- Wang, L., Zhou, G., Shi, W., Shi, J., Tian, X., Gao, C., Zhang, L., Zhu, S.E., Zhang, T.T., Zeng, S.M., & Liu, G.S. (2012). First live offspring born in superovulated sika deer (*Cervus nippon*) after embryo vitrification. *Theriogenology*, 7, 1627-1632.
- Wiedemann, C., Hribal, R., Ringleb, J., Bertelsen, M.F., Rasmusen, K., Andersen, C.Y., Kristensen, S.G., & Jewgenow, K. (2012). X Preservation of primordial follicles from lions by slow freezing and xenotransplantation of ovarian cortex into an immunodeficient mouse. *Reprod Domest Anim*, 47, 300-304.
- Wiedemann, C., Zahmel, J., & Jewgenow, K. (2013). Short-term culture of ovarian cortex pieces to assess the cryopreservation outcome in wild felids for genome conservation. *BMC Vet Res*, 22, 37-48.
- Woelders, H., & Chaveiro, A. (2004). Theoretical prediction of optimal freezing programmes. *Cryobiology*, 49, 258-271.
- Wolvekamp, M.C.J., Cleary, M.L., Cox, S.L., Shaw, J.M., Jenkin, G., & Trounson, A.O., (2001). Follicular development in cryopreserved common wombat ovarian tissue xenografted to Nude rats. *Anim Reprod Sci*, 65, 135-147.
- Wong, P.B., Wiley, E.O., Johnson, W.E., Ryder, O.A., O'Brien, S.J., Haussler, D., Koepfli K.P., Houck, M.L., Perelman, P., Mastromonaco, G., Bentley, A.C., Venkatesh, B., Zhang, Y., & Murphy, R.W. (2012). Tissue sampling methods and standards for vertebrate genomics. *Giga Sci*, 1, 1-12.
- Xia, Z., Xin, D., Murray, D., Triffitt, J.T., & Price, A.J. (2013). A method of isolating viable chondrocytes with proliferative capacity from cryopreserved human articular cartilage. *Cell Tissue Bank*, 14, 267-276.
- Yamini, N., Pourmand, G., Amidi, F., Salehnia, M., Ataei Nejad, N., & Noori Mougahi, S.M.H. (2016). Developmental potential of vitrified mouse testicular tissue after ectopic transplantation. *Cell J*, 18, 74-82.
- Yeoman, R.R., Wolf, D.P., & Lee, D.M. (2005). Coculture of monkey ovarian tissue increases survival after vitrification and slow-rate freezing. *Fertil Steril*, 83, 1248-1254.
- Yokonishi, T., Sato, T., Komeya, M., Katagiri, K., Kubota, Y., Nakabayashi, K., Hata, K., Inoue, K., Ogonuki, N., Ogura, A., & Ogawa, T. (2014). Offspring production with sperm grown *in vitro* from cryopreserved testis tissues. *Nat Commun*, 5, 54320-54326.
- Zhang, Y.L., Liu, F.J., Zhuang, Y.F., Wang, X.A., Zhai, X.W., Li, H.X., & Zhang, W.C. (2012). Blastocysts cloned from the Putian Black pig ear tissues frozen without cryoprotectant at -80 and -196 degrees Celsius for 3 yrs. *Theriogenology*, 78, 1166-1170.
- Zhmakin, A.I. (2008). Physical aspects of cryobiology. *Uspekhi Fizicheskikh Nauk*, 178, 243-266.
- Zieger, M.A.J., Tredget, E.E., & McGann, L.E. (1996). Mechanisms of cryoinjury and cryoprotection in split-thickness skin. *Cryobiology*, 33, 376-389.

BIOGRAPHICAL SKETCH

Name: Alexsandra Fernandes Pereira

Affiliation: Laboratory of Animal Germplasm Conservation, Federal Rural University of Semi-Arid – UFERSA

Education: Graduate in Chemistry, Doctor in Veterinary Sciences

Research and Professional Experience: Adjunct Professor of UFERSA, Member of the Brazilian Society of Embryo Transfer

Professional Appointments: Expertise in studying the assisted techniques applied to the animal conservation and reproduction (*in vitro* culture, parthenogenesis, *in vitro* fertilization and somatic cell nuclear transfer

Publications Last 3 Years:

- Santos, M.L.T.; Borges, A.A.; Queiroz Neta, L.B.; Santos, M.V.O.; Oliveira, M.F.; Silva, A.R.; Pereira, A.F. *In vitro* culture of somatic cells derived from ear tissue of collared peccary (*Pecari tajacu* Linnaeus, 1758) in medium with different requirements. *Pesquisa Veterinária Brasileira (Online)*, 2016.
- Santos, M.V.O.; Queiroz Neta, L.B.; Borges, A.A.; Pereira, A.F. Influence of commercially available follicle stimulating hormone on the *in vitro* maturation of bovine oocytes. *Semina. Ciências Agrárias (Online)*, 2016.
- Sousa, F.C.; Melo, C.H.S.; Batista, R.I.T.P.; Sanchez, D.J.D.; Souza, J.M.G.; Pereira, A.F.; Melo, L.M.; Freitas, V.J.F.; Teixeira, D.I.A. Resultados ultrassonográficos de glândula mamária, fígado, vesícula biliar, baço e rins de caprinos transgênicos para o hG-CSF durante lactação induzida. *Semina. Ciências Agrárias (Online)*, 2016.
- Campelo I.S.; Pereira, A.F.; Alcântara Neto, A.S.; Canel, N.G.; Souza-Fabjan, J.M.G.; Teixeira, D.I.A.; Camargo, L.S.A.; Melo, L.M.; Rádis-Baptista, G.; Salamone, D.F.; Freitas, V.J.F. Effect of crotamine, a cell penetrating peptide, on blastocyst production and gene expression of *in vitro* fertilized bovine embryos. *Zygote (Cambridge. Print)*, v. 24, p. 48-57, 2016.
- Oliveira, Moacir Franco De; Oliveira, Gleidson Benevides De; Bezerra, Ferdinando Vinícius Fernandes; Câmara, Felipe Venceslau; Pereira, Alexsandra Fernandes; Silva, Alexandre Rodrigues; Miglino, Maria Angélica. The placental barrier of *Kerodon rupestris* (Rodentia: caviidae). *Bioscience Journal (Online)*, v. 32, p. 208-218, 2016.
- Pereira, A.F.; Melo, L.M.; Freitas, V.J.F.; Salamone, D.F. Phosphorylated H2AX in parthenogenetically activated, *in vitro* fertilized and cloned bovine embryos. *Zygote (Cambridge. Print)*, v. 23, p. 485-493, 2015.
- Moura, R.R.; Souza-Fabjan, J.M.G.; Fonseca J.F.; Melo, C.H.S.; Sanchez, D.J.D.; Vieira M.P.; Almeida, T.M.; Serov, I.A.; Serov, O.; Pereira, A.F.; Teixeira, D.I.A.; Melo, L.M.; Freitas, V.J.F. Reproductive parameters and the use of MOET in transgenic founder goat carrying the human granulocyte colony-stimulating factor (hG-CSF) gene. *Animal Reproduction*, v. 11, p. 37-43, 2014.

- Oliveira, Carla Rozilene Guimarães Silva; Melo, Carlos Henrique De Sousa; Souza-Fabjan, Joanna Maria Gonçalves De; Teles, Antônio Carlos De Albuquerque; Batista, Ribrio Ivan Tavares Pereira; Pereira, Alexsandra Fernandes; Melo, Luciana Magalhães; Freitas, Vicente José De Figueirêdo; Teixeira, Dárcio Ítalo Alves. Ultrasonographic evaluation of hG-CSF transgenic goat conceptus. *Revista Brasileira de Ciência Veterinária (Impresso)*, v. 21, p. 53-59, 2014.
- Freitas, Vicente J.F.; Serova, Irina A.; Andreeva, Lyudmila E.; Melo, Luciana M.; Teixeira, Dárcio I.A.; Pereira, Alexsandra F.; Moura, Raylene R.; Lopes-Jr, Edílson S.; Souza-Fabjan, Joanna M.G.; Batista, Ribrio I.T.P.; Serov, Oleg L. The comparison of two embryo donor breeds for the generation of transgenic goats by DNA pronuclear microinjection. *Animal Production Science (Print)*, v. 54, p. 564-568, 2014.
- Pereira, A.F.; Santos, M.L.T.; Borges, A.A.; Queiroz Neta, L.B.; Santos, M.V.O.; Feitosa, A.K.N. Isolamento e caracterização de células doadoras derivadas da pele para a transferência nuclear. *Acta Veterinaria Brasilica (UFERSA)*, v. 8, p. 311-316, 2014.
- Sanchez, Deisy J.D.; Melo, Carlos H.S.; Souza-Fabjan, Joanna M.G.; Sousa, Francisco C.; Rocha, Amanda A.; Campelo, Iana S.; Teixeira, Dárcio I.A.; Pereira, Alexsandra F.; Melo, Luciana M.; Freitas, Vicente J.F. Repeated hormonal treatment and laparoscopic ovum pick-up followed by *in vitro* embryo production in goats raised in the tropics. *Livestock Science (Print)*, v. 165, p. 217-222, 2014.
- Freitas, V.J.F.; Alcântara Neto, A.S.; Pereira A.F.; Campelo I.S.; Melo, L.M.; Rádis-Baptista, G. Assessing the complex formation between crotamine – a natural cell- penetrating peptide – and dna using high sensitive fluorescence exclusion assay. *Cloning and Transgenesis*, v. 3, p. 128-132, 2014.
- Moura, Raylene R.; Albuquerque, Erica S.; Melo, Carlos Henrique S.; Alcântara-Neto, Agostinho S.; Batista, Ribrio Ivan T.P.; Nunes-Pinheiro, Diana Célio S.; Pereira, Alexsandra F.; Teixeira, Darcio Ítalo A.; Melo, Luciana M.; Serova, Irina A.; Andreeva, Lyudmila E.; Serov, Oleg L.; Freitas, Vicente José F. Dynamics of recombinant hG-CSF in transgenic goat: preliminary study in the founder during hormonally induced lactation. *Animal Biotechnology*, v. 24, p. 10-14, 2013.
- Souza-Fabjan, J.M.G.; Pereira, A.F.; Melo, C.H.S.; Sanchez, D.D.; Oba, E.; Mermillod, P.; Melo, L.M.; Teixeira, D.I.A.; Freitas, V.J.F. Assessment of the reproductive parameters, laparoscopic oocyte recovery and the first embryos produced *in vitro* from endangered Canindé goats (*Capra hircus*). *Reproductive Biology*, v. 13, p. 1-8, 2013.
- Pereira, A.F.; Feltrin, C.; Almeida, K.C.; Carneiro, I.S.; Avelar, S.R.G.; Neto, A.S. Alcântara; Sousa, F.C.; Melo, C.H.S.; Moura, R.R.; Teixeira, D.I.A.; Bertolini, L.R.; Freitas, V.J.F.; Bertolini, M. Analysis of factors contributing to the efficiency of the *in vitro* production of transgenic goat embryos (*Capra hircus*) by handmade cloning (HMC). *Small Ruminant Research*, v. 109, p. 163-172, 2013.

Chapter 43

**MITOCHONDRIAL GENE DIVERSITY OF THE MEGA-
HERBIVOROUS SPECIES OF THE GENUS *TAPIRUS*
(TAPIRIDAE, PERISSODACTYLA) IN SOUTH AMERICA
AND SOME INSIGHTS ON THEIR GENETIC
CONSERVATION, SYSTEMATICS AND THE
PLEISTOCENE INFLUENCE ON THEIR
GENETIC CHARACTERISTICS**

***Manuel Ruiz-García^{1,*}, Armando Castellanos²,
Luz Agueda Bernal¹, Diego Navas¹,
Myreya Pinedo-Castro¹ and Joseph Mark Shostell³***

¹Laboratorio de Genética de Poblaciones-Biología Evolutiva,
Unidad de Genética, Departamento de Biología,
Facultad de Ciencias, Pontificia Universidad Javeriana,
Bogotá DC., Colombia

²Andean Bear Foundation, La Isla, Quito, Ecuador

³Department of Math Science and Technology,
University of Minnesota Crookston, Crookston, MN, US

ABSTRACT

We sequenced mitochondrial genes (*COI*, *COII*, *Cyt-b*) of accepted Latin America tapir species (*Tapirus pinchaque*, *T. terrestris* and *T. bairdii*) as well as an alleged new species, *T. kabomani*.

The mountain tapir (*T. pinchaque*) is a relatively rare large mammal species. Some population censuses indicate that no more than 2,000 mountain tapirs are left in the wilderness areas of Colombia, Ecuador and Peru. Our results showed that the gene diversity

* Corresponding Author's Email: mruizgar@yahoo.es, mruiz@javeriana.edu.co.

levels are medium to low with respect to other mammals sequenced for the same or similar genes. However, these gene diversity levels are not impoverished, which means that the genetic situation of this species is not as critical as its population censuses suggest. It will be crucial to determine the gene diversity levels in certain populations not included in the current work (the eastern and, possibly, western Andean Cordilleras in Colombia as well as the Tabaconas Namballe National Sanctuary in Peru), because they are probably the smallest populations of this species. On the other hand, the lowland tapir (*T. terrestris*), the species with the largest geographical distribution in Latin America, showed the highest gene diversity levels of all the other tapir species studied. Additionally, the genetic structure of *T. terrestris* is clearly more robust than that of *T. pinchaque*. Different geographic populations of both species showed different demographic trends throughout time. Our results including five samples of *T. kabomani* showed this taxon to be a haplogroup within *T. terrestris*, reducing the likelihood of *T. kabomani* being a new full species. Finally, we also analyzed the influence of diverse Pleistocene climatic changes on the mitochondrial haplotype diversification of *T. terrestris* and *T. pinchaque*. The Pleistocene Refugia and the Recent Lake hypotheses probably played integral roles in the evolutionary history of *T. terrestris*. In contrast, the Pleistocene Refugia hypothesis involving the Andes, which probably played an important part in the genetic diversification of other mammals, did not have a significant impact on *T. pinchaque*.

Keywords: mitochondrial genes (*COI*, *COII* and *Cyt-b*), *Tapirus*, *T. pinchaque*, *T. terrestris*, *T. bairdii*, “*T. kabomani*,” conservation genetics, speciation, pleistocene biodiversity hypotheses

INTRODUCTION

Large mammals are often considered umbrella or emblematic species whose presence may greatly affect community structure (Meffe and Carrol, 1997). In general, the extraction of large vertebrates, especially those at the top of the trophic chain, can provoke strong extinction waves in complex natural systems (Pimm, 1991). These species can be considered as key to overall community structure because they live sympatrically with many other species. For example, East (1981) showed that the preservation of viable populations of the African wild dog (*Lycaon pictus*) and the cheetah (*Acinonyx jubatus*), due to their large hunting territories, helped to preserve the communities of other large African mammals.

In Latin America, three wild large perissodactyla species have been traditionally accepted: the Baird or Central America tapir (*Tapirus bairdii*), the mountain tapir (*T. pinchaque*) and the lowland or Brazilian tapir (*T. terrestris*).

The fossil evolution of the Tapiridae is old and very interesting. Perissodactyla is a very old group of mammals that arose around 60 Million years ago (MYA). In the fossil record, there are representative specimens from five main superfamilies (Tapiroidea, Rhinoceroidea, Chalicotheroidea, Equoidea and Brontotheroidea) and 14 different families (Savage and Long, 1986; Holbrook, 1999). The order reached its maximum diversity peak during the Eocene and then declined from 14 to 4 families during the upper Oligocene (Radinsky, 1969; Froehlich, 1999; MacFadden, 1992; Metais et al., 2006). The Tapiridae family (Gray 1821) is currently composed of one unique genus, *Tapirus* (Brisson 1762). Colbert (2005) defined the Tapiridae family as the clade conformed by the most recent common ancestor of *Protapirus*. The oldest record of *Tapirus* comes from the European Oligocene (33-37 MYA), where the fossil remains

are found until the Pleistocene (McKenna and Bell, 1997). In North America, the *Tapirus* records indicate that they were present in the Middle Miocene through the present (Hulbert, 1995), whereas for Asia the records indicate that *Tapirus* has been in existence since the Lower Miocene (Deng, 2006). Currently, additional to the three quoted Latin American tapir species, there is a fourth species in Southern Eastern Asia, the Malayan tapir (*T. indicus*). Two of these species are exclusively distributed in South America and they are the subject of this work, from a genetics point of view. The mountain tapir (*T. pinchaque*) has a unique and critical role because it is an indispensable seed disperser in the Andean mountains (Colombia, Ecuador, Northern Peru) but its conservation status is in jeopardy because of its relatively low numbers (less than 2,500 individuals; Cavellier et al., 2010). Additionally, this is an umbrella species with an average territory size per tapir of 880 ha (Downer, 1996). This is a substantially large home range. Downer's work in the Sangay National Park suggested that a minimum of 1,000 mountain tapirs were needed to ensure a reproductively viable population. Such a population would need at least 293,500 ha. The necessary protection of this area would positively benefit many other sympatric animal species, like the Andean bear, puma, and various deer species (Downer, 1996). Indeed, the cloud forests of the Andean region, where the mountain tapir lives, contain some of the greatest organismic diversity on Earth (Brehm et al., 2005; Krömer et al., 2005), with a noteworthy number of endemic species (Kessler, 2002). However, about 90 % of the mountain Andean forests have been deforested (Henderson et al., 1991) and a great proportion of páramos has been transformed via burns into extensive ranchlands and croplands (potatoes) (Verweij, 1995). Additionally, around 83 % of Colombia's mountain forests have been highly affected by human activity (Cavellier and Etter, 1995). Besides ranching and agriculture, the mountain tapir is also highly threatened by poaching and illegal trade of its body parts (Downer, 2003). In Andean areas, the introduction of livestock has also spread new diseases and attracted more potential predators of the mountain tapir.

The lowland tapir has a wide geographical distribution across a great part of South America. From an ecological perspective, the lowland tapir is considered the "architect" of their Neotropical habitats because it is a very efficient disperser and seed predator (Wallace et al., 2010). Cabrera (1961) considered four morphological and geographical subspecies of *T. terrestris*: *T. t. colombianus*, *T. t. terrestris*, *T. t. aenigmaticus*, and *T. t. spegazzinni*. *T. t. colombianus* (Hershkovitz 1954; Type locality: El Salado, between Valencia and Pueblo Bello on the eastern slope of the Sierra Nevada de Santa Marta, in Magdalena, Colombia) inhabits Northern Colombia and around Maracaibo Lake in Venezuela. *T. t. terrestris* (Linnaeus 1758; Type locality: Pernambuco, Brazil) lives in Venezuela, Guyana, Suriname, French Guyana, with its major fraction in Brazil (including Amazonas) up to the Misiones within Argentina. *T. t. aenigmaticus* (Gray 1872; Type locality: Macas, Eastern Ecuador) is distributed in Southeastern Colombia, Eastern Ecuador and Northern/Eastern Peru (Amazonian area). *T. t. spegazzinni* (Ameghino 1909; Type locality: Rio Pescado, Orán, Salta, Argentina) is distributed in Southeastern Brazil and Mato Grosso, Paraguay, Eastern Bolivia and Northwestern Argentina.

Recently, Cozzuol et al., (2013) claimed the existence of a new tapir species ("*T. kabomani*") in South America based on a comparison of mitochondrial Cytochrome-b (*Cyt-b*) gene sequences of four Amazon tapir individuals (two Brazilian animals sampled by these authors and two Colombian specimens sampled by the first author of the present study, M. R-G) to the 45 *Cyt-b* gene sequences published by Thoisy et al., (2010). They also enclosed the results of two additional mitochondrial genes (mitochondrial Cytochrome Oxidase subunit I,

COI, and subunit II, *COII*) for six *T. terrestris*, one *T. pinchaque* and three individuals of the alleged new species “*T. kabomani*.” They concluded that the three mitochondrial genes supported “*T. kabomani*” as a full species. Voss et al., (2014), however, questioned the validity of “*T. kabomani*” as a different species from *T. terrestris*.

The current work provides some insight about the genetic structure and heterogeneity, demographic evolution, spatial structure, biological conservation, and systematics of the Latin American tapirs. Moreover, this work also explores the influence of the Pleistocene period on the Latin American tapirs. And, all of these insights are based on the analysis and study of mitochondrial genes *Cyt-b*, *COI*, and *COII*. These new data should be helpful in validating or rejecting the claim of a new species, “*T. kabomani*”.

MATERIALS AND METHODS

Two sets of data were used in this study. The first set consisted of 93 Latin American tapir individuals analyzed for three specific concatenated mitochondrial genes (*Cyt-b* + *COI* + *COII*). They were the same genes analyzed by Cozzuol et al. (2013) and whose mitochondrial sequences are accessible in GenBank for the two Brazilian specimens they classified as an alleged new species “*T. kabomani*.” Breaking the first group down by taxa, it had 46 *T. pinchaque*, 29 *T. terrestris*, 13 *T. bairdii*, and 5 *T. kabomani*. *T. pinchaque* was constituted by 26 from Ecuador representing three populations, Northern Ecuador, Sangay National Park and Podocarpus National Park, and the remainder of the 46 (20) was from Colombia representing three populations: Los Nevados National Park, Tolima and Purace National Park.

Another 29 of the first set were *T. terrestris*. Five of these individuals were from Ecuador, four were from Colombia, seven from Peru, three from Bolivia, three from Brazil, three from Surinam, and one was from Argentina. Three animals of the 29 were from American zoos in Cincinnati, Milwaukee and Columbus. Another 13 of the 93 were *T. bairdii*. Nine of these animals were from Panama, three were from Costa Rica and one was from Colombia. Five of the 93 were *T. kabomani*- including the two Brazilian specimens of Cozzuol et al., (2013). We detected three additional specimens with mitochondrial haplotypes of the alleged *T. kabomani*. There were from 1) San Martín de Amacayacu along the Amazon River in Colombia, 2) the Mazan River which is a tributary of the Napo River, in the Peruvian Amazon, and 3) near Tena along the upper Napo River in the Ecuadorian Amazon.

However, at least, two of these animals (the Colombian and the Ecuadorian ones) presented typical morphotypes of *T. terrestris*, although they showed “*T. kabomani*” haplotypes (see Figure 1). The Peruvian sample consisted of skin obtained from hunters and we could not determine the morphotype of the specimen.

The second set of individuals contained 141 *T. terrestris*, sequenced at the *Cyt-b* gene. Forty-one of these came from different regions of Colombia. Of these, 1 was from Bajo Sinú-Tierra Alta, (Córdoba Department), 2 were from Mesay River (Caquetá Department), and another was from Fondo Canaima, (Vichada Department). Of the 41, 18 were also from Leticia to San Juan de Atacuarí (Amazonas Department), 7 from the Eastern Colombian Llanos (Meta Department), 3 from Pto. Inirida (Guania Department), 3 from Palomino River-Sierra Nevada de Santa Marta and 6 from the Antioquia Department. Five of the 141 were from Venezuela (El Zulia, Maracaibo). Eleven of the 141 were from French Guiana (Carnopi River). Seven

animals were from Ecuador (4 from Limoncocha, Sucumbios and 3 from Coca, Sucumbios). Thirty animals were from Peru [(one from Arica (Curaray River), 2 from Napo River (Nueva Vida and Mazán), 7 from Nanay River, one from Requena (Ucayali River), one from Breña (Canal del Puhinauva-Ucayali River), 4 from Pucallpa (Ucayali River), and 15 from Pto. Maldonado (Madre de Dios River)]. Eleven animals were from Bolivia (9 from Mamoré River, one from Chimoré River and one from Villa Bella at the Beni River). Twenty four animals were from Brazil [2 from Yavarí River, 12 from Tabatinga, 2 from Negro River, 2 from Santarém (Pará state) and 6 from the Amazon's mouth (Pará state)]. One animal was from Paraguay (from Hernandarias). And, 4 animals were from Argentina (the Yungas area in Salta-Jujuy). Additionally, one animal was from the Barcelona Zoo (Spain) and 5 animals were from US zoos. One animal of unknown origin was also analyzed.



(A)



(B)

Figure 1. An individual with morphotype of *Tapirus terrestris* but with “*T. kaboman*” mitochondrial haplotype from San Martín de Amacayacu, Amazon River, Colombian Amazon (A); An individual with morphotype of *Tapirus terrestris* but with “*T. kaboman*” mitochondrial haplotype from near to Tena, upper Napo River, Ecuadorian Amazon (B).

Molecular Procedures

DNA from teeth, bones, muscle, and skins, were obtained with the phenol-chloroform procedure (Sambrook et al., 1989), whereas DNA samples from hair and blood were obtained with 10% Chelex® 100 resin (Walsh et al., 1991). Amplifications for *Cyt-b* gene were achieved using primers L7 (5' ACC AAT GAC ATG AAA AAT CAT CGT T 3') and H6 (5' TCT CCA TTT CTG GTT TAC AAG AC 3'), that had been designed for perissodactyles (Tougard et al., 2001). The PCR reactions were performed in a 50- μ l volume, including 10 μ l of 10x Buffer, 7 μ l of 3 mM MgCl₂, 2 μ l of 10 mM dNTPs (dNTP Mix Promega), 4 μ l (15 pmol) of each primer, one unit of Taq DNA polymerase (genTaq polimerasa), 2 μ l of DNA from blood, skin or muscle tissue (50-200 ng/ μ l) or 4-10 μ l of DNA from hair and teeth (6-35 ng/ μ l) and a variable quantity of ddH₂O. PCR reactions were carried out in a Geneamp PCR system 9600 (Perkin Elmer) and in an iCycler™ BioRad thermocycler. We used the following temperatures: 94 °C for 5 minutes, 35 cycles of 50 s at 94 °C, 50 s at 53 °C and 1.5 minutes at 72 °C and a final extension time for 10 minutes at 72 °C. For the *COI* and *COII* genes, the amplification conditions followed Ashley et al., (1996) and Hebert (2003, 2004). All amplifications, including positive and negative controls, were checked in 2 % agarose gels, employing the molecular weight marker ϕ X174 DNA digested with *Hind* III, *Hinf* I and HyperLadder IV. The gels were visualized in a Hoefer UV Transilluminator. Those samples that amplified were purified using membrane-binding spin columns (Qiagen). The double-stranded DNA was directly sequenced in a 377A (ABI) automated DNA sequencer. The samples were sequenced in both directions and all the samples were repeated to ensure sequence accuracy.

Data Analyses

Genetic Diversity, Linkage and Heterogeneity Analyses

The sequences were edited and aligned with BioEdit Sequence Alignment Editor (Hall, 2004) and DNA Alignment (Fluxus Technology Ltd).

We used the following statistics to determine the genetic diversity at the three concatenated mitochondrial genes for *T. terrestris*, *T. pinchaque*, *T. bairdii* and the alleged "*T. kabomani*": the number of polymorphic sites (S), the number of haplotypes (H), the haplotypic diversity (H_d), the nucleotide diversity (π), the average number of nucleotide differences (k) and the θ statistic by sequence.

The possible linkage disequilibrium within the *Cyt-b* gene for *T. terrestris* was evaluated by two different statistics. We obtained the average value of R^2 for all the comparison pairs among nucleotide sites (ZnS) (Kelly, 1997) as well as among the adjacent polymorphic nucleotide sites (Za), and ZZ (Rozas et al., 2001). This is equal to $Za - ZnS$, using the software DNAsp 5.10 (Librado and Rozas, 2009).

Different tests were carried out to measure genetic heterogeneity, and possible gene flow estimates, among the four Latin American tapir taxa studied. These tests were those of Hudson et al., (1992a,b) (H_{ST} , K_{ST} , K_{ST}^* , Z , Z^*), Hudson (2000)'s Snn test and the chi-square test on the haplotypic frequencies with permutation tests using 10,000 replicates. We also estimated the G_{ST} statistic from the haplotypic frequencies and the γ_{ST} , N_{ST} and F_{ST} statistics (Hudson et al., 1992a) from the nucleotide sequences.

We carried out two different AMOVA analyses at the *Cyt-b* gene for *T. terrestris* to determine if the gene diversity within this species was distributed in different hierarchical geographical levels (Excoffier et al., 1992). The first analysis was carried out taking into account the six haplogroups found in the phylogenetic analysis clustered in three main population sets: northern, Amazonian, and southern. Thus, this analysis was applied to the overall geographical *T. terrestris* range analyzed. The second analysis was completed taking into account six different geographical areas within the Amazon basin: 1-Northwestern Amazon in Colombia and Brazil, 2- Central and Eastern Brazilian Amazon, 3-Napo River and a tributary in Peru and Ecuador, 4- Ucayali River in Peru, 5-Madre de Dios River in Peru and 6- Mamore River and a tributary in Bolivia. These six geographical areas were clustered into two main groups: northern Amazon (the first three areas) and southern Amazon (the last three areas). This analysis was carried out to determine the gene diversity structure in the Amazon region. The fixation indices of Wright (1951) were estimated: Φ_{sc} (variation of populations within the groups), Φ_{ct} (variation among groups) and Φ_{st} (variation among individuals). These analyses were carried out by means of the software ARLEQUIN 3.0 (Excoffier et al., 2005).

Phylogenetic Analyses

The Modeltest (Posada and Crandall, 1998) and the Mega 5.1 software (Tamura et al., 2011) were applied to determine the best evolutionary nucleotide model for the analyzed concatenated gene sequences for all the *Tapirus* taxa studied.

To determine if *T. pinchaque* or *T. terrestris*, contained a more conspicuous genetic structure, we used a consensus maximum parsimony (MP) tree for the three concatenated mitochondrial genes. We also obtained a maximum likelihood (ML) tree to determine how many significant clades were within *T. terrestris*, at the *Cyt-b* gene. In this analysis, 80 *T. terrestris* haplotypes were considered (including two haplotypes later classified as “*T. kabomani*”). To carry out this task, we tested the hypothesis that the *T. terrestris* sequences fall into one to 10 different groups. These classifications were contrasted with the ML tree. For this, we performed parametric bootstrapping and a posteriori significance test with the Swofford-Olsen-Waddell-Hillis test (SOWH; Huelsenbeck and Bull, 1996; Swofford et al., 1996). The 10 hypotheses were used as a model tree for parameter estimation and for generating 100 replicate data sets in the software Seq-Gen 1.2.5 (Rambaut and Grassly, 1997) which presented a uniform base composition. Goldman et al., (2000) demonstrated that this procedure can increase power in rejecting the null hypothesis and is better than typical nonparametric tests for comparisons of a posteriori hypotheses. The same was performed with the Shimodaira and Hasegawa (1999) test (a nonparametric SH test).

We conducted two additional phylogenetic trees. One was a ML tree with the 93 tapirs sequenced at the three concatenated mitochondrial genes to determine the relationship of the five “*T. kabomani*” with regard to the other tapir taxa. The second was an ML tree with some animals only sequenced at the *COI* gene since this gene is used as a barcode to discriminate different species (Herbert et al., 2003, 2004). Similarly, we estimated the Kimura 2P genetic distance (Kimura, 1980), for the three mitochondrial genes, among all the Neotropical *Tapirus* taxa because this genetic distance is a standard measurement for barcoding tasks (Hebert et al., 2003, 2004).

Possible divergence times were obtained using a Median Joining Network (MJN) (Bandelt et al., 1999). These were applied to all the haplotypes of the three concatenated mitochondrial

genes for all the tapir taxa considered as well as to the *T. terrestris* haplotypes at the *Cyt-b* gene. These were constructed with Network 4.6 software (Fluxus Technology Ltd). The ρ statistic (Morral et al., 1994) was estimated and transformed into years. The standard deviation of ρ was also calculated (Saillard et al., 2000), which is unbiased and highly independent of past demographic events. Thoisie et al., (2010) used two different mutation rates: 5.6×10^{-3} and 2.5×10^{-2} substitutions/site/million years respectively. Relative to our work with the genus *Tapirus*, the first rate translated into around one mutation each 537,634 years. The second rate was equal to about one mutation each 120,482 years. This last mutation rate agrees quite well with that reported by Nabholz et al., (2008, 2009) for mammals at the *Cyt-b* gene. We employed an intermediate value of 1.5×10^{-2} substitutions/site/million years (around one mutation each 200,000 years), closer to the second value used by Thoisie et al., (2010).

Demographic Changes

We used two methods to determine possible demographic changes across the natural history of *T. pinchaque*, *T. terrestris* and “*T. kabomani*” for the three concatenated mitochondrial genes as well as for the *Cyt-b* gene in the lowland tapir. (1) Following the method of Rogers and Harpending (1992) and Rogers et al., (1996) we used a mismatch distribution (pairwise sequence differences). The raggedness rg statistic (Harpending et al., 1993; Harpending, 1994) was used to determine the similarity between the observed and the theoretical curves. (2) We used the Fu and Li D and F tests (Fu and Li, 1993), the Fu F_s statistic (Fu, 1997), the Tajima D test (Tajima, 1989) and the R_2 statistic (Ramos-Onsins and Rozas, 2002) to determine possible changes in population size (Simonsen et al., 1995; Ramos-Onsins and Rozas, 2002). All of these statistics and tests were obtained by means of the DNAsp 5.10 and Arlequin 3.0 programs.

Spatial Structure

Garnering this information can assist in our understanding of the evolutionary events that have determined the natural history of the tapir species. A Mantel's test (Mantel, 1967) was used to detect possible overall relationships between a genetic matrix among individuals (Log-Det genetic distance; Nei and Kumar, 2000) and the geographic distance matrix among the individuals analyzed. In this study, Mantel's statistic was normalized according to Smouse et al., (1986). This procedure transforms the statistic into a correlation coefficient. The geographic distances were measured with the Spuhler's (1972) procedure, where $D = \arccos(\cos X_{(i)} \cdot \cos X_{(j)} + \sin X_{(i)} \cdot \sin X_{(j)} \cos |Y_{(i)} - Y_{(j)}|)$, where $X_{(n)}$ and $Y_{(n)}$ are the latitude and longitude of the n th individual sampled, respectively. The significance of the correlations obtained was tested using a Monte Carlo simulation with 1,000 permutations. This test was undertaken by means of the software NTSYS v. 2.1 (Rohlf, 2000). This analysis was carried out for *T. pinchaque*, *T. bairdii* and the overall sample of *T. terrestris* for the three concatenated mitochondrial genes as well as for the different haplogroups found and for different geographical regions at the *Cyt-b* gene for *T. terrestris*. For all the Amazon *T. terrestris* individuals sequenced at the *Cyt-b* gene, an AIDA analysis was used (Bertorelle and Barbujani, 1995). The expressions of the respective AIDAs coefficients are as follows:

$$H = (n \sum_{i=1 \text{ to } n} \sum_{j>i \text{ to } n} w_{ij} \sum_{k=1 \text{ to } S} (p_{ik} - p_k) (p_{jk} - p_k)) / (W \sum_{i=1 \text{ to } n} \sum_{k=1 \text{ to } S} (p_{ik} - p_k)^2)$$

and

$$CC = ((n - 1) \sum_{i=1 \text{ to } n} \sum_{j > i \text{ to } n} w_{ij} \sum_{k=1 \text{ to } S} (p_{ik} - p_{jk})^2) / (2W \sum_{i=1 \text{ to } n} \sum_{k=1 \text{ to } S} (p_{ik} - p_k)^2)$$

where n is the sample size, W is the number of pairwise comparisons of a distance class given, p_{ik} and p_{jk} are the haplotypes of the i th and j th individuals, respectively. At the k th nucleotide site, p_k is the k th element of the average vector and w_{ij} is one if individuals i and j are within the same distance class; otherwise it is zero. Summation includes the Σ nucleotide sites for all the n individuals analyzed. To connect the individuals sequenced within each distance class, the Gabriel-Sokal network (Gabriel and Sokal, 1969; Matula and Sokal, 1980) and the Delaunay's triangulation with elimination of the crossing edges (Ripley, 1981; Upton and Fingleton, 1985; Isaaks and Srivastava, 1989) were used. However, the results were very similar in each case. The Bonferroni (Oden, 1984), Oden's Q and the Kooijman's tests were estimated to determine the statistical significance of the autocorrelation coefficients.

RESULTS

Gene Diversity, Genetic Heterogeneity and Phylogenetic Considerations for the Latin American Tapir Taxa

The GTR with gamma distributed rate variation among sites for both the maximum likelihood and the AIC criteria was the best fit nucleotide substitution model at the three concatenated genes.

Out of 4,753 comparisons among the total polymorphic sites at the *Cyt-b* gene in *T. terrestris*, only 303 comparisons were statistically significant by means of a Fisher exact test (6.4 %), but they were not different from the Type I error of 5 %. Similarly, the values of the statistics of Kelly (1997) ($ZnS = 0.0219$; 95 % of Confidence interval, $CI = 0.01498 - 0.31650$), of Rozas et al., (2001) ($Za = 0.0244$; $CI = 0.01706 - 0.32607$) and $ZZ = 0.0026$ ($CI = -0.05363 - 0.05625$) also showed that there was no significant association among the polymorphic sites at this mitochondrial gene. Therefore, no evidence of linkage among polymorphic sites was detected at this mitochondrial gene. Out the four taxa of *Tapirus* considered at the three concatenated mitochondrial genes, clearly *T. terrestris* showed the highest levels of gene diversity ($H_d = 0.985$; $\pi = 0.0094$; $k = 22.98$), followed by *T. pinchaque* ($H_d = 0.957$; $\pi = 0.0041$; $k = 9.98$). In contrast, "*T. kabomani*" ($H_d = 0.900$; $\pi = 0.0033$; $k = 8.00$) and *T. bairdii* ($H_d = 0.910$; $\pi = 0.0017$; $k = 4.08$) yielded the lowest levels of gene diversity (Table 1). The "*T. kabomani*" gene diversity should be enclosed within the gene diversity of *T. terrestris* (proof 1). The genetic heterogeneity among these four taxa was highly significant and of a large magnitude for all the statistics used (for instance, $\gamma_{ST} = 0.808$ and $F_{ST} = 0.903$; Table 2a), with the associated indirect gene flow values practically non-existent ($N_m = 0.05-0.12$).

Table 1. Gene diversity statistics for *Tapirus terrestris*, *T. pinchaque*, *T. bairdii* and the alleged new species “*T. kabomani*” at the mitochondrial genes sequenced (*Cyt-b* + *COI* + *COII*). The statistics estimated were the number of haplotypes (NH), the haplotypic diversity (H_d), the nucleotide diversity (π), the average number of nucleotide differences (K) and the θ statistic ($= 2N_e\mu$; N_e = effective female population size; μ = mutation rate per generation) by sequence

	NH	H_d	π	K	θ per sequence
<i>Tapirus terrestris</i>	24	0.985 ± 0.014	0.0094 ± 0.0007	22.985 ± 10.42	29.538 ± 9.448
<i>Tapirus pinchaque</i>	31	0.957 ± 0.021	0.0041 ± 0.0005	9.981 ± 4.648	13.652 ± 4.161
“ <i>Tapirus kabomani</i> ”	4	0.900 ± 0.161	0.0033 ± 0.0009	8.000 ± 4.483	7.680 ± 4.165
<i>Tapirus bairdii</i>	9	0.910 ± 0.068	0.0017 ± 0.0003	4.077 ± 2.173	5.156 ± 2.268

Table 2. Genetic heterogeneity statistics applied to Latin American tapirs: among four Latin American tapir taxa (*T. pinchaque*, *T. terrestris*, *T. bairdii* and the alleged new species “*T. kabomani*”) (2A); between *T. pinchaque* and *T. terrestris* (2B). *Significant Probability ($P < 0.01$)

Table 2A. Among four Latin American tapir taxa

Genetic differentiation estimated	P	Gene flow	
$\chi^2 = 279.000$ df = 201	0.0002*		
$H_{ST} = 0.0287$	0.0000*	$\gamma_{ST} = 0.8079$	Nm = 0.12
$K_{ST} = 0.8026$	0.0000*	$N_{ST} = 0.9078$	Nm = 0.05
$K_{ST}^* = 0.3433$	0.0000*	$F_{ST} = 0.9034$	Nm = 0.05
$Z_S = 941.8106$	0.0000*		
$Z_S^* = 6.4554$	0.0000*		
$S_{nn} = 1.0000$	0.0000*		

Table 2B. *T. pinchaque* sample vs. *T. terrestris* sample

Genetic differentiation estimated	P	Gene flow	
$\chi^2 = 104.041$ df = 70	0.0052*	$G_{ST} = 0.0411$	Nm = 11.68
$H_{ST} = 0.0414$	0.0004*	$\gamma_{ST} = 0.4389$	Nm = 0.64
$K_{ST} = 0.4167$	0.0000*	$N_{ST} = 0.5158$	Nm = 0.47
$K_{ST}^* = 0.2421$	0.0000*	$F_{ST} = 0.5145$	Nm = 0.47
$Z_S = 678.8785$	0.0000*		
$Z_S^* = 6.1920$	0.0000*		
$S_{nn} = 0.7876$	0.0000*		

In another heterogeneity analysis, *T. pinchaque* was only analyzed with regard to *T. terrestris* (Table 2b). In this case, the seven genetic heterogeneity tests were also highly significant at the $P < 0.05$ level. The relative genetic differentiation tests yielded elevated amounts of genetic heterogeneity between the *Tapirus* taxa ($\gamma_{ST} = 0.439$ and $F_{ST} = 0.514$). The gene flow estimates were clearly lower than 1 ($Nm = 0.47$ - 0.64), which showed that these taxa

are genetically disconnected. Although, “*T. kabomani*” showed significant differences with *T. terrestris* as well as with *T. pinchaque*, the genetic differences with *T. terrestris* are considerably lower than with *T. pinchaque* (six significant heterogeneity tests, $\gamma_{ST} = 0.151$ and $F_{ST} = 0.454$, $Nm = 2.82-0.60$ versus seven significant heterogeneity tests, $\gamma_{ST} = 0.300$ and $F_{ST} = 0.711$, $Nm = 1.16-0.20$, respectively). Therefore, “*T. kabomani*” is less differentiated genetically speaking from *T. terrestris* than from *T. pinchaque* and it showed less differentiation in regard to *T. terrestris* than the genetic differentiation between *T. terrestris* and *T. pinchaque*. This provides strong evidence (proof 2) that “*T. kabomani*” is more closely related to *T. terrestris* than to the other tapir species, which disagrees with the point of view of Cozzuol et al., (2013).

The genetic structure within *T. terrestris* is much more robust (greater) than within *T. pinchaque*. This is verified by the consensus MP tree (Figure 2). *T. pinchaque* showed some small geographical clusters with several individuals of nearby geographical regions. Within *T. pinchaque* there were five Colombian individuals (four from Los Nevados National Park and one from Gaitania, Tolima) and two individuals from the Alto Papallacta River in the Napo Province in Ecuador. It also contained seven Ecuadorian individuals (six from the Napo Province and one from Sangay National Park), two from the Napo Province and an Ecuadorian cluster of six.

The Ecuadorian cluster represented three different Ecuadorian populations. Two *T. pinchaque* sequences were more related to those from *T. terrestris* than to those of the other mountain tapirs. These were the cases of one Colombian individual from La Planada, Tolima (Colombia) and one Ecuadorian individual from Chaco, Oyacachi, Napo. However, the relationships among the *T. pinchaque* were considerably less structured than those observed within *T. terrestris*. This second species was composed by more robust haplogroups than those detected in the mountain tapir. Even an exact probability test with a Markov chain length of 100,000 steps only revealed one significant population comparison pair: Los Nevados vs. Sangay National Park ($p = 0.0297 \pm 0.0013$) out of the six populations of *T. pinchaque* studied. This may indicate that (proof 3) *T. terrestris* is an older species than *T. pinchaque*.

The most developed genetic structure in *T. terrestris* can be seen in the AMOVA analyses carried out. The first AMOVA, with the six haplogroups, showed that the major part of the genetic variance was among lineages (64%; $\Phi_{sc} = 0.762$, P-value = 0.000 ± 0.00), followed by the gene variance among the individuals (36%; $\Phi_{st} = 0.641$, P-value = 0.000 ± 0.000). In contrast, the genetic variance among the three main groups (North, Amazonian and South) did not show significant variance (0%, $\Phi_{ct} = -0.510$, P-value = 1.000 ± 0.000). However, in the second AMOVA, by geographical areas within the Amazon basin, the major part of the genetic variance was explained by the difference among the individuals (90%, $\Phi_{st} = 0.499$, P-value = 0.000 ± 0.000). The genetic variance among the populations, although significant, only explained 10 % of the genetic variance ($\Phi_{sc} = 0.148$, P-value = 0.000 ± 0.000). The gene variance between the two main groups was again not significant (0%, $\Phi_{ct} = -0.057$, P-value = 0.913 ± 0.008).

The lower gene variance among geographical regions relative to among haplogroups is explained because the diverse haplogroups coexisted sympatrically at the same points of the Northern Peruvian Amazon (four out of six different haplogroups). This was also true for the Colombian Amazon (six of six), Northwestern Brazilian Amazon (three out of six), Southern Peruvian Amazon (three out of six) and Bolivian Amazon (three out of six). The existence of

extremely, well-differentiated haplogroups within *T. terrestris* could help us to understand why “*T. kabomani*” is an additional haplogroup within “*T. kabomani*” rather than a full species (proof 4). An ML tree, applied to *T. terrestris* at the *Cyt-b*, showed that six haplogroups were detected within this species (Figure 3). The SOWH tests (and also the SH tests) indicated that there was neither support for the taxonomic schemes of 1 to 5 clusters nor for the 7 to 10 clusters. The maximum likelihood trees were significantly different at the 0.001 level (since 12,569 to 689,999 log likelihood units). However, the scheme suggested that six clusters did not significantly deviate from the tree we obtained (1,455 log likelihood units, $p < 0.55$).

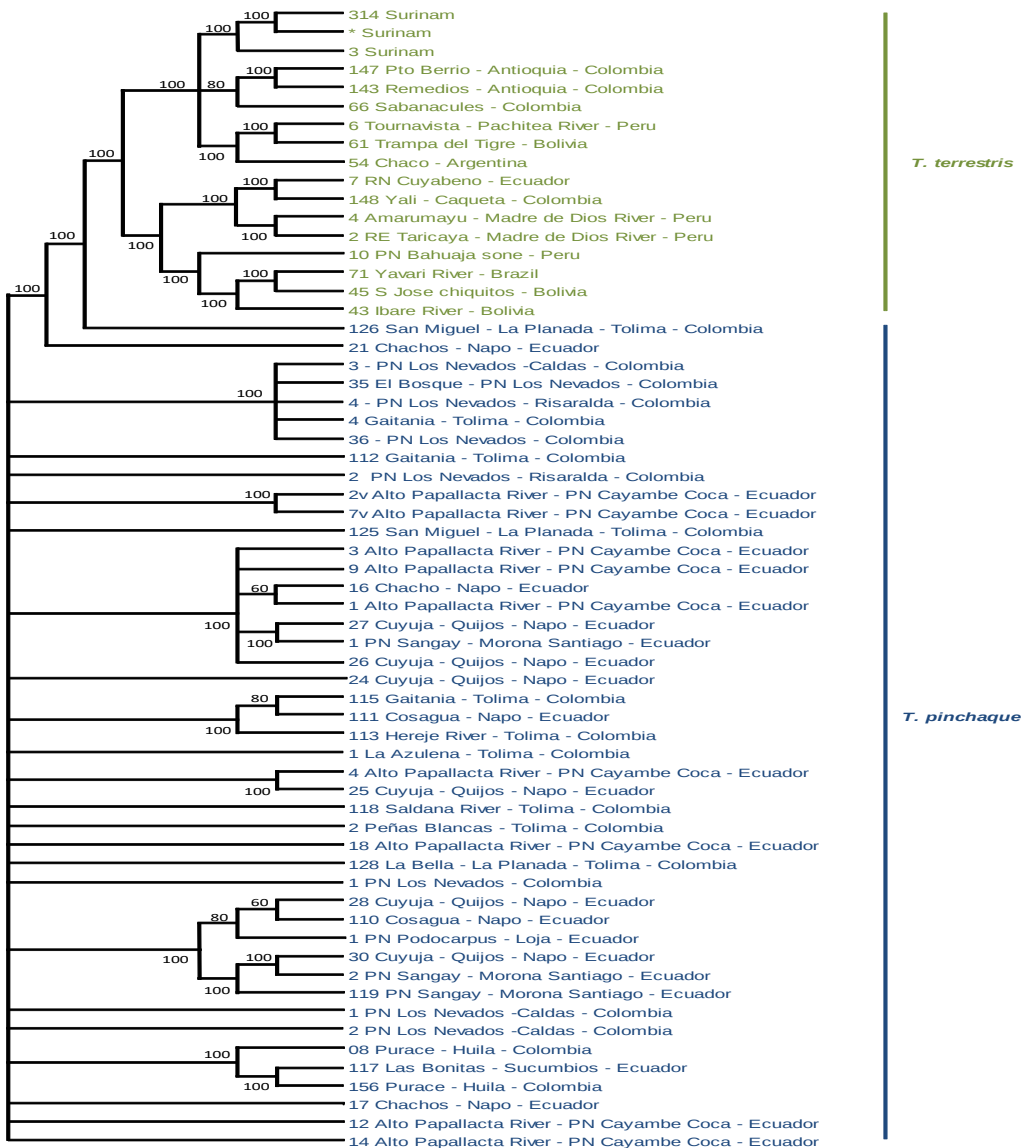


Figure 2. Consensus maximum parsimony tree with 45 *Tapirus pinchaque*, plus 17 *T. terrestris* by using the concatenated sequences of three mitochondrial genes (*Cyt-b* + *COI* + *COII*).

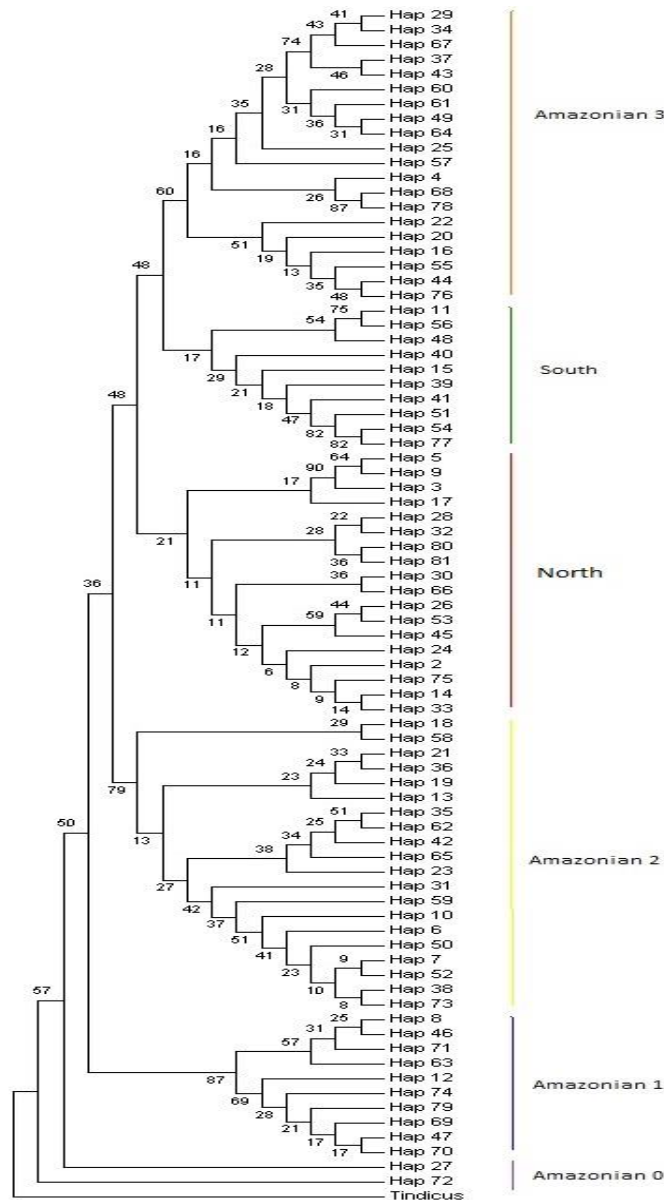


Figure 3. Maximum likelihood (ML) tree with the 80 haplotypes (from haplotype 2 to haplotype 81) found at the *Cyt-b* gene (1,140 bp) for the 141 lowland tapirs sequenced analyzed. Numbers in the nodes are bootstrap percentages.

Table 3 shows the Kimura 2P genetic distances among all of the Neotropical *Tapirus* taxa. For all three genes, “*T. kabomani*” yielded lower genetic distances with regard to *T. terrestris* than did *T. pinchaque* (*Cyt-b*: 0.018 ± 0.003 vs. 0.024 ± 0.003 ; *COI*: 0.005 ± 0.002 vs. 0.010 ± 0.004 ; *COII*: 0.013 ± 0.003 vs. 0.017 ± 0.004 , respectively). The genetic distances between “*T. kabomani*” and *T. terrestris*, typical of different populations within a species, are not surprising results. However, the small genetic distances between what are traditionally considered full species (*T. terrestris* and *T. pinchaque*) are (proof 5).

Table 3. Kimura 2P genetic distance (Kimura, 1980), in percentages, with standard deviations among Neotropical *Tapirus* taxa pairs for three mitochondrial genes (*COI*, *COII* and *Cyt-b*)

COI				
Taxa	<i>T. terrestris</i>	" <i>T. kabomani</i> "	<i>T. pinchaque</i>	<i>T. bairdii</i>
<i>T. terrestris</i>	-			
" <i>T. kabomani</i> "	0.5 % $\pm$ 0.2	-		
<i>T. pinchaque</i>	1.0 % $\pm$ 0.4	1.2 % $\pm$ 0.4	-	
<i>T. bairdii</i>	5.8 % $\pm$ 1.0	6.0 % $\pm$ 1.0	5.9 % $\pm$ 1.0	-
COII				
Taxa	<i>T. terrestris</i>	" <i>T. kabomani</i> "	<i>T. pinchaque</i>	<i>T. bairdii</i>
<i>T. terrestris</i>	-			
" <i>T. kabomani</i> "	1.3 % $\pm$ 0.3	-		
<i>T. pinchaque</i>	1.7 % $\pm$ 0.4	1.2 % $\pm$ 0.4	-	
<i>T. bairdii</i>	7.8 % $\pm$ 1.1	7.7 % $\pm$ 1.0	8.0 % $\pm$ 1.1	-
Cyt-b				
Taxa	<i>T. terrestris</i>	" <i>T. kabomani</i> "	<i>T. pinchaque</i>	<i>T. bairdii</i>
<i>T. terrestris</i>	-			
" <i>T. kabomani</i> "	1.8 % $\pm$ 0.3	-		
<i>T. pinchaque</i>	2.4 % $\pm$ 0.3	2.3 % $\pm$ 0.4	-	
<i>T. bairdii</i>	12.5 % $\pm$ 1.2	12.1 % $\pm$ 1.2	12.4 % $\pm$ 1.1	-

Figure 4 shows the ML tree for the three concatenated mitochondrial genes with the "*T. kabomani*" sequences (both Brazilian individuals reported by Cozzuol et al., 2013 and three specimens sampled by us). Clearly, our tree with 93 specimens showed reciprocal monophyly between *T. terrestris* and *T. pinchaque*. Moreover, the alleged "*T. kabomani*" is, mitochondrially speaking, a clade within *T. terrestris*. Thus, our results do not provide positive data in favor of *T. kabomani* as a new and full tapir species (proof 6). The ML tree with only the barcoding gene *COI* employed for species discrimination, also showed "*T. kabomani*" as a clade within *T. terrestris* (Figure 5; proof 7). This demonstrates reciprocal monophyly between *T. terrestris* and *T. pinchaque*.

The MNJ procedure including the four possible Latin American tapir taxa for the three concatenated genes (Figure 6) showed that the haplotype sets of *T. bairdii*, *T. terrestris* and *T. pinchaque* are well defined. However, the "*T. kabomani*" haplotypes seem to be an extreme extension from those of *T. terrestris* (proof 8).

An individual of *T. terrestris* from the Chiquitania in Bolivia (this was not reflected in any phylogenetic tree) connected the "*T. kabomani*" and *T. pinchaque* haplotypes. Some temporal split estimations were estimated by this procedure and they are interesting and cited here. For example, the temporal split between *T. bairdii* and *T. terrestris* and *T. bairdii* and "*T. kabomani*" were $7,960,000 \pm 69,282$ YA ($\rho = 39.8 \pm 0.346$) and $9,560,000 \pm 144,222$ YA ($\rho = 47.8 \pm 0.721$), respectively. Therefore, on average, the time split between *T. bairdii* and *T. terrestris* + "*T. kabomani*" was around 8.76 MYA. The temporal split between *T. terrestris* and *T. pinchaque* was around $3,420,000 \pm 509,117$ YA ($\rho = 17.1 \pm 2.545$) and the haplotype diversifications within both species were $3,300,000 \pm 141,421$ YA ($\rho = 16.5 \pm 0.707$) and $1,208,696 \pm 194,635$ YA ($\rho = 6.04 \pm 0.973$), respectively. In consideration of these data, the mitochondrial diversification of *T. terrestris* preceded *T. pinchaque* and/or being the subsistence of the original haplotypes in *T. terrestris* higher than in *T. pinchaque*. This last possibility agrees quite well with the fact that the gene diversity is lower in *T. pinchaque* than

in *T. terrestris*. This is probably more due to the greater action in gene drift associated with considerably small population sizes in the first species rather than in the second species. The temporal split between *T. terrestris* and “*T. kabomani*” was around $1,300,000 \pm 360,555$ YA ($\rho = 6.5 \pm 1.802$), considerably lower than that between *T. terrestris* and *T. pinchaque*, ratifying “*T. kabomani*” as a clade within *T. terrestris* (proof 9).

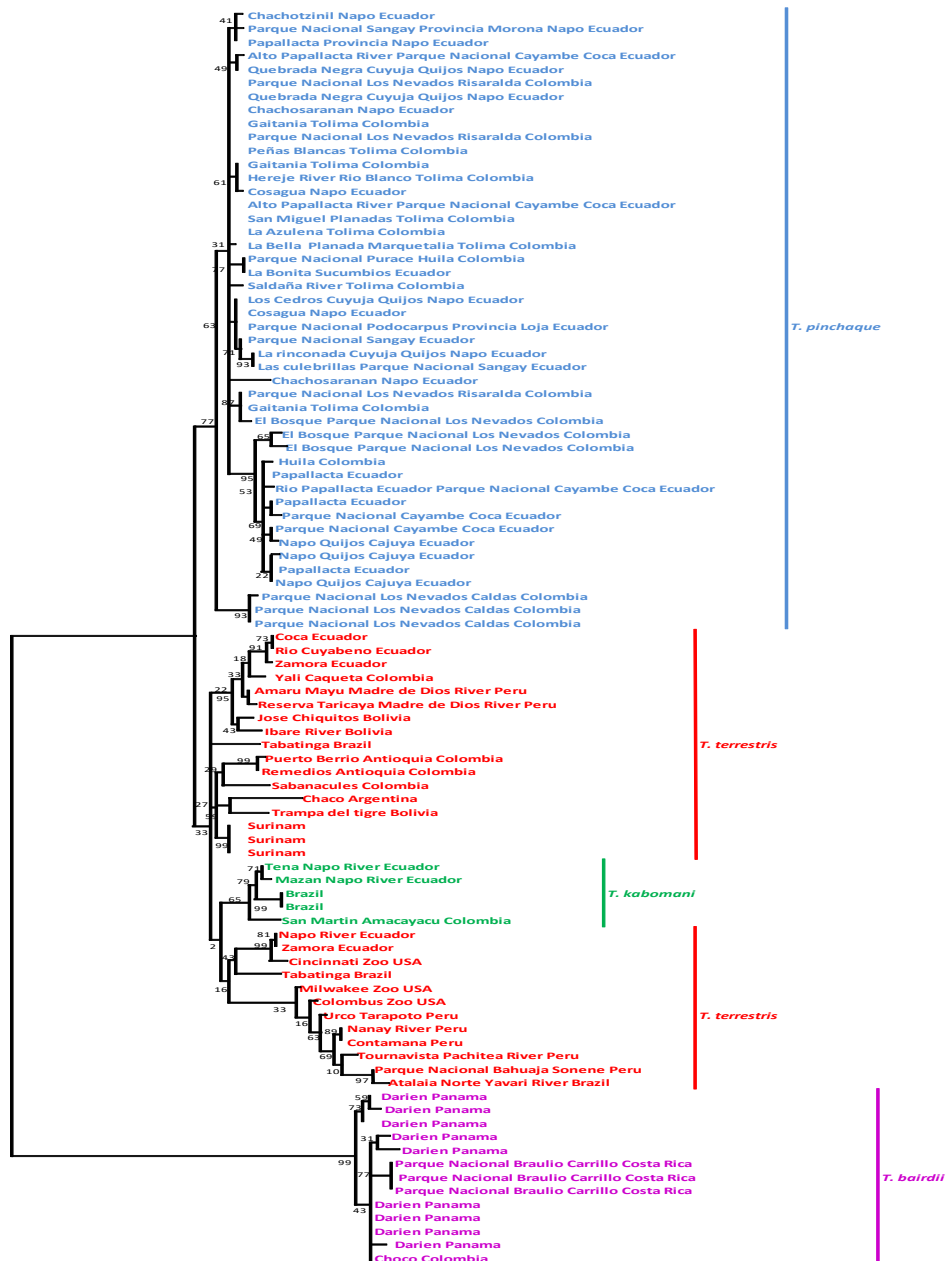


Figure 4. Maximum likelihood (ML) tree for 93 Neotropical tapir specimens (including *T. pinchaque*, *T. terrestris*, *T. bairdii* and the alleged new species, “*T. kabomani*”, reported by Cozzuol et al., 2013) for three concatenated mitochondrial genes (*Cyt-b* + *COI* + *COII*). *T. kabomani* was within the *T. terrestris* clade and therefore these mitochondrial genes did not provide positive evidence for *T. kabomani* as a full species.

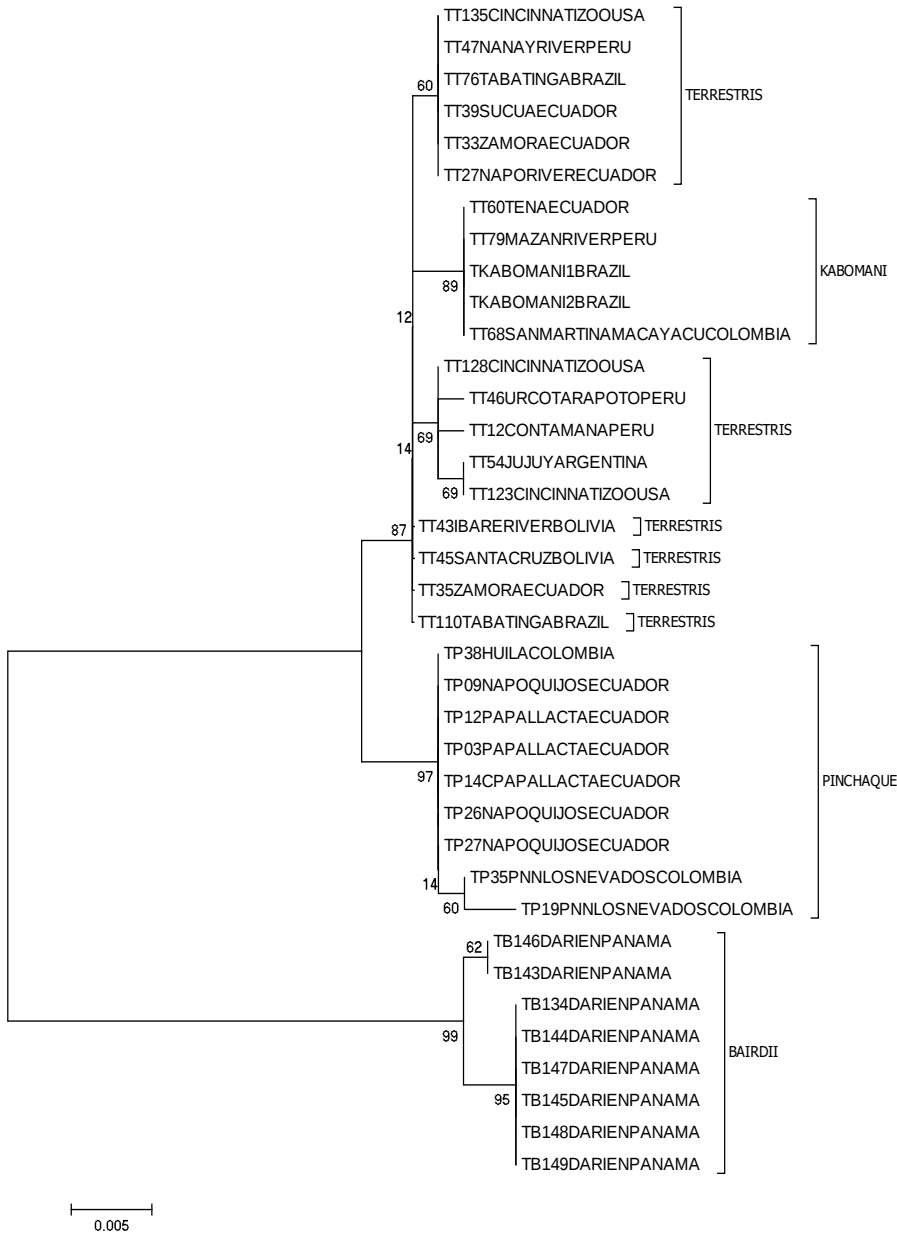


Figure 5. Maximum likelihood (ML) tree for different individuals of the three Neotropical tapir species recognized and the alleged new species reported as “*T. kabomani*” by Cozzuol et al., (2013) at the *COI* gene. “*T. kabomani*” was a clade within *T. terrestris*.

The MJN applied only to *T. terrestris* at the *Cyt-b* gene showed the following picture (Figure 7). There were six well-defined haplogroups. Amazon 0 (which corresponds to “*T. kabomani*” in the other analyses) and Amazon I haplogroups were the most differentiated ones.

The average temporal split among the six haplogroups was around 1.6 ± 0.34 MYA. The beginnings of the temporal diversification within each haplogroup were as follows: 0.59 MYA for the Amazon I, 0.42 MYA for Amazon II and Amazon III, 0.55 MYA for the North and 0.33 MYA for the South.

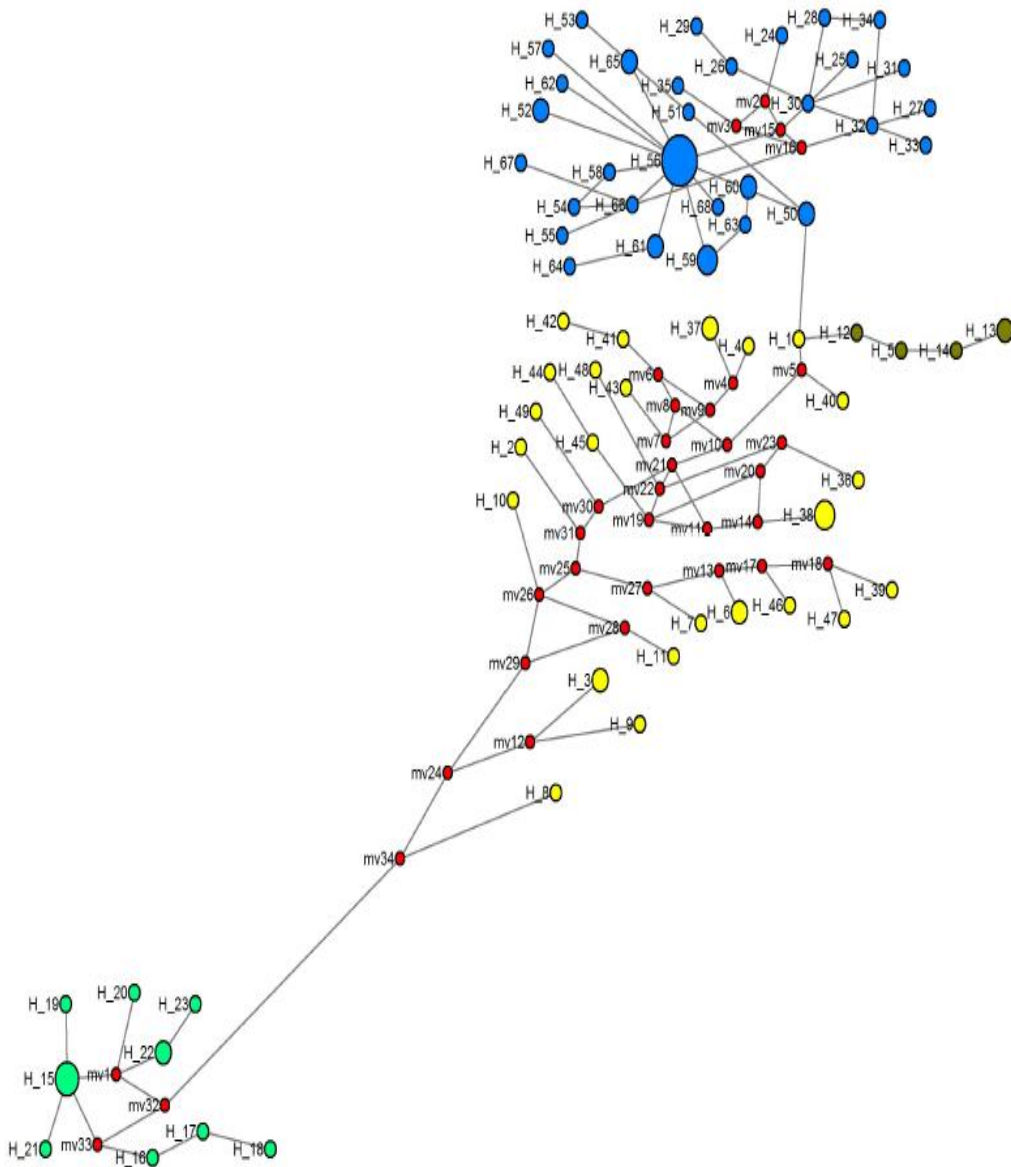


Figure 6. Median Joining Network (MJN) with the mitochondrial haplotypes found for *Tapirus bairdii* (green), *T. terrestris* (yellow), “*T. kabomani*” (brown) and *T. pinchaque* (blue) with the concatenated sequences of three genes (*Cyt-b* + *COI* + *COII*). Red circles are intermediate haplotypes not found. “*T. kabomani*” was a prolongation of *T. terrestris*.

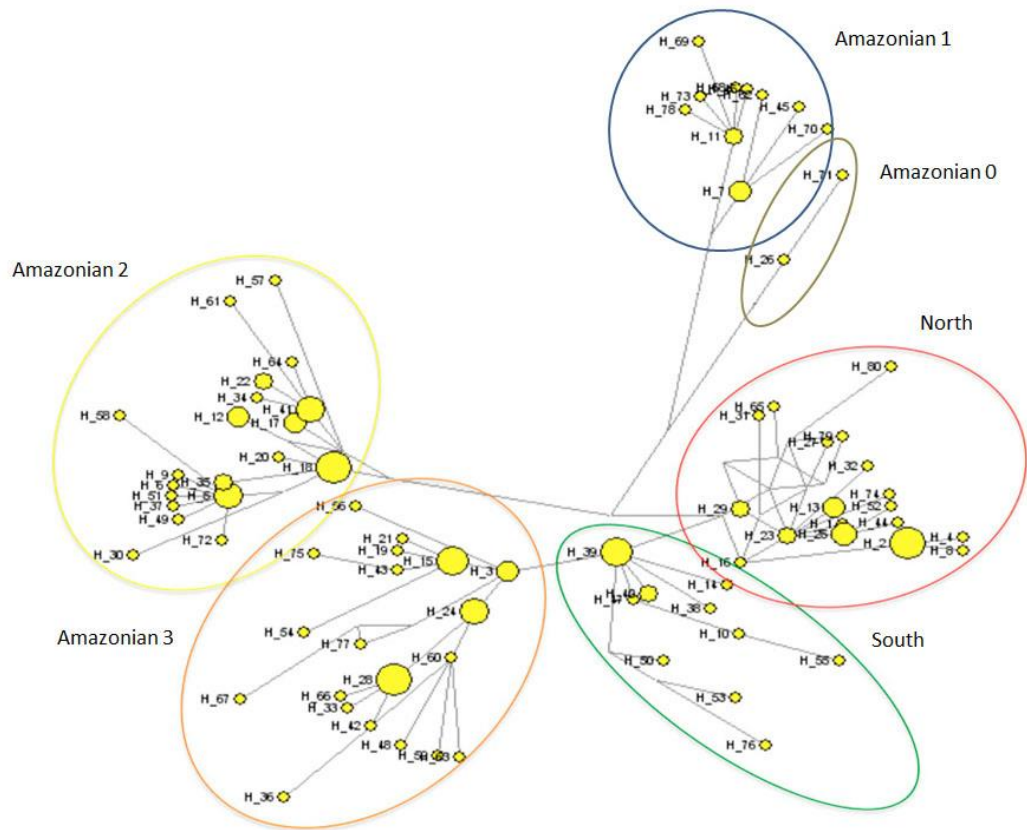


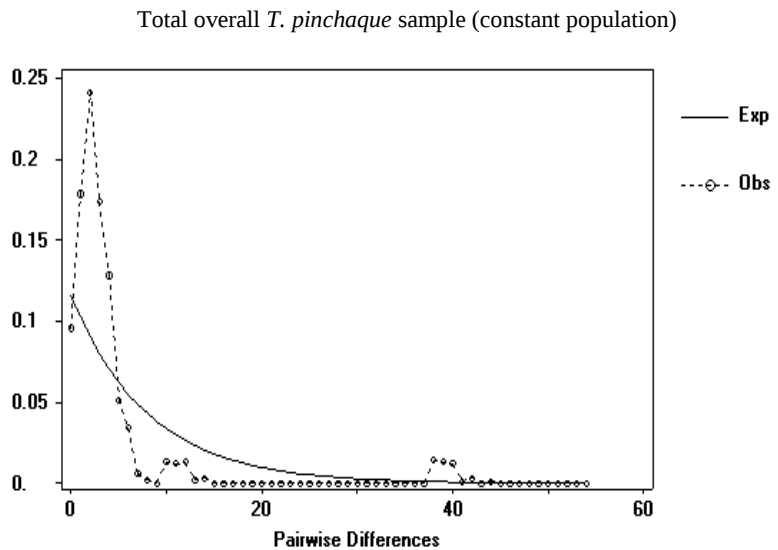
Figure 7. Median Joining Network (MJN) with the 80 haplotypes found at the Cyt-b gene for the 141 lowland tapirs analyzed. Six different mitochondrial haplogroups were found (Amazon 0, I, II, III, North and South). The Amazon 0 haplogroup corresponds to “*T. kabomani*”. Red circles indicate missing intermediate haplotypes.

Table 4. Demographic statistics applied to the overall *Tapirus pinchaque* sample studied, to the Colombian sample and to the Ecuadorian sample. ⁺ P < 0.05; * P < 0.01, significant population expansions

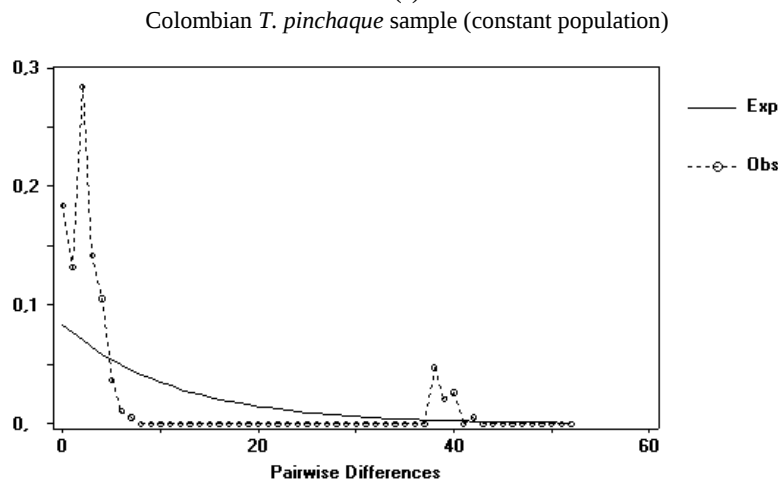
	Tajima D	Fu & Li D*	Fu & Li F*	Fu's Fs	raggedness rg	R2
Overall <i>Tapirus pinchaque</i> sample	P[D ≤ -2.502] = 0.0000*	P[D* ≤ -4.49] = 0.0008*	P[F* ≤ -4.49] = 0.0006*	P[Fs ≤ -7.02] = 0.0032*	P[rg ≤ 0.0251] = 0.1213	P[R2 ≤ 0.0814] = 0.2091
Colombian <i>T. pinchaque</i> sample	P[D ≤ -2.279] = 0.0013*	P[D* ≤ -2.337] = 0.0263 ⁺	P[F* ≤ -2.551] = 0.0240 ⁺	P[Fs ≤ -6.126] = 0.0074*	P[rg ≤ 0.0363] = 0.2276	P[R2 ≤ 0.0923] = 0.0917
Ecuadorian <i>T. pinchaque</i> sample	P[D ≤ -1.825] = 0.0144 ⁺	P[D* ≤ -2.337] = 0.0253 ⁺	P[F* ≤ -2.551] = 0.0227 ⁺	P[Fs ≤ -6.126] = 0.0029*	P[rg ≤ 0.0061] = 0.0111 ⁺	P[R2 ≤ 0.0944] = 0.1517

Demographic Changes in Latin American Tapirs

The mismatch distribution for *T. pinchaque* (Figure 8) showed no clear demographic changes of the total sample or in the Colombian population. But, it did show a significant curve related to a population expansion for the Ecuadorian sample ($rg = 0.006$, $p = 0.011$). For the Ecuadorian population, this expansion began around 2,900 years ago and the initial population size could have been around 4,770–6,691 females. Four of the five demographic change tests (Tajima's D , Fu & Li D^* and F^* , Fu's F) were significant for the total sample. However, the Ramos-Onsins & Rozas R_2 test was not significant for any of the three samples (Table 4).



(a)



(b)

Ecuadorian *T. pinchaque* sample (population expansion)

Figure 8. (Continued).

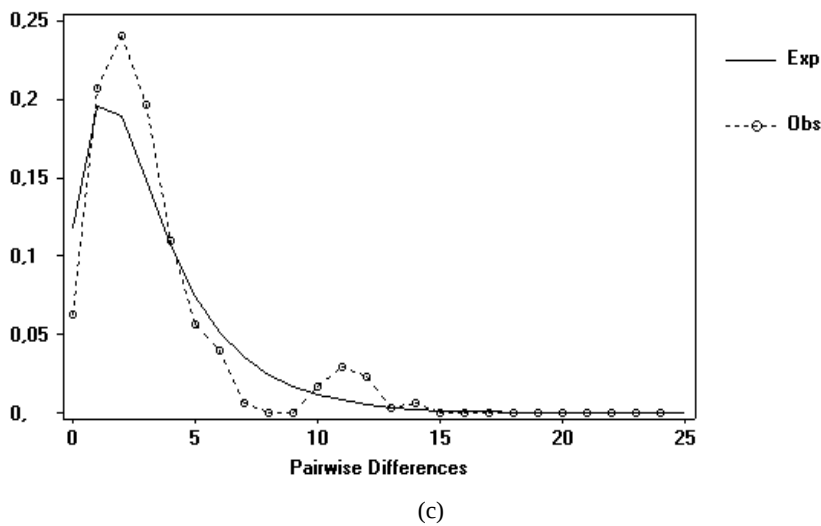


Figure 8. Historical demographic analyses by means of the mismatch distribution procedure (pairwise sequence differences) for the mitochondrial DNA studied in *Tapirus pinchaque*. These analyses were applied to: Overall *Tapirus pinchaque* sample (A); Colombian mountain tapir sample (B); Ecuadorian mountain tapir sample (C).

The mismatch distribution and the six demographic tests carried out showed a clear population expansion for the overall *T. terrestris* sample studied (Figure 9 and Table 5). All six tests were highly significant. There was clear evidence of population expansion in different haplogroups. For Amazon I, the mismatch distribution and two out of six tests were significant.

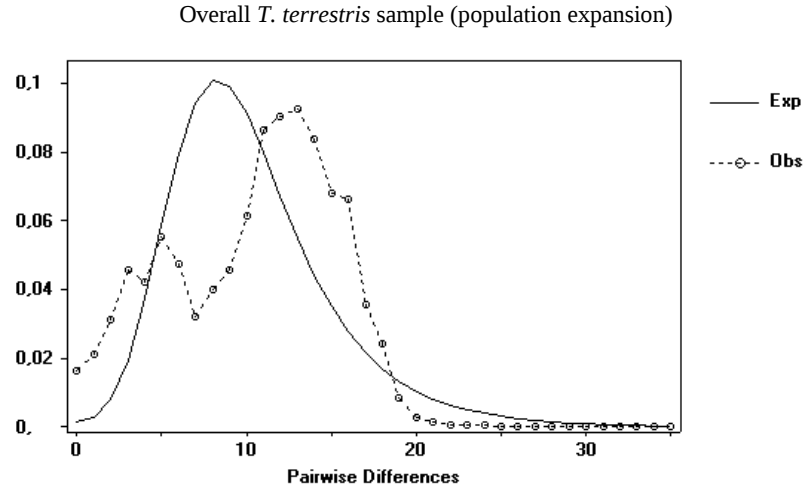
This was the haplogroup where there was less evidence of demographic change. For Amazon II, III and North, the mismatch distribution and four out of six tests were significant. For South, the mismatch distribution and five out of six tests were significant. Thus, we determined noteworthy evidence of population expansion within each haplotype lineage as well as for the total *T. terrestris* distribution. When we only considered two geographical groups (North and South of the Amazon River), there was also strong evidence of population expansion. For the Southern Amazon group, the mismatch distribution and six out of six tests were significant, whereas for the Northern Amazon group, the mismatch distribution and four out of six tests were significant.

Although, the sample of “*T. kabomani*” is modest, we also analyzed possible demographic changes in this taxon (Figure 9 and Table 5). The mismatch distribution and two out of six demographic tests showed evidence of population expansion such as was observed in *T. terrestris*, especially in the Amazon I haplogroup.

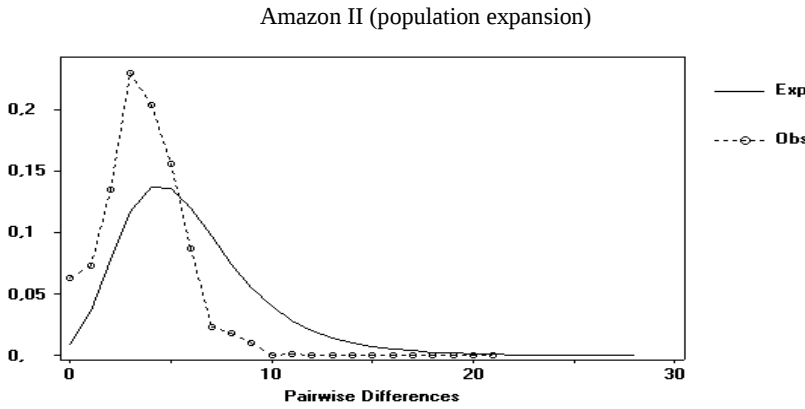
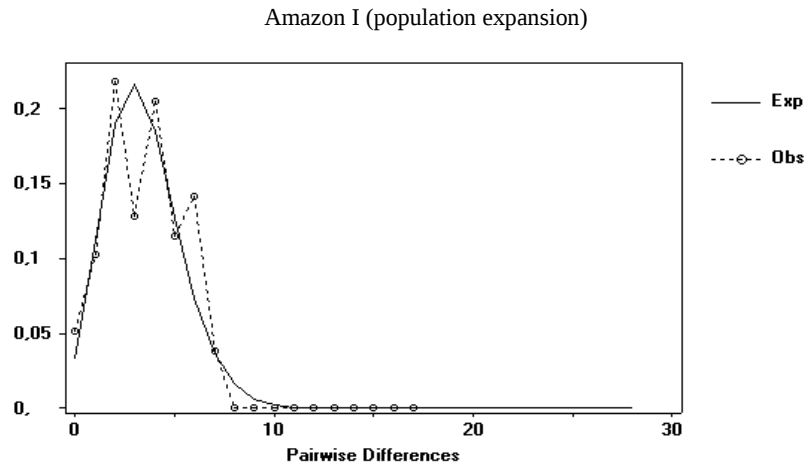
Thus, “*T. kabomani*” should be a dynamic demographic extension of *T. terrestris*, characterized by historical population expansions (proof 10).

Table 5. Demographic statistics applied to the overall *Tapirus terrestris* sample studied, to the haplogroups, to the population separated by the Amazon River and to “*T. kabomani*”. * $P < 0.05$; ** $P < 0.01$, significant population expansions

	Tajima D	Fu & Li D*	Fu & Li F*	Fu's Fs	raggedness rg	R2
Overall <i>Tapirus terrestris</i> sample	$P[D \leq -1.642] = 0.018^*$	$P[D^* \leq -4.00] = 0.004^{**}$	$P[F^* \leq -3.59] = 0.004^{**}$	$P[Fs \leq -34.02] = 0.000^{**}$	$P[rg \leq 0.0035] = 0.0009^{**}$	$P[R2 \leq 0.0453] = 0.031^*$
Haplogroups						
Amazon I	$P[D \leq -0.909] = 0.176$	$P[D^* \leq -1.563] = 0.089$	$P[F^* \leq -1.611] = 0.101$	$P[Fs \leq -4.517] = 0.008^{**}$	$P[rg \leq 0.051] = 0.145$	$P[R2 \leq 0.089] = 0.003^{**}$
Amazon II	$P[D \leq -1.639] = 0.046^*$	$P[D^* \leq -3.307] = 0.002^{**}$	$P[F^* \leq -3.244] = 0.003^{**}$	$P[Fs \leq -9.990] = 0.000^{**}$	$P[rg \leq 0.051] = 0.346$	$P[R2 \leq 0.088] = 0.256$
Amazon III	$P[D \leq -1.818] = 0.015^*$	$P[D^* \leq -3.152] = 0.005^{**}$	$P[F^* \leq -3.198] = 0.013^*$	$P[Fs \leq -9.691] = 0.001^{**}$	$P[rg \leq 0.050] = 0.360$	$P[R2 \leq 0.089] = 0.235$
North	$P[D \leq -1.738] = 0.028^*$	$P[D^* \leq -3.242] = 0.006^{**}$	$P[F^* \leq -3.242] = 0.010^*$	$P[Fs \leq -6.728] = 0.008^{**}$	$P[rg \leq 0.051] = 0.422$	$P[R2 \leq 0.091] = 0.202$
South	$P[D \leq -1.714] = 0.033^*$	$P[D^* \leq -1.687] = 0.050^*$	$P[F^* \leq -1.954] = 0.041^*$	$P[Fs \leq -2.965] = 0.047^*$	$P[rg \leq 0.060] = 0.204$	$P[R2 \leq 0.049] = 0.019^*$
Amazon River as a barrier						
North to the Amazon River	$P[D \leq -1.215] = 0.095$	$P[D^* \leq -2.489] = 0.020^*$	$P[F^* \leq -2.345] = 0.026^*$	$P[Fs \leq -30.424] = 0.000^{**}$	$P[rg \leq 0.006] = 0.033^*$	$P[R2 \leq 0.060] = 0.112$
Southern to the Amazon River	$P[D \leq -1.322] = 0.042^*$	$P[D^* \leq -2.872] = 0.013^*$	$P[F^* \leq -2.769] = 0.012^*$	$P[Fs \leq -6.695] = 0.024^*$	$P[rg \leq 0.006] = 0.003^{**}$	$P[R2 \leq 0.060] = 0.025^*$
“ <i>T. kabomani</i> ”	$P[D \leq 0.30550] = 0.6511$	$P[D^* \leq 0.311] = 0.6404$	$P[F^* \leq 0.3259] = 0.6397$	$P[Fs \leq 1.2411] = 0.6601$	$P[rg \leq 0.01532] = 0.0000^{**}$	$P[R2 \leq 0.079401] = 0.0024^{**}$

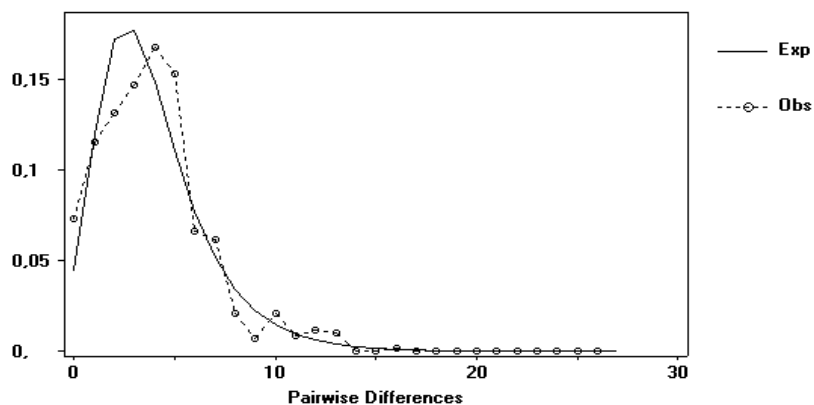


(a)

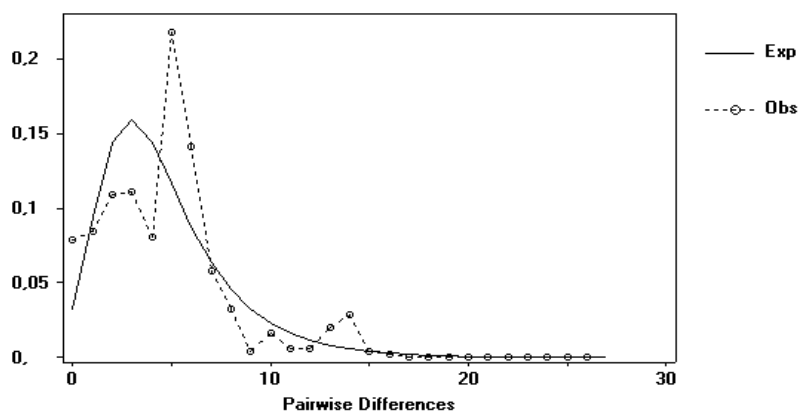


Amazon III (population expansion)

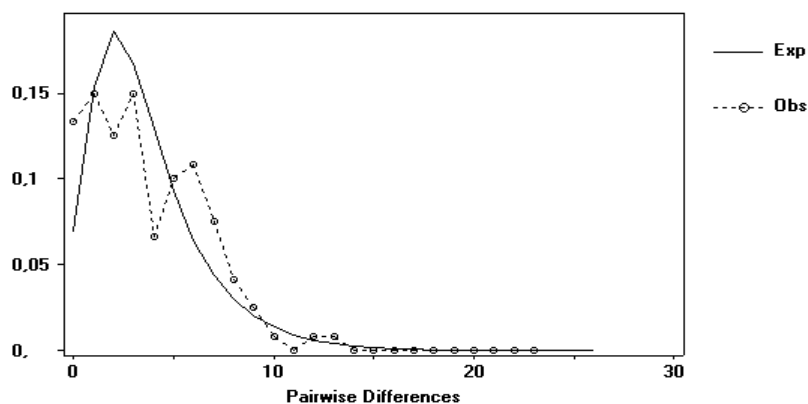
Figure 9. (Continued).



North (population expansion)



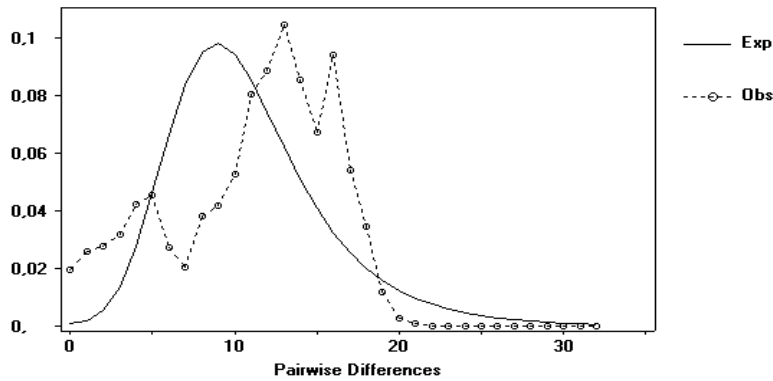
South (population expansion)



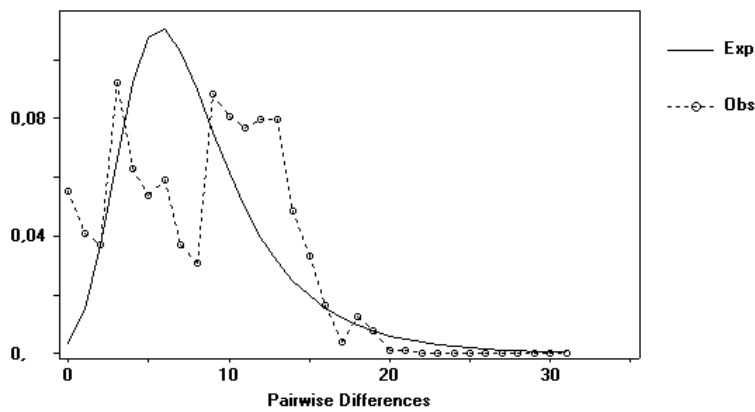
(b)

Northern Amazon River bank (population expansion)

Figure 9. (Continued).

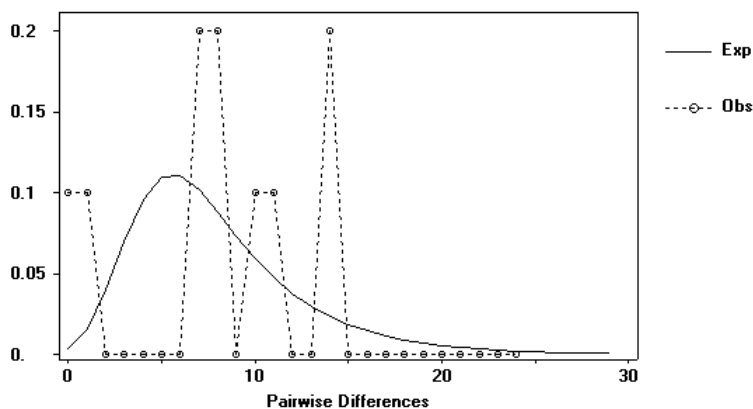


Southern Amazon River bank (population expansion)



(c)

"T. kabomani" (population expansion)



(d)

Figure 9. Historical demographic analyses by means of the mismatch distribution procedure (pairwise sequence differences) for the mitochondrial DNA studied in *Tapirus terrestris* and "*T. kabomani*". These analyses were applied to: Total *T. terrestris* sample (A); Amazon I, Amazon II, Amazon III, North, and South (B); Northern bank of the Amazon River and Southern bank of the Amazon River (C); "*T. kabomani*" (D).

Spatial Structure

For *T. pinchaque*, the overall Mantel test between the log (genetic distance) and the log (geographic distance) showed $r = -0.127$ ($p < 0.691$ from 10,000 randomizations). Therefore, there is no measureable spatial genetic structure in *T. pinchaque*, at least, in the areas sampled in Colombia and Ecuador for mitochondrial DNA. Similarly, for *T. terrestris*, the Mantel test did not detect any significantly positive correlation between the genetic distances and the geographical distances among the exemplars sampled at the global level, within different haplogroups and within certain geographical regions at the *Cyt-b* gene. Some correlations were significant but negative (for the global tapir population studied, $r = -0.169$, $p = 0.001$; Amazon I, $r = -0.439$, $p = 0.001$; Amazon II, $r = -0.213$, $p = 0.001$; Amazon III, $r = -0.107$, $p = 0.023$; North, $r = 0.007$, $p = 0.439$; South, $r = -0.525$, $p = 0.001$; total Colombia, $r = -0.070$, $p = 0.007$; Northern Colombia-Venezuela, $r = -0.134$, $p = 0.001$). Therefore, no isolation by distance patterns were found in any of the *T. terrestris* groups analyzed. Nevertheless, for *T. terrestris*, an AIDA analysis for the entire Amazon basin showed the 1 DC (0-350 km) to be significantly positive. The 4 DC (1,150-1,350 km) had a significantly negative autocorrelation and again a significantly positive autocorrelation at 6 DC (1,500-1,700 km). This correlogram showed the typical resemblance of a double circular cline (Sokal and Oden, 1978), which implied a complex pattern of colonization of this area by the lowland tapir.

In contrast, the *T. bairdii*'s haplotypes showed a very significant spatial genetic trend ($r = 0.819$, approximate Mantel t-test = 3.611, $p = 0.0002$; out of 5000 random permutations, 5000 were $< Z$, $0 = Z$ and $0 > Z$; one-tail probability $p = 0.0004$), an effect of isolation by distance.

DISCUSSION

Genetic Conservation of the Latin American Tapirs

Clearly, *T. terrestris* had the highest gene diversity level of all tapir species. This finding agrees quite well with the fact that the Latin American *Tapirus* species had the widest geographical distribution and therefore the highest potentially effective numbers. Curiously, *T. pinchaque*, a species with a very restrictive geographic distribution and a small population size, presented a gene diversity that was more than three times higher than that of *T. bairdii*. The surprise is due to the fact that *T. pinchaque*, may be on the brink of extinction (Ashley et al., 1996) whereas *T. bairdii*, has historically occupied a distribution from Southern Mexico to the Pacific area of Colombia and maybe Ecuador (Tirira, 2011). Thus, although *T. pinchaque* has a very small census population size and a very restrictive geographical distribution, within disturbed and fragmented ecosystems, the species is not impoverished from a genetics point of view. This is good news from conservation perspective.

The geographical distribution of *T. pinchaque* is still not completely known and there is disagreement over its distribution within the scientific literature. In Colombia, populations of the mountain tapirs are well described in the Central and Eastern Cordilleras (Lizcano et al., 2002; Cavelier et al., 2010). Lizcano et al., (2002) estimated around 2,500 mountain tapirs in Colombia with around 1,874 animals in seven Colombian National Parks. However, Cavelier et al., (2010) estimated a minimum of 543 specimens and a maximum of 579 animals in the

eight Colombian National Parks where they considered that this species still lives. However, the presence of the mountain tapir in the Western Colombian Cordillera is disputed. Acosta et al., (1996) and Downer (1997) recorded locations in this Cordillera where the species could have lived in the past. Sarriá (1993) recorded comments of staff working at Farallones de Cali National Park who observed *T. pinchaque* near Alto de Pance and Alto del Hambre (3,730 masl). Another possible record of the presence of *T. pinchaque* within the Western Cordillera was made by De Wilde (1994), who wrote about footprints in the Caramanta Cerro between the Departments of Antioquia and Risaralda. However, Lizcano et al., (2002) suggested that this species does not inhabit the Western Cordillera. Furthermore, Lizcano and Cavelier (2000) considered that such records within the Western Cordillera may indicate the Baird's tapir (*T. bairdii*) and not *T. pinchaque*. The Baird's tapir could also live within the montane forest of Central America, up to 3,000 masl. However, Arias-Alzate et al., (2010) recently reported the existence of a skull of a *T. pinchaque* in a Medellín museum sampled in 1911 from Páramo de Frontino in the Antioquia Department of Western Cordillera. This provides evidence of *T. pinchaque* living within the northern area of the Western Cordillera.

Lizcano et al., (2002) recorded the presence of the mountain tapir within the Central Cordillera from Los Nevados National Park (Caldas and Risaralda Departments) to the South. They also noted the mountain tapir within the Eastern Cordillera from Páramo de Sumapaz to the South. However, some populations of mountain tapir have been recently detected north of Páramo de Sumapaz within the Eastern Cordillera, including at Chingaza National Park and 12 other northern localities. Páramo de Pisba may be one of these sites (Montenegro, 2002). The current absence of mountain tapirs from Northern Central and Eastern Cordilleras could be the result of hunting activity (Lizcano et al., 2002). The extreme decline of the female population we detected for the Colombian *T. pinchaque* population at the genetic level with a Bayesian sky plot analysis (Ruiz-García et al., 2015b) correlated quite well with the fact of an extreme reduction in the potential habitat space of the Andean tapirs. Cavelier et al., (2010) showed that the current distribution for Andean tapirs covers 31,400 km², compared to 205,000 km² in the past. Similarly, the trend in distribution size for tapirs in Colombia is much lower today (14,385 km²) than in the past (74,556 km²). That is, the current value corresponds to 19% of the past distribution. The area where the species occurs is actually a collection of 35 forest patches. Most of the fragments are small and only four are larger than 1,000 km² (Lizcano et al., 2002). These authors showed that only five to six fragments have the minimum necessary size (826 km²) to maintain at least 150 individuals, the estimated number to maintain a viable population in the short term. Thus, the situation of *T. pinchaque* is critical in Colombia.

In Ecuador, the areas with the highest *T. pinchaque* population sizes are the Cayambe-Coca National Park (Northern Eastern Cordillera) and the Sangay National Park (Central Eastern Cordillera). There are other areas of Ecuador, such as the Reserva Ecológica El Ángel (Carchi Province), Cajas National Park-Bosque Protector Mazán (Azuay Province) and the Condor Cordillera, where the mountain tapirs have lived but where there are no records of them in the last few decades (Downer, 1997; Tirira and Castellanos, 2001).

Cavelier et al., (2010) estimated a population size of 543-579 individuals for Colombia and 983-1,047 individuals for Ecuador. However, the gene diversity of *T. pinchaque* in Colombia is higher than in Ecuador. Dobzhansky (1971) showed that ancestral populations retain more elevated gene diversities than do derived populations. This may support that the original *T. pinchaque* population inhabited the Colombian Andean Mountains and later it expanded toward the southern mountains (Ecuador today). Our analyses detected the population expansion of the

Ecuadorian population. However, more recently, the Colombian population decreased due to hunting and habitat fragmentation (high population decrease in the last 5,000 YA; Ruiz-García et al., 2015b). The Ecuadorian population has been less affected by human activity. Thus, the Colombian population could be in a more dangerous situation than the Ecuadorian population. Another *T. pinchaque* population in critical situation should be the Peruvian population. Cavelier et al., (2010) only estimated between 41 to 43 mountain tapirs in the Peruvian protected Tabaconas Namballe National Sanctuary. Downer's (1997) work suggests the possibility of mountain tapirs inhabiting Eastern Tapal in the Cajamarca Province, south of the Jaén Province, and at the Piura Province. The Peruvian population needs a molecular population analysis.

The gene diversity levels of *T. terrestris* are the highest estimated for the Latin America tapirs and this species seems to not be significantly endangered from a genetics point of view. Also, the major part of the haplogroups showed historical population expansions. However, *T. terrestris* is a species threatened due to the subsistence hunting carried out by the Indian and "colonos" communities in South America. Another threat comes from the deforestation of the Neotropical forests due to extensive agriculture and ranching (Constantino et al., 2006). In many areas of South America, this species is more targeted than other animals by hunters. Here we provide three examples of hunting: 1- Our first example is the case of 22 Indian Izoceño-Guaraní communities in the "chaqueña" region of Bolivia, who intensively hunt tapirs (Barrientos and Cuéllar, 2003); 2-A second example is the case reported by Romero et al., (2013) in the community of Caura (Guarataro), along the Southern Orinoco River in Venezuela. There humans consume 40% of their proteins from hunted animals. Of the 196.6 g of proteins consumed by a human per week, 77 g originated from wild hunts and of these, 14 were from tapirs. Out of 184 tons of wild meat consumed in a year, 44 tons were from tapirs. This means that in one year this community hunted 274 tapirs. 3- Third, it is clear that the tapir provided the most meat in the Peruvian rainforest compared to all other mammals, but unfortunately the hunting model used is non-sustainable. Bodmer et al., (1997a) calculated the percentage of production taken by hunters in the Pacaya-Samiria National Park within the Peruvian Amazon. For example, the hunted red deer (*Mazama americana*), collared peccary (*Pecary tajacu*) and white lipped peccary (*Tayassu pecari*) yielded values of 6.8%, 4.3% and 19.8%, respectively. The authors considered that a value higher than 20% could be dangerous to the continued subsistence of a population. The value for the lowland tapir in that Peruvian area was 166%, which showed that the probability of subsistence of that population was completely unsustainable. Also, Bodmer et al., (1997b) showed that in one year for three areas of the Pacaya-Samiria National Park, 53 *Tayassu pecari*, 12 *Pecari tajacu*, 13 *Mazama Americana* and 20 *Tapirus terrestris* were killed. But, tapir species have the highest biomass (4,000 kg) of all mammals, making up 48% of the ungulate biomass extracted and 35% of the total mammal biomass extracted. Thus, tapirs are critically affected by hunting. Furthermore, this species has a long gestation period of around 13 months, only producing one offspring by litter (Padilla and Dowler, 1994). Tapirs reach sexual maturity at two years of age (Yamini and Schillhorn, 1988), and females only have one offspring every three years. Tapirs also have a very low population density (Bodmer and Brooks, 1997; Bodmer et al., 2000). Bodmer and Brooks (1997) reported an overall density of 0.4 ind/km² for the Peruvian Amazon forests, while Ayala (2003) found a value of 0.5 ind/km² for the Bolivian Chaco. Scientists have offered a range of estimations from 0.028 ind/km² in the Northern Colombian Amazon (Chamorro-Rengifo and Cubillos-Rodriguez, 2007) to 1.2-1.6 ind/km² in the Chiquitanian forest of Bolivia (Arispe et

al., 2007) and Brazil (Schaller, 1983). For this reason, the conservation of tapirs is crucial within Neotropical environments.

The reduction of the lowland tapir populations is observable because of its markedly reduced distribution. One example of this is the putative subspecies *T. t. colombianus*, that has been extirpated from the lowlands of the Caribe region as well as from the Andean valleys of Colombia. This putative subspecies has been isolated in some localities of the Sierra Nevada de Santa Marta (Constantino et al., 2006). *T. terrestris* is classified in the Appendix II of CITES and the IUCN (2003) classified it as Vulnerable. Also, the subspecies, *T. t. colombianus*, is considered to be Critically Threatened (IUCN, 2004). In northern Argentina, the tapir distribution range has been reduced by 46% in the last century by hunting and habitat destruction (Chalukian et al., 2013).

Thoisy et al., (2010) affirmed that the Amazon River acted as a barrier to gene flow in *T. terrestris*. Our results showed that the Amazon River was only a partial barrier for haplotype dispersion for *T. terrestris*. An interesting additional comment related with this is that the jaguar corridor initiative proposed by Rabinowitz and Zeller (2010) and Zeller et al., (2013) could be very useful for the lowland tapirs too. The jaguar conservation units could be very similar and overlap the region with the largest tapir populations (they are the largest carnivore and herbivore mammals in South America). The 182 jaguar corridors proposed by these authors (2.6 million km²) could also be useful in connecting tapir populations especially in the northern and southern geographical areas of South America where the habitat destruction carried out by humans is more intense.

The case of *T. bairdii* seems to be even more dramatic. Although its geographic distribution is wide, it has been dramatically reduced and fragmented in the last two centuries and today no more than 6,000 individuals are left in the wild (Ashley et al., 1996). Its mitochondrial genetic diversity levels were extremely low compared with other Neotropical tapirs. This could mean that this species suffered from a bottleneck and/or the gene drift has been more intense on this species by natural or human constrictions. Indeed, this species has intensely declined in the last century by habitat destruction and hunting and has been extinct in El Salvador and in a major fraction of its original distribution range in Colombian and probably completely extinct in Ecuador (Bodmer and Brooks, 1997).

Systematics of the Neotropical Tapirs and the Case of the Alleged “*T. Kabomani*”

All the molecular studies on tapirs, with the exception of that of Couzzol et al., (2013) and one of the analyses of Thoisy et al., (2010), determined the split between *T. terrestris* and *T. pinchaque* to have occurred during the last Pliocene or at the beginning of the Pleistocene. Ashley et al., (1996), analyzed the *COII* gene and determined the temporal split between the ancestors of *T. terrestris* and *T. pinchaque* to be around 3 MYA. The authors concluded that this coincided with the disappearance of the Bolivar Trough and the apparition of the modern Panamanian isthmus. Norman and Ashley (2000) added more *Perissodactyla* species and other mitochondrial gene (*12S rRNA*) and estimated new temporal splits. The *COII* and *12SrRNA* genes supported divergence times between *T. terrestris* and *T. pinchaque* of 2.5-2.7 MYA and 1.5-1.6 MYA, respectively. More recently, Ruiz-García et al., (2012), analyzed the *Cyt-b* gene and showed a Bayesian tree where the ancestors of *T. terrestris* and *T. pinchaque* diverged

around 3.8 MYA (95% High Posterior Density, HPD: 2.1-4.7 MYA). When a ρ statistic was used on a MNJ network, the two most frequent *T. terrestris*'s haplotypes diverged from the main *T. pinchaque*'s haplotype around 1.58 ± 0.30 MYA and 1.53 ± 0.34 MYA, respectively. Ruiz-García et al., (2015a) showed a Bayesian tree that supported a temporal split of 3.33 MYA between both *T. pinchaque* and *T. terrestris*. Also, Ruiz-García et al., (2015b) performed a mitogenomic study of *T. pinchaque* and determined the temporal split between the ancestors of *T. terrestris* and *T. pinchaque* to oscillate from 7.42 to 3.27 MYA (95% HPD) in a Bayesian tree. In that study also, with the ρ statistic from the MNJ analysis, and with no priors from other paleontological or molecular studies, assuming that *T. pinchaque* was an ancestral taxon with regard to *T. terrestris* (the Hershkovitz, 1954's and Haffer 1970's hypotheses), a split occurred between the taxa around 2.47 ± 0.42 MYA. In contrast, if *T. terrestris* is older than *T. pinchaque*, (Ruiz-García et al., 2012's hypothesis), the temporal divergence between them would have occurred around 3.59 ± 0.79 MYA. Our current work suggests that the split between species occurred 3.4 MYA.

These temporal estimates absolutely disagree with what Cozzuol et al., (2013) claimed. They reported a temporal split within the clade "*T. kabomani*"-*T. pinchaque*- *T. terrestris* ranging from 0.65 to 0.29 MYA (95% HPD, with the split between *T. pinchaque* and *T. terrestris* around 0.1-0.3 MYA following these authors). In another approach, Thoisy et al., (2010), estimated the divergence between *T. terrestris* and *T. pinchaque* to have occurred around 0.33 MYA.

Nevertheless, we are more confident that the first estimates are more accurate than the second for three main reasons. (1) In the current work we used more individuals of all the Latin American tapir taxa and more genes (better gene diversity estimations in each taxa) than the Thoisy et al., (2010)' work. However, we used a mutation rate, in our MJN analysis that was close to the second mutation rate used by Thoisy et al., (2010). They obtained a very recent split between *T. pinchaque* and *T. terrestris* (0.33 MYA). The temporal split estimates provided by Ruiz-García et al., (2015a,b), ranged from 7.4 to 2.5 MYA. They were obtained without any prior or constriction made by the researchers on divergence times or mutation rates. Our current estimate of 3.4 MYA is within of the quoted range. (2) The extremely reduced divergence time put forth by Cozzuol et al., (2013) for the split among "*T. kabomani*"-*T. pinchaque*- *T. terrestris*, and which influences the appearance of "*T. kabomani*", is easily explained by a temporal constriction imposed by these authors in their data. They applied a constriction for the mitochondrial haplotype diversification in *T. terrestris* of 0.13 ± 0.1 MYA because they affirmed that the oldest fossil records of *T. terrestris* date back to the beginning of the fourth Pleistocene glaciation (0.13 MYA). This paleontological constriction helps to explain the difference in temporal divergence between South American tapir species noted by Cozzuol et al. (2013) relative to other studies. These authors claimed that no clear *T. terrestris* fossil records dating older than 130,000 YA have been located, following Tonni (1992). He described a *T. terrestris*'s mandible from the Lujanense age (upper Pleistocene, 130,000 YA) that he collected from the Colon Department at the Entre Rios Province (Argentina). Also, Noriega et al., (2004) and Ferrero et al., (2007), found fossil fragments of *T. cf. terrestris* corresponding to the El Palmar Formation at the El Boyero locality (upper Pleistocene) at the Entre Rio Province (Argentina). But, this does not mean that they don't exist, simply that they have yet to be found. Recently, Holanda and Rincón (2012) reported two tapir remains in Venezuela. One of them was classified as *T. terrestris* from the Zumbador Cave from the upper Pleistocene,

while the second fossil from El Breal de Orocuál was classified as *Tapirus sp.* It's unclear if this last fossil was from the Pliocene or from the early Pleistocene. For example, if this last remain is classified as *T. terrestris*, the Cozzuol et al., (2013)'s constrictor doesn't make sense and therefore the divergence within the clade "*T. kabomani*"-*T. pinchaque*- *T. terrestris* should be considerably older. Indeed, we believe that some of the oldest South American tapir fossil remains classified as *Tapirus sp.* could belong to *T. terrestris*, especially if a population view is adopted more than a typological one, which is more typical of paleontologists. Hulbert et al., (2009) showed that the discovery of 75 individuals of *T. polkensis* in the Gray Fossil Site in Eastern Tennessee indicated a unique species. However, it had considerable intraspecific variation in the development of the sagittal crest, outline shape of the nasals and the number and relative strength of lingual cusps on the P1. These authors concluded that if these fossil remains had been found in diverse geographical areas, they would have been considered different species. Similarly, Perini et al., (2011) demonstrated that the lower molariform teeth size and proportions, used by many authors to define different tapir fossil species, are unreliable because they have great population variability. Thus, some early and middle Pleistocene tapir fossils should be re-assigned and some of them could be un-differentiated from *T. terrestris*. The findings of Perini et al., (2011) don't support the conclusion that all of these *Tapirus* fossils are *T. terrestris*, they simply indicate that some are not, but that others could be. Indeed, in our population criteria and in a context of anagenesis, or phyletic evolution, some fossil of tapirs, like *T. rondoniensis*, *T. cristatellus* and, even, *T. mesopotamicus* could be interpreted as *T. terrestris* individuals with some small morphological differentiated traits. These traits would probably be due to the exposure to different environmental conditions in each moment and in each Pleistocene refugium as well as due to phenotypic plasticity. Another controversial question between genetics and paleontological findings in regard to tapirs is the fact that several authors, from a paleontological point of view, claimed that the South American tapirs do not form a monophyletic group (Holanda and Ferrero, 2013). These authors maintain that *T. pinchaque*, *T. terrestris*, *T. mesopotamicus*, *T. rondoniensis* and *T. cristatellus* are paraphyletic with evolution in situ in South America—deriving from a similar form such as *T. webbi* from North America during the Miocene. They also claimed that another dispersal event occurred from South America to North America by means of a form similar to *T. cristatellus*, which gave origin to the derived forms of North America. This contrasts with the opinion of other authors such as Hulbert and Wallace (2005) and Hulbert (2010), who maintained that the North American forms were more primitive. From a molecular genetics perspective, Ruiz-García et al., (2012), and the current work, found that the two extant tapir species in South America only belonged to a unique migration wave, whilst *T. bairdii* belonged to another different molecular group, older than the first and probably more related to the fossil tapirs from North America. Thus, the molecular results seem to agree better with the second paleontological hypothesis than with the first one.

3- Within the Perissodactyla order, there are no cases of an appearance of new species in the last 0.1-0.3 MYA. In Equidae, Xu (1996), took into account all of the mitochondrial DNA and estimated the initial splitting within *Equus* to have occurred approximately 9 MYA. Tougaard et al., (2001), by means of the 12S rRNA and Cyt-b genes, determined that time split to be 12.2 ± 2.2 MYA. Recently, Orlando et al., (2013), sequencing more than 5,000 genes, determined that all contemporary horses, zebras and donkeys originated 4.0–4.5 MYA. Furthermore, they determined the divergence time between populations of Przewalski's and domestic horses to be approximately 0.38-0.72 MYA (different horse subspecies). In Rhinocerotidae, the split between the current species is estimated to be around

26 MYA (isoenzymes; Merenlender et al., 1989) and 21.5 MYA (12S rRNA and 16S rRNA genes; Morales and Melnick, 1994). The molecular temporal divergence between the two Asian rhinoceros genera was around 25.9 ± 1.9 MYA (Tougard et al., 2001). Their paleontological estimates oscillated from 16 to 23 MYA (Carroll, 1988). The molecular split for the two Asian rhinoceros species of the genus *Rhinoceros* (*R. sondaicus* and *R. unicornis*) was molecularly estimated to have occurred around 11.7 ± 1.9 MYA (Tougard et al., 2001). Their paleontological split has been estimated to have occurred 1.6-3.3 MYA (Carroll, 1988). The two traditional subspecies of the white rhinoceros (*Ceratotherium simum simum* and *C. simum cottoni*) have been estimated to diverge around 1 MYA (Groves et al., 2010). Henceforth, the claim by Cozzuol et al., (2013) of a divergence time of around 0.1-0.3 MYA between *T. terrestris* and *T. pinchaque* does not agree with that determined for other current Perissodactyla taxa.

With regard to the systematics of *T. pinchaque*, our genetics data make a remarkable contribution. Hershkovitz (1954) found a variable trait in the *T. pinchaque* dentition, as manifested by the first upper premolar, which is variable. Ecuadorian specimens show the simple condition, with the cinguloid shelf absent, whereas the first Colombian *T. pinchaque* skull examined shows the premolar as another *Tapirus* species. For this reason, he distinguished two subspecies (*T. p. pinchaque* and *T. p. leucogenys*) within the mountain tapir. However, our molecular results showed that Colombian and Ecuadorian specimens were mixed independently of their geographical origins disagreeing with the fact that they are different subspecies.

With regard to the systematics of *T. terrestris*, the existence of the six mitochondrial haplotype lineages did not agree at all with the four morphological and geographical subspecies determined by Cabrera (1961) and Hershkovitz (1954). That is, there was no correspondence among the mitochondrial haplogroups and the morphological subspecies. In the geographical area of *T. t. aenigmaticus*, we found the six different haplogroups detected in this study. In the wide distribution of *T. t. terrestris*, four haplogroups were found (Amazon II and III, North and South). Thus, there was no correspondence among the subspecies *T. t. anigmaticus* and *T. t. terrestris* and the mitochondrial lineages herein showed. We tentatively negate the validity of these two subspecies. In the territory of *T. t. colombianus*, two haplogroups were detected, Amazonian II and North. Theoretically, the animals sampled in the Colombian Departments of Antioquia, Córdoba and Magdalena (Sierra Nevada de Santa Marta) and at the Zulia (Maracaibo Lake) in Venezuela belonged to *T. t. colombianus*. However, one animal from the Antioquia Department belonged to the Amazonian II haplogroup. Animals from Meta, Vichada and Guainia Departments (Colombian Eastern Llanos and Amazon), French Guyana, Madre de Dios River (southern Peruvian Amazon), Yavari River (Western Brazilian Amazon) and even animals from Northern Argentina shared haplotypes with the North haplogroup. Therefore, there is no linear correspondence between *T. t. colombianus* and the North haplogroup, which questions the reality of this subspecies. The case of these animals from Northern Argentina within this haplogroup is extremely strange. It could be that animals with the same haplotypes migrated from the upper part of the Western Amazon to the northern and southern distribution ranges. And, for this reason, both extreme geographical areas (Northern Colombia and Northern Argentina) share mitochondrial haplotypes. Another possibility, which must be remotely considered, is the fact that during the XIX century and the beginning of the XX century, people imported tapirs from northern areas of South America and liberated them to Northern Argentina

for game hunting purposes (Martínez, personal communication). This could explain the presence of haplotypes of the North haplogroup in Northern Argentina.

All the animals we sampled in Southern Brazil, Eastern Bolivia and Paraguay, which “a priori” belonged to *T. t. spegazzinni* by its geographical distribution and morphological features, belonged to the South haplogroup. Therefore all the southern animals belonged to a unique mitochondrial haplogroup, with the exception of the Argentinian individuals. In this case, it could be a linear correspondence between the South haplogroup and *T. t. spegazzinni*. However, some animals from the Central Brazilian Amazon and from the Colombian Amazon also showed haplotypes of this South haplogroup. It is also interesting to note that traditionally only *T. t. spegazzinni* has been cited in Bolivia (Anderson 1997). In Bolivia, 35 specimens have been located representing 24 localities (Salazar-Bravo et al. 2003). We found many tapirs of Bolivia belonging to the South haplogroup (related with *T. t. spegazzinii*), but we also determined exemplars of the Mamore River which belonged to the Amazonian II lineage typical of more Northern Amazon areas.

Due to the complex spatial distribution pattern of *T. terrestris* garnered through analysis of the *Cyt-b* gene it is impossible to assign tapirs with unknown origin and living in captivity to precise geographical locations. However, we can assign them to one of the six haplogroups we found. For instance, the five tapirs from the US zoos belonged to three out of the six haplogroups (Amazon I, III and South). Also, the exemplar sampled at the Barcelona Zoo belonged to the South haplogroup. Moreover, the animal of unknown origin we analyzed was assigned to the North haplogroup. It was a unique tooth that had no recorded origin. This complex spatial structure could complicate the application of some biological conservation concepts, such as ESUs (Evolutionary Significant Units) and MUs (Management Unit) (Moritz and Faith, 1998) to *T. terrestris*.

Our mitochondrial data totally disagree with the results of Cozzuol et al., (2013). They claimed that “*T. kabomani*” was a full species. We provide 10 population genetics proofs against this claim. All of our trees, as well as those from Ruiz-García et al., (2015a,b,c) showed that “*T. kabomani*” is a particular lineage within *T. terrestris*. In a mitogenomic study with 15 genes and more than 250 *T. terrestris* individuals, Ruiz-García et al. (2015c) showed reciprocal monophyly between *T. terrestris* and *T. pinchaque*. Also, “*T. kabomani*”, together with the Amazon I haplogroup, were the first haplogroups to diverge within *T. terrestris*. In the current work, with three mitochondrial genes, the genetic distances between *T. terrestris* and “*T. kabomani*” were always lower than the genetic distances obtained between *T. terrestris* and *T. pinchaque*.

Kartavtsev (2011) analyzed sequences of the *COI* gene from 20,731 vertebrate and invertebrate animal species. He estimated the average distance data for five different groups. He obtained $0.89\% \pm 0.16\%$ for populations within species, $3.78\% \pm 1.18\%$ for subspecies or semispecies, $11.06\% \pm 0.53\%$ for species within a genus; $16.60\% \pm 0.69\%$ for species from different genera within a family and $20.57\% \pm 0.40\%$ for species from separate families within an order. For this gene, the genetic differentiation between *T. terrestris* and “*T. kabomani*” was only 0.5%, clearly within the status of populations within species. Ascunce et al., (2003) and Collins and Dubach (2000) reported for Primates at the *COII* gene, an average genetic distance around 5.82% among species within a genus and around 2-4% for subspecies. For this gene, the genetic differentiation between *T. terrestris* and “*T. kabomani*” was only 1.3%, clearly within the status of populations within species. Bradley and Baker (2001) claimed, for the *Cyt-b* gene, that values $< 2\%$ would equal intra-specific variation, values between 2% and 11%

would merit additional study, and values >11% would be indicative of specific recognition. For this marker, the genetic differentiation between *T. terrestris* and “*T. kabomani*” was 1.8%, clearly within the status of intra-specific variation. The surprising question is the small genetic distances between *T. pinchaque* and *T. terrestris*, because these taxa have traditionally been considered full species. Thus, considering molecular genetic distances, *T. pinchaque*, should be considered a sub-species of *T. terrestris*. However, we consider *T. pinchaque* a full species because no natural or captivity hybridization has been reported due to reproductive barriers via stasipatric speciation with chromosomal changes (White, 1968, 1978).

The inaccurate result obtained by Cozzuol et al., (2013) was probably due to the fact that these authors only analyzed five samples of *T. pinchaque* at the *Cyt-b* gene and only one sample at the *Cyt-b* + *COI* + *COII* genes. Indeed, they only analyzed three samples at the *Cyt-b* gene, because two *T. pinchaque* samples were repeated due to that the two samples of *T. pinchaque* that M. Ruiz-García donated to the Cozzuol’s team were also donated by another scientist (that shared samples of the same animals with M. Ruiz-García) to B de Thoisy, who latter added his *T. pinchaque*’ results to those of Cozzuol et al., (2013). The very small *T. pinchaque* sample used by Cozzuol et al., (2013) probably did not represent all of the mitochondrial gene diversity of *T. pinchaque*. This contrasts with the larger *T. terrestris* sample containing animals from a very wide geographical area and thus retaining a major fraction of the mitochondrial gene diversity of this last species, and as the genetic distances between both *T. terrestris* and *T. pinchaque* were very small, by chance the no representative mitochondrial gene diversity of *T. pinchaque* of the small sample of Cozzuol et al., (2013) was nested within the gene diversity of *T. terrestris*. Nevertheless, as soon as we enlarged the *T. pinchaque* sample, collected from a wider sampling area, this phenomena disappeared.

Cozzuol et al., (2013) also provided alleged evidence of morphological and morphometric differences of “*T. kabomani*” from the remaining living and fossil South American tapir species. In this work, we provide molecular proofs that do not support the highly speculative and inconclusive conclusions of Cozzuol et al., (2013) that favor “*T. kabomani*.” We also provide several comments on the alleged morphological differences between “*T. kabomani*” and *T. terrestris*. In a recent study, Ruiz-García et al., (2015c) analyzed several *T. terrestris* populations throughout Northern Colombia and different areas of the Amazon basin in Colombia, Peru, Brazil and Bolivia from a craniometric perspective (around 160 skulls). Most of these populations showed significant statistical differences regarding the two areas of the skull where Cozzuol et al., (2013) found the greater divergence of “*T. kabomani*” with regard to *T. terrestris* (the position of the frontal-parietal suture related with the beginning of the sagittal crest and frontals broad and more inflated behind the nasals in “*T. kabomani*”). The significant differences of these skull from those studied by Ruiz-García et al., (2015c) within diverse *T. terrestris* populations were related to different ontogenetic patterns within each population as well as to the age composition of each population. The statistical differences were extreme among some of the seven *T. terrestris* populations studied for different morphometric traits (1- the Napo River area in the Northern Peruvian Amazon, 2- Yavari River and neighborhoods in the Amazonian border of Colombia, Brazil and Peru, 3- Sierra Nevada de Santa Marta in Northern Colombia, 4- Meta, 5- Caqueta, and 6- Vichada departments, all three in Colombia, and 7- the Mamore River in the Bolivian Amazon). However, we don’t claim that each one of these populations was a different tapir species, such as Cozzuol et al., (2013) claimed with “*T. kabomani*”.

Additionally, we show two other morphological proofs which don’t support the claims of Cozzuol et al., (2013). For example, supposedly *T. kabomani*,” with its lower sagittal crest and wider post-nasal exposition of frontals, is much smaller and has a darker color than the “typical” *T. terrestris*. In Figure 1A, we

show a tapir captured in the Amacayacu River (Colombian Amazon) with a “*T. kabomani*” mitochondrial haplotype. We also show, in Figure 1B, a “typical” morphological *T. terrestris* individual from Tena, at the upper Napo River in Ecuador. But, it had a “*T. kabomani*” haplotype. Both animals exhibited different morphotypes to that claimed by Cozzuol et al., (2013) for “*T. kabomani*”. For example, the Ecuadorian animal showed a considerable size and its color was brown. In contrast, as we show in Figure 10, in our travels to obtain tapir samples, we have found small-sized lowland tapirs. Figure 10A shows a pair of small-sized tapirs sampled in the Colombian Amazon, whereas Figure 10B shows a small tapir sampled in the Ecuadorian Amazon. We inspected the dentition of all three exemplars and determined them to be adults because each had at least two erupted molar. Nevertheless, no one showed “*T. kabomani*” haplotypes. The Colombian exemplars belonged to the Amazon II and III haplogroups, while the Ecuadorian individual belonged to the Amazon I haplogroup. Indeed, these animals were of a size smaller than the two individuals in the photo published by Cozzuol et al., (2013) and that were defined as “*T. kabomani*.” Additionally, and mistakenly, they affirmed the existence of photos of “*T. kabomani*” from Colombia. This was a mistake because we provided the authors with two Colombian samples that they claimed to be “*T. kabomani*.” Furthermore, we did not provide any photos of these animals. Most likely, they never asked for the morphology of these exemplars to verify a link between these “*T. kabomani*” haplotypes and the size and dark coat color of these individuals. Complementary, we have observed some lowland tapir adults with small sizes and certain bones deformations as if they were affected by malnutrition or some kind of genetic syndrome (for instance, small head and body but large feet).



(A)

Figure 10. (Continued).



(B)

Figure 10. Some *T. terrestris* adult exemplars with very small sizes but without “*T. kabomani*” mitochondrial haplotypes. Two small adult individuals from Leticia, Colombian Amazon (A); one small adult individual from Macas, Ecuadorian Amazon sampled by the first author (B).

Dobzhansky (1937), in his seminal book for the Neodarwinism synthesis, showed that in evolution the approaches should be of a population nature rather than a typological one. It is quite understandable that a paleontologist (such as is the first author of Cozzuol et al., 2013), working with the scarce material that provides the fossil record on most of the occasions, keeps a typological morphology vision of the species concept. Indeed, many paleontologists only accept a cladogenetic mechanism of speciation following the principle of the punctuated equilibrium theory (Eldredge and Gould, 1972; Gould and Eldredge, 1993). This means that the punctuated equilibrium can become tautological because paleontologists define fossil species on the basis of morphological change. So, it is a trivial problem to observe a strong correlation between speciation and morphological changes. However, population geneticists and evolutionary biologists (including Darwin) know the existence of rapid morphological evolution in a hypothetical ancestral form without speciation (phyletic transformation or anagenesis). Therefore, the existence of a certain morphological and morphometric skull variability in the lowland tapirs did not mean the real existence of different species. Furthermore, the “*T. kabomani*” case is a current living case, and not a paleontological one. With a living species, there are many other aspects to be analyzed in addition to changes in morphology before concluding on the existence of a new species. And, changes in morphology are the only aspects that can be detected in the fossil record. This project and other studies show that the samples sizes and the interpretation of mitochondrial and craniometric data by Cozzuol et al., (2013) are insufficient to define a new tapir species in the Amazon. Additionally and unfortunately, Cozzuol et al., (2013) did not provide sufficient geographical (they contrarily argue that “*T. kabomani*” is sympatrically extended by all the Amazon with *T. terrestris*),

ecological, ethological or reproductive (post and/or prezygote) barrier information to claim to “*T. kabomani*” as a new full species. This does not mean that other tapir species do not exist, but that the provided data are insufficient and misunderstood. In what-ever-the-case, nuclear DNA and especially comprehensive karyological studies (due to the possibility of stasipatric speciation; White, 1968, 1978) should be carried out before claiming the existence of new species of tapirs.

The Impact of the Pleistocene Changes on the Phylogeography of *T. pinchaque* and *T. terrestris*

The temporal split between the ancestors of *T. terrestris* and *T. pinchaque* occurred around 3.4 MYA. And, the beginning of the *T. terrestris* mitochondrial diversification was around 3.1-3.3 MYA. These times match up well with the end of the Pliocene and the beginning of the Pleistocene (2.6-1.8 MYA; Van der Hammen, 1992), which were extremely cold. The average temperature descended $4^{\circ}\text{C} \pm 2^{\circ}\text{C}$ and rainfall declined (500-1000 mm less than today) (Van der Hammen, 1992, 2001). The Pleistocene had 25 main glacial-interglacial cycles with climatic cycles changing every 20,000 Y for the first million years and thereafter every 100,000 Y. These climatic cycles had oscillations in temperature of 10°C at 2,500 m above sea level (masl). This first lowland tapir haplotype split (3.0-3.3 MYA) also coincided with the last formational phase of the Central Andes. The entire Andean chain between Cajamarca and Huancavelina was formed during this epoch by volcanic activity. The lagoons in Huancho (Northern Lima) and the bay in La Ventanilla Beach also formed during this time (León-Canales, 2007). The diverging process of the Amazon 0 haplogroup or “*T. kabomani*” and the haplotype diversification within *T. pinchaque* began around 1.3-1.2 MYA. This coincides with the Pre-Pastonian glacial period (1.3-0.8 MYA), which was extremely dry and cold. While the Buenos Aires province where mammalian fauna (*Categorius*, *Cyonasua*, *Clyomys*, etc) lived was basically tropical from 2-1.3 MYA, this climate later changed. This area was Patagonian Steppe-like around 1.3-0.8 MYA which agrees quite well with a dry and cold climate (Forasiepi et al., 2007). Examples of mammalian fauna living at this time were guanaco, *Lestodelphys* and *Lyncodon*. The diversification process within some of the *T. terrestris* haplogroups occurred around 0.5-0.3 MYA which showed that the same climatic events affected these lowland tapir haplogroups. This correlates to the beginning of the Kansas-Mindel glacial period, 0.4 MYA. Much of the haplotype proliferation we note in the haplogroups identified occurred around 0.2-0.06 MYA and in 0.03-0.01 MYA. This can be explained by the extreme climatic changes in these epochs. The last glacial-interglacial super-cycle began around 0.13-0.15 MYA (Würm. Wisconsin, Vistula, Gamblense). From 130,000 to 80,000 YA (Eemense period), the climate was hot, the precipitation greater, and there were extensive swampy forests of *Aliso*, *Vallea* and *Weinmannia* (Van der Hammen, 1992). But, an extremely cold climate (upper Pleniglacial period) began around 70,000 YA which could have had a fragmentation effect on the lowland tapir haplogroups. From 30,000 YA to 10,000 YA there were many periods of extreme dry and cold conditions. For example from 30,000 YA to 23,000 YA, there was a very cold period in Europe named Dryas I. During this period, the Fuquene Lagoon near the current Bogota (Colombia) dried (29,000 YA; Van der Hammen, 1992) and Mac Neish (1979). This determined changes in the soil acidity and the types of pollen at the Pikimachay Cove found within Peru which are related to the extreme cold 23,000 YA. Later, around 19,000-16,000 YA,

the Last Glacial Maximum (LGM or Dryer II) occurred, which had the most extensive distribution of snow and ice in the Central Andes within the last 200,000 YA. Metivier (1998) estimated that about 18,000 YA, the glacial extension in North and Central Andes was around 371,306 km², whereas it is currently around 3,220 km² (only 1% of the LGM). Also, there were intense cold periods from 14,000 YA to 10,000 YA (with alternating hot periods). Rodbell and Seltzer (2000) found moraines (the front of a glacial) up to 3,170 masl (12,000 YA) at the Cordillera Blanca (San Martin Department, Peru). Comparatively, today the glacial front is 4,600 masl. Also, Rodbell (1991) determined that the other areas of the Cordillera Blanca within the limits of the Departments of Huánuco and Ancash (17 glaciers) had snow lines at 4,200 masl, whereas today this limit is 500-900 m higher. Similarly, Seltzer (1987, 1990) determined the snow lines in the Huaytapallama, Junin and Huancavelina Departments (Peru) 10,950 YA to be 1,400 m lower than today. These climatic conditions could have caused the last lowland tapir haplotypes to diversify.

Several authors (Ab'Saber, 1982; Brown et al., 1974; Fjeldsä, 1994; Haffer, 1969, 1974, 1982, 1987, 1997, 2008; Prance, 1982, 1996; Prance and Lovejoy, 1985; Terborgh, 1992; Van der Hammen, 1975; Vanzolini, 1970, 1973, 1992; Vanzolini and Williams, 1970; Whitmore and Prance, 1987) claimed that these climatic events and subsequent dry/humid cycles created in the Amazon basin are the result of Milankovitch cycles. During the dry periods rainforest communities split into isolated refugia separated by savannah or arid pampas. This hypothesis was originally established for the Pleistocene, but later expanded to include the Miocene-Pliocene periods as well and it was named the Refugia Hypothesis. This hypothesis could explain the appearances of the different haplogroups within *T. terrestris*. During each dry period the haplogroups were formed, but later during the humid periods the tapirs migrated allowing its different haplogroups to diversify in different geographic areas. Along with this, the relationship between the original refugia areas was lost which helped create the geographical distribution of tapirs we see today. However, some Pleistocene refugia could have played an important role in the origins of the haplogroups. For example, the Amazon I and the "*T. kabomani*" haplogroups could have radiated from the Napo refugia (Northwestern Amazon), while the Amazon II and III haplogroups could have radiated from the Napo and Inambari refugia (Southwestern Amazon). The South haplogroup could have expanded from the Inambari or from the Rondonia refugia (between the Madeira and the Tapajos rivers), whilst the North haplogroup could have originated from the Bolivar refugia within the Guiana area. Another hypothesis, called the Recent Lake hypothesis (Klammer, 1984; Marroig and Cerqueira, 1997; Nores, 1999, 2004; Sombroek, 1966), could also support the formation of these lowland tapir haplogroups. This hypothesis claims that most of Amazonia was covered by a huge lake or lagoon during the Pliocene, and successively smaller portions of Amazonia were covered during a series of assumedly high sea level events during the Pleistocene. These created different islands and archipelagos within this Amazon lake. The different haplogroups could have been created on these islands. However, like the tapirs, who are exceptionally good swimmers, these haplogroups could have migrated throughout aquatic systems and expanded their original geographical areas. Therefore, both the Refugia and the Recent Lake hypotheses could support the existence of lowland tapir haplogroups. The significant and complex spatial structure found for the lowland tapirs in the Amazon basin by means of AIDA agrees quite well with both of these hypotheses.

Using Bayesian sky plot analyses, we determined the population changes over the last few thousand years for some of the lowland tapir haplogroups and geographical areas studied.

Amazon I suffered a population decline 5,500 YA coinciding with one of the dry periods of the Holocene after the Optimum Climaticum (OP). Following Rothlisberger (1987), around 6,300 YA, there was a significant increase in temperature especially in Southern South America as well as in the Central Andes, evidenced by O_{18} levels in Huascaran snow. This dry period around 5,500 YA was detected in the Amazon, Caqueta and lower Magdalena River basins as well as in Andean lagoons in Colombia and Peru (Thompson et al., 1995; Van der Hammen & Cleef, 1992; Van der Hammen, 2001). However, in the last 1,400 years, this haplogroup again increased. Amazon II and III revealed a population decrease 75,000 YA coinciding with the ending of the Eemian inter-glacial period and the beginning of the upper Pleniglacial period (Würm I; Van der Hammen, 1992). These two Amazon haplogroups began to again expand around 12,000 and 14,000 YA respectively. This could have been the result of finishing the coldest and driest period of the fourth Pleistocene glacial period, a period that is also related to the last major extinction period for mammals.

In contrast, North decreased 14,000 YA, agreeing with the massive extinction of mammals across the Earth including in South America. This corresponds with the Younger Dryas (Dryas III), typical of Northern Europe and Scandinavia (Clapperton, 1993). This means that the end of this extremely cold and dry period was not the same for all of South America and therefore the diverse haplogroups of tapirs were differentially affected. The Amazon's climate was not very affected by these drastic conditions—leading to an increase in some Amazon haplogroups. However, in Northern South America, the Younger Dryas was more drastic and the North haplogroup was negatively affected. Nevertheless, this lineage increased around 3,000 YA, just when the average temperature reached a value similar to what it is today (Van der Hammer, 1992). Finally, South suffered a population declination around 6,000 YA coinciding with the drier period between 7,000–5,500 YA that we commented on above.

When the data were analyzed by geographical area, only one region showed an important population declination in the last 5,000 years. This was the lowland tapir population of Northern Colombia and Northwestern Venezuela. This strong population decrease, detected by genetic methods, agrees quite well with the situation of this population, which was considered Critically Threatened in 2004 by the IUCN (Constantino et al., 2006).

In contrast to what was expected, in the Andes Mountains, the Pleistocene climatic changes did not create a phylogeographic structure in *T. pinchaque*. This is a conundrum because in other Andean mammals, such as the Pampas cat (*Leopardus pajeros*) (Cossíos et al., 2010; Ruiz-García et al., 2013), the Andean cat (*L. jacobita*) (Cossíos et al., 2012; Ruiz-García et al., 2013) or the spectacled bear (*Tremarctos ornatus*) (Ruiz-García, 2003, 2013; Ruiz-García et al., 2003, 2005), the phylogeographic structure for different kinds of molecular markers were significantly marked. This may support that the haplotype diversification within the current *T. pinchaque* occurred more recently than it did in *T. terrestris*. It's also possible that the gene flow capacity of *T. pinchaque* was relatively higher for this species in the Andean cordillera than for *T. terrestris* in the Amazon lowlands. The population expansion detected in the Ecuadorian *T. pinchaque* population could correlate with this last idea. If the mitochondrial diversification of *T. pinchaque* occurred much more recently than in *T. terrestris*, this could agree with an event of anagenesis, or phyletic evolution, with the original ancestor more related to *T. terrestris* than to *T. pinchaque*. *T. pinchaque* could have appeared via adaptation and exploitation of a new habitat within the Andean Highlands from a lineage, or population, of *T. terrestris*. It could have also appeared via stasipatric speciation. However, all of our phylogenetic trees do not agree with this perspective.

To complement these new mitochondrial results future studies should include sequences of HLA markers, autosomic and sexual chromosome introns and other nuclear DNA genes. Also, future studies focus on sequencing of tapir fossil remains may be helpful in our understanding of the evolution of the current mega-herbivores of South America.

ACKNOWLEDGMENTS

The authors are grateful for help and permissions from the Ecuadorian Ministerio del Ambiente, Ecofondo, Ecociencia, Zoological Foundation of Ecuador, Instituto para la Conservación y Capacitación Ambiental (ICCA), Tapir Specialist Groups-from Colombia and Ecuador, Andean Bear Foundation in USA and Ecuador and to the Von Humboldt Institute in Villa de Leyva (Colombia). Thanks are also given to Carlos Urgiles, Freddy Gallo, Leopoldo Gómez, Fabian Ascantá and Olimpo Gómez to help to obtain samples of *T. pinchaque* in Cuyuja and in Oyacachi (Ecuador). Also thanks to Dr. Diana Alvarez, Pablo Escobar-Armel, Luisa Fernanda Castellanos-Mora, the Peruvian Ministry of Environment, PRODUCE (Dirección Nacional de Extracción y Procesamiento Pesquero from Peru), Consejo Nacional del Ambiente and the Instituto Nacional de Recursos Naturales (INRENA) and to the Colección Boliviana de Fauna and CITES Bolivia (Dr. Julieta Vargas) for their respective help in obtaining *T. terrestris* samples during the last 18 years in Colombia, Perú and Bolivia. Also thanks go to the Ticuna, Yucuna, Yaguas, Witoto and Cocama Indian communities in the Colombian Amazon, Bora, Ocaina, Shipibo-Comibo, Capanahua, Angoteros, Orejón, Yaguas, Cocama, Kishuarana and Alama in the Peruvian Amazon, to the Sirionó, Canichana, Cayubaba and Chacobo in the Bolivian Amazon and Marubos, Matis, Mayoruna, Kanaimari, Kulina, Maku and Waimiri-Atroari communities in the Brazilian Amazon for helping to obtain *T. terrestris* samples.

REFERENCES

- Ab'sáber, A. N. (1982). The paleoclimate and paleoecology of Brazilian Amazonia. In Prance, G. T. (Ed.). *Biological Diversification in the Tropics*. New York: Columbia University. Pp. 41-49.
- Acosta, H., Cavelier, J., Londoño, S. (1996). Aportes al conocimiento de la biología de la danta de montaña, *Tapirus pinchaque* en los Andes Centrales de Colombia. [Contributions to the knowledge of the biology of the mountain tapir, *Tapirus pinchaque* in the Central Andes of Colombia]. *Biotropica* 28: 258-266.
- Ashley, M. V., Norman, J. E. Stross, L. (1996). Phylogenetic analysis of the perissodactylan family Tapiridae using mitochondrial cytochrome c oxidase (COII) sequences. *Journal of Mammalian Evolution* 3: 315-326.
- Anderson, S. (1997). Mammals of Bolivia, Taxonomy and Distribution. *Bulletin of the American Museum of Natural History* 231: 1-652.
- Arias-Alzate, A., Downer, C. C., Delgado, C. A., Sánchez-Londoño, J. D. (2010). Un registro de Tapir de montaña (*Tapirus pinchaque*) en el norte de la Cordillera Occidental de

- Colombia. [A record of mountain tapir (*Tapirus pinchaque*) in the northern Cordillera Occidental of Colombia]. *Mast. Neotrop.* 17: 111-116.
- Arispe, R., Rumiz, D. I., Venegas C. (2007). *Censo de jaguares (Panthera onca) y otros mamíferos con trampas cámara en la concesión temporal forestal El Encanto*. [Census of jaguars (*Panthera onca*) and other mammals with camera traps in the El Bosque forest temporary concessio]. Technical Inform 173. Wildlife Conservation Society, Santa Cruz de la Sierra, Bolivia. Pp. 1-37.
- Ascunce, M.S., Hasson, E., Mudry, M.D. (2003). COII: a useful tool for inferring phylogenetic relationships among New World monkeys (Primates, Platyrrhini). *Zool. Scripta* 32: 397-406.
- Ayala, G. (2003). *Monitoreo de Tapirus terrestris en el Izozog (Cerrado Cortado), mediante el uso de telemetría como base para un plan de conservación*. [Monitoring of *Tapirus terrestris* in the Izozog (Cerrado Cortado), using telemetry as the basis for a conservation plan]. Master Thesis. Universidad Mayor de San Andrés, La Paz, Bolivia. Pp. 1-96.
- Bandelt, H-J., Forster, P., Rohl, A. (1999). Median-joining networks for inferring intraspecific phylogenies. *Molecular Biology and Evolution* 16: 37-48.
- Barrientos, J., Cuéllar, E. (2003). Radiotelemetria de antas (*Tapirus terrestris*) en el Chaco seco, Izozog, Santa Cruz, Bolivia. In *Manejo de Fauna Silvestre en Amazonía y Latinoamérica*. [Radiotelemetry of antas (*Tapirus terrestris*) in the dry Chaco, Izozog, Santa Cruz, Bolivia. In *Wildlife Management in Amazon and Latin America*]. Polanco-Ochoa, R (Edit.). CITES, Fundación Natura, Bogotá DC., Colombia. Pp. 1-446.
- Bertorelle, G., Barbujani, G. (1995). Analysis of DNA Diversity by Spatial Autocorrelation. *Genetics* 140: 811-819.
- Bodmer, R. E., Brooks, D. M. (1997). Status and action plan of the lowland tapir (*Tapirus terrestris*). In Brooks, D. M., Bodmer, R. E., Matola, S., (Eds). *Tapir-status survey and conservation action plan*. IUCN/SSC Tapir Specialist Group, IUCN Gland & Cambridge. Pp. 46-56.
- Bodmer, R. E., Aquino, R., Puertas, P. (1997a). Alternativas de Manejo para la Reserva Nacional Pacaya-Samiria: Un análisis sobre el uso sostenible de la caza. In *Manejo de la Fauna Silvestre en la Amazonia*. [Management Alternatives for the Pacaya-Samiria National Reserve: An analysis on the sustainable use of hunting. In *Management of Wildlife in the Amazon*]. Fang, T. G., Bodmer, R. E., Aquino, R. & Valqui, M. H. (Eds.). Instituto de Ecología, La Paz, Bolivia. Pp.65-74.
- Bodmer, R. E., Puertas, P. E., Reyes, C., García, J. E., Diaz, D. (1997b). Animales de caza y palmeras: Integrando la socio-economía de extracción de frutos de palmera y carne de monte con el uso sostenible. In *Manejo de la Fauna Silvestre en la Amazonia*. [Game animals and palm trees: Integrating the socio-economy of extraction of palm fruits and bush meat with sustainable use. In *Management of Wildlife in the Amazon*]. Fang, T. G., Bodmer, R. E., Aquino, R. & Valqui, M. H. (Eds.). Instituto de Ecología, La Paz, Bolivia. Pp.75-86.
- Bodmer, R. E., Aquino, R., Navarro, J. (2000). Sustentabilidad de la caza de mamíferos en la cuenca del río Samiria, Amazonía Peruana. [Sustainability of mammal hunting in the Samiria river basin, Peruvian Amazon]. In Cabrera, E., Mercolli, C. & Resquin, R. (Eds.). *Manejo de Fauna Silvestre en Amazonía y Latinoamérica*. [Wildlife Management in the Amazon and Latin America]. CITES, Paraguay, Fundación Moises Bertoni & University of Florida, Asunción, Paraguay. Pp. 447-449.

- Bradley, R.D., Baker, R.J. (2001). A test of the genetic species concept: cytochrome-b sequences and mammals. *J. Mammal.* 82: 960–973.
- Brehm, G. L., Pitkin, M., Hilt, N., Fiedler, K. (2005). Montane Andean rain forests are a global diversity hotspot of geometrid moths. *J. Biogeog.* 32: 1621–1627.
- Brown Jr., K. S., Sheppard, P. M., Turner, J. R. G. (1974). Quaternary refugia in tropical America: evidence from race formation in *Heliconius* butterflies. *Proc. R. Soc. Lond. B.* 187: 369–378.
- Cabrera A. (1961). Catálogo de los mamíferos de América del Sur. Revista del Museo Argentino de Ciencias Naturales. [Catalog of mammals of South America. Journal of the Argentine Museum of Natural Sciences]. *Ciencias Zoológicas* 4: 309 – 732.
- Carroll, R. L. (1988). Ungulates, Edentates and Whales. In Carroll, R. L. (Ed.). *Vertebrate Paleontology and Evolution*. Freeman, New York. Pp. 502–568.
- Cavelier, J., Etter, A. (1995). Deforestation of montane forest in Colombia as result of illegal plantations of opium (*Papaver somniferum*). In Churchill, S. P., Balsley, H., Forero, E., Luteyn, J. L. (Eds.). *Biodiversity and Conservation of Neotropical montane forest*. New York Botanical Garden, New York. Pp. 125–137.
- Cavelier, J., Lizcano, D., Yerena, E., Downer, 967 C. (2010). The mountain tapir (*Tapirus pinchaque*) and Andean bear (*Tremarctos ornatus*): Two charismatic, large mammals in South American tropical montane cloud forests. In Bouijnacel L. A., Scabena, F. N., Hamilton, L. S (Eds.). *Tropical Montane Cloud Forests: Science for Conservation and Management*. Cambridge University Press, Cambridge, USA. Pp. 172–181.
- Chalukian, S. C., de Bustos, M. S., Lizárraga, L., Varela, D., Paviolo, A., Juliá, J. P., Quse, V. (2013). Plan de Acción para la conservación del Tapir (*Tapirus terrestris*) en Argentina: Estado de Conservación y Perspectivas. *Libro de Resúmenes del I Congreso Latinoamericano de Tapires y II Congreso Ecuatoriano de Mastozoología* [Action Plan for the conservation of the Tapir (*Tapirus terrestris*) in Argentina: State of Conservation and Perspectives. *Book of Abstracts of the 1st Latin American Congress of Tapirs and II Ecuadorian Congress of Mastozoology*], Puyo, Ecuador, May 8–11, 2013. Pp. 95–96.
- Chamorro-Rengifo, J., Cubillos-Rodriguez, P. A. (2007). Fichas de Fauna y Flora proyecto “Biodiversidad y desarrollo en ecoregiones estratégicas de Colombia-Orinoquía”. Sistema de Información sobre Biodiversidad de Colombia (SIB), [Fauna and Flora fiches project “Biodiversity and development in strategic ecoregions of Colombia-Orinoquía”. Information System on Biodiversity of Colombia (SIB)]. Bogotá, Colombia. <http://www.siac.net.co/sib/catalogoespecies>.
- Clapperton, C. (1993). *Quaternary geology and geomorphology of South America*. Elsevier, Amsterdam, The Netherlands. Pp. 1–489.
- Colbert, M. (2005). The facial skeleton of the Early Oligocene *Colodon* (Perissodactyla, Tapiroidea). *Palaeontologia Electronica* 8: 1–27.
- Collins AC, Dubach JM. (2000). Biogeographic and Ecological forces responsible for speciation 916 in *Ateles*. *Int. J. Primat.* 21: 421–444.
- Constantino E., Lizcano D., Montenegro O., Solano C. (2006). Danta Común. In: *Libro Rojo de los Mamíferos de Colombia*. Conservación Internacional Colombia, Ministerio de Ambiente, Vivienda y Desarrollo Territorial. [Red Book of the Mammals of Colombia. Conservation International Colombia, Ministry of Environment, Housing and Territorial Development] Bogotá, Colombia. Pp. 1–433.

- Cossíos, E.D., Lucherini, M., Ruiz-García, M., Angers B. (2009). Influence of ancient glacial periods on the Andean fauna: the case of the Pampas cat (*Leopardus colocolo*). *BMC Evolutionary Biology* 9: 68-79.
- Cossíos, E.D., Walker S, Lucherini M, Ruiz-García M, Angers B. (2012). Between high-altitude islands and high-altitude corridors. The population structure of the Andean cat (*Leopardus jacobita*). *Endangered Species Research* 16: 283-294.
- De Wilde, A.H. (1994). Caramanta, un proyecto para la creación de un nuevo parque nacional natural. [Caramanta, a project for the creation of a new natural national park]. Pp.33, In Cavelier, J., Uribe, A (eds.). *Resúmenes del Simposio Nacional—Diversidad biológica, conservación y manejo de los ecosistemas de montaña en Colombia*. [Summaries of the National Symposium-Biological diversity, conservation and management of mountain ecosystems in Colombia]. Universidad de los Andes, Bogota, Colombia.
- Deng, T. (2006). Paleoecological comparison between late Miocene localities of China and Greece based on *Hipparion* faunas. *Geodiversitas* 28: 499-516.
- Dobzhansky, T. (1937). *Genetics and the Origin of Species*. Columbia University Press, New York, USA. Pp. 1-364.
- Dobzhansky, T. (1971). Evolutionary oscillations in *Drosophila pseudoobscura*. In: Creed, R. (Eds.). *Ecological genetics and evolution*. Blackwell Scientific, Oxford, UK. Pp. 109–133.
- Downer, C.C. (1996). The mountain tapir, endangered 'flagship' species of the high Andes. *Oryx* 30: 45-58.
- Downer, C.C. (1997). *Status and action plan of the mountain Tapir (Tapirus pinchaque)*. *Tapirs, status survey and conservation action plan* (eds D.M. Brooks, R.E. Bodmer and S. Matola), pp. 10-22. IUCN/SSC Tapir specialist group, IUCN, Gland, Switzerland and Cambridge, UK.
- Downer, C. C. (2003). Ámbito hogareño y utilización de hábitat del Tapir Andino e ingreso de ganado en el Parque Nacional Sangay, [Home environment and habitat use of the Andean Tapir and livestock income in Sangay National Park]. Ecuador. *Lyonia* 4: 31-34.
- East, R. (1981). Species–area curves and populations of large mammals in African savanna reserves. *Biological Conservation* 21: 111–126.
- Eldredge, N., Gould, S. J. (1972). Punctuated equilibria: An alternative to phyletic gradualism. Schopf, T. J. M. (Ed.) in *Models in Paleobiology*. Ed. Freeman, San Francisco. Pp. 82-115.
- Excoffier, L., Smouse, P.E., Quattro, J.M. (1992). Analysis of molecular variance inferred from metric distances among DNA haplotypes: application to human mitochondrial DNA restriction data. *Genetics* 131: 479-491.
- Excoffier, L., Laval, G. & Schneider, S. (2005). Arlequin (version 3.0): An integrated software Packaged for population genetics data analysis. *Evolutionary Bioinformatics* 1: 47 –50.
- Ferrero, B. S., Noriega, J. I. (2007). A new upper Pleistocene tapir from Argentina: Remarks on the phylogenetics and diversification of neotropical Tapiridae. *Journal of Vertebrate Paleontology* 27: 504-511.
- Fjeldsa, J. (1994). Geographical patterns for relict and young species of birds in Africa and South America and implications for conservation priorities. *Biodiversity Conserv.* 3: 207-226.
- Forasiepi, A., Martinelli, A., Blanco J. (2007). Bestiario Fósil. Mamíferos del Pleistoceno de la Argentina. [Fossil bestiary. Pleistocene mammals of Argentina]. *Editorial Albatros SACI*, Buenos Aires, Argentina. Pp. 1-190.

- Froehlich, D. J. (1999). Phylogenetic systematics of basal perissodactyls. *Journal of Vertebrate Paleontology* 19: 140-159.
- Fu, Y-X. (1997). Statistical tests of neutrality against population growth, hitchhiking and background selection. *Genetics* 147: 915-925.
- Fu, Y., Li, W. (1993). Statistical Tests of Neutrality of Mutations. *Genetics* 133: 693-709.
- Gabriel, K.R., Sokal, R.R. (1969). A new statistical approach to geographic variation analysis. *Systematic Zoology* 18: 259-278.
- Goldman, N., Anderson, P., Rodrigo, A.G. (2000). Likelihood-based tests of topologies in phylogenetics. *Systematic Biology* 49: 652-670.
- Gould, S. J., and Eldredge, N. (1993). Punctuated equilibrium comes of age. *Nature* 366: 223-227.
- Groves C. P., Fernando, P., Robovsky, J. (2010). The sixth Rhino: A taxonomic reassessment of the critically endangered northern White rhinoceros. *PLoS ONE* 5: e9703.
- Haffer, J. (1969). Speciation in Amazonian forest birds. *Science* 165: 131-137.
- Haffer, J. (1970). Geologic-climatic history and zoogeographic significance of the Uraba region in northwestern Colombia. *Caldasia* 10: 603-636.
- Haffer, J. (1974). Avian speciation in tropical South America. *Publ. Nuttall. Ornith. Club.* 14: 1-390.
- Haffer, J. (1982). General aspects of the refuge theory. In Prance, G. T., (Ed.). *Biological diversification in the tropics*. New York: Columbia University. Pp. 6-24.
- Haffer, J. (1987). Quaternary history of tropical America. In Whitmore, T.C., and Prance, G.T. (Eds.). *Biogeography and Quaternary history in tropical America*. Oxford: Clarendon and Oxford University Press. Pp. 1-18.
- Haffer, J. (1997). Alternative models of vertebrate speciation in Amazonia: an overview. *Biodiversity and Evolution* 6: 451-476.
- Haffer, J. (2008). Hypotheses to explain the origin of species in Amazonia. *Braz. J. Biol.* 68: 917-947.
- Hall, T. (2004). *Bioedit Sequence Alignment Editor*. Version 7.0.0.
- Harpending, H.C. (1994). Signature and ancient population growth in a low-resolution mitochondrial DNA mismatch distribution. *Human Biol.* 66: 591-600.
- Harpending, H.C., Sherry, S.T., Rogers, A.R., Stoneking, M. (1993). Genetic structure of ancient human populations. *Current Anthropol.* 34: 483-496.
- Hebert, P.D.N., Ratnasingham, S., de Waard, J.R. (2003). Barcoding animal life: cytochrome c oxidase subunit 1 divergences among closely related species. *Proc. R. Soc. Lond. B* 270 (Suppl.): S96-9.
- Hebert, P.D.N., Stoeckle, M.Y., Zemlak, T., Francis, C.M. (2004). Identification of birds through DNA barcodes. *PLoS Biol.* 2: 1657-63.
- Henderson, A., Churchill, P., Luteyn, J. (1991). Neotropical plant diversity. *Nature* 229: 44-45.
- Hershkovitz, P. (1954). Mammals of Northern Colombia. Preliminary Report no. 7. Tapirs (genus *Tapirus*), with a Systematic Review of American Species. *Proceedings of the United States National Museum* 103: 465-496.
- Holanda, E. C., Rincón, A. D. (2012). Tapirs from the Pleistocene of Venezuela. *Acta Palaeontologica Polonica* 57: 463-472.

- Holanda, E. C., Ferrero, B. C. (2013). Reappraisal of the genus *Tapirus* (Perissodactyla, Tapiridae): Systematics and phylogenetic affinities of the South American tapirs. *Journal of Mammalian Evolution* 20: 33-44.
- Holbrook, L. T. (1999). The phylogeny and classification of tapiromorph perissodactyls (Mammalia). *Cladistics* 15: 331-350.
- Hudson, R.R. (2000). A new statistic for detecting genetic differentiation. *Genetics* 155: 2011-2014.
- Hudson, R.R., Boss, D.D., Kaplan, N.L. (1992a). A statistical test for detecting population subdivision. *Molecular Biology and Evolution* 9: 138-151.
- Hudson, R.R., Slatkin, M., Maddison, W.P. (1992b). Estimations of levels of gene flow from DNA sequence data. *Genetics* 132: 583-589.
- Huelsenbeck, J.P., Bull, J.J. (1996). A likelihood ratio test to detect conflicting phylogenetic signal. *Systematic Biology* 45: 92-98.
- Hulbert, R. C. (1995). The giant tapir, *Tapirus haysii*, from Leisey Shell Pit 1A and other Florida Irvingtonian localities. *Bull. Fla. Mus. Nat. Hist.* 37: 515-551.
- Hulbert, R. C. (2010). A new early Pleistocene tapir (Mammalia: Perissodactyla) from Florida, with a review of Blancan tapirs from the state. *Bulletin of the Florida Museum of Natural History* 49: 67-126.
- Hulbert, R. C., Wallace, S. C. (2005). Phylogenetic analysis of late Cenozoic *Tapirus* (Mammalia, Perissodactyla). *Journal of Vertebrate Paleontology* 25 (supplement to 3): 72A.
- Hulbert, R. C., Wallace, S. C., Klippel, W. E., Parmalee, P. W. (2009). Cranial morphology and systematics of an extraordinary sample of the late Neogene Dwarf Tapir, *Tapirus polkensis* (Olsen). *Journal of Paleontology* 83: 238-262.
- Isaaks EH, Srivastava RM. (1989). *An introduction to applied geostatistics*. Oxford University Press, New York. Pp 1-561.
- Kartavtsev, Y. (2011). Divergence at Cyt-b and Co-1 mtDNA genes on different taxonomic levels and genetics of speciation in animals. *Mitochondrial DNA* 22: 55-65.
- Klammer, G. (1984). The relief of the extra-Andean Amazon basin. In Sioli, H. (Ed.). *The Amazon. Limnology and landscape ecology of a mighty tropical river and its basin*. Dordrecht: Junk Publishers. Pp. 47-83.
- Kelly J. (1997). A test of Neutrality Based on Interlocus Associations. *Genetics* 146: 1197-1206.
- Kessler, M. (2002). The elevational gradient of Andean plant endemism: varying influences of taxon-specific traits and topography at different taxonomic levels. *J. Biogeog.* 29: 1159-1165.
- Kimura M. (1980). A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences. *J. Mol. Evol.* 16: 111-120.
- Krömer, T., Kessler, M., Gradstein, S. R., Acebey, A. (2005). Diversity patterns of vascular epiphytes along an elevational gradient in the Andes. *J. Biogeogr* 29: 1159-1165.
- León-Canales, E. (2007). *Orígenes humanos en los Andes peruanos*. [Human origins in the Peruvian Andes]. Universidad de San Martín de Porres. Lima, Perú. Pp. 1-328.
- Librado, P., Rozas, J. (2009). DnaSP v5: A software for comprehensive analysis of DNA polymorphism data. *Bioinformatics* 25: 1451-1452 | doi: 10.1093/bioinformatics/btp187.
- Lizcano, D.J., Cavelier, J. (2000). Densidad poblacional y disponibilidad de habitat de la danta de montaña (*Tapirus pinchaque*) en los andes centrales de Colombia. [Population density

- and habitat availability of the mountain tapir (*Tapirus pinchaque*) in the central Andes of Colombia]. *Biotropica* 31: 165-173.
- Lizcano, D. Pizarro, J. V., Cavelier, J., Carmona, J. 2002. Geographic distribution and population size of the mountain tapir (*Tapirus pinchaque*) in Colombia. *J. Biogeog.* 28: 1-9.
- MacFadden, B. J. (1992). *Fossil horses: Systematics, paleobiology, and evolution of the family Equidae*. Cambridge University Press, New York.
- MacNeish, R. S. (1979). The early man remains from Pikimachay cave, Ayacucho basin, highlands Peru. In Humprey, R. L. & Standford, D., (Eds.). *Pre-Llano cultures of the Americas: Paradoxes and possibilities*. Anthropological Society of Washington, Pp. 1-47.
- Mantel, N.A. (1967). The detection of disease clustering and a generalized regression approach. *Cancer Research* 27: 209-220.
- Matula, D.W., Sokal, R.R. (1980). Properties of Gabriel graphs relevant to geographic variation research and the clustering of points in the plane. *Geographic Analysis* 12: 205-222.
- Marroig, G., Cerqueira, R. (1997). Plio-Pleistocene South American history and the Amazon Lagoon hypothesis: a piece in the puzzle of Amazonian diversification. *J. Comp. Biol.* 2: 103-119.
- McKenna, M. C., Bell, S. K. (1997). *Classification of Mammals-Above the species level*. Columbia University Press, New York. Pp. 631.
- Meffe GK, Carroll CR. (1997). *Principles of Conservation Biology*. Second Edition. Sinauer Associates, INC. Publishers. Sunderland, Massachusetts.
- Merelender, A. M., Woodruff, D. S., Ryder, O. A., Kock, R., Vahala, J. (1989). Allozyme variation and differentiation in African and Indian rhinoceroses. *J. Hered.* 80: 377-382.
- Metais, G., Soe, A. N., Ducrocq, S. (2006). A new basal tapiromorph (Perissodactyla, Mammalia) from the middle Eocene of Myanmar. *Geobis* 39: 513-519.
- Metivier, S. P. (1998). *A reconstruction of glacial extent, temperature and precipitation in South America at the time of the Last Glacial Maximum*. PhD Thesis, Syracuse University, Syracuse, USA. Pp. 1-116.
- Montenegro, O. (2002). *Evaluación del estado actual de la danta o tapir de páramo (Tapirus pinchaque) en la región Andina Oriental, con base en una recopilación y verificación de registros de campo y una aproximación preliminar al estado de su hábitat en la región. [Evaluation of the current status of tapir tapir (Tapirus pinchaque) in the Eastern Andean region, based on a compilation and verification of field records and a preliminary approximation of the state of its habitat in the region]*. Informe final. CORPOCHIVOR, CAR, CORPOGUAVIO, CORPOBOYACA & Ministerio del Medio Ambiente. Garagoa. Colombia.
- Morales, J. C., Melnick, D. J. (1994). Molecular systematics of the living rhinoceros. *Mol. Phylogenet. Evol.* 3: 128-134.
- Moritz, C., Faith, D.P. (1998). Comparative phylogeography and the identification of genetically divergent areas for conservation. *Molecular Ecology* 7:419-429.
- Morral, N., Bertranpetit, J., Estivill, X., et al., (1994). The origin of the major cystic fibrosis mutation (delta F508) in European populations. *Nature Genetics* 7: 169-175.
- Nabholz, B., Glemin, S., Galtier, N. (2008). Strong variations of mitochondrial mutation rate 1224 across mammals—the longevity hypothesis. *Mol. Biol. Evol.* 25: 120-130.

- Nabholz, B., Glemin, S., Galtier, N. (2009). The erratic mitochondrial clock: variations of mutation rate, not population size, affect mtDNA diversity across birds and mammals. *BMC Evol. Biol.* 9: 54-67.
- Nei, M., Kumar, S. (2000). *Molecular Evolution and Phylogenetics*. Oxford University Press, New York. Pp.1-333.
- Nores, M. (1999). An alternative hypothesis for the origin of Amazonian bird diversity. *J. Biogeogr.* 26: 475-485.
- Nores, M. (2004). The implications of Tertiary and Quaternary sea level rise events for avian distribution patterns in the lowlands of northern South America. *Global Ecology and Biogeography* 13: 149-162.
- Noriega, J., Carlini, A. A., Brandoni, O., Ferrero, B. S., Vassalo, C., Cettour de Soto, S. (2004). Mamíferos del Cuaternario de la cuenca del río Uruguay, Departamento de Concordia, Entre Ríos, Argentina. [Mammals of the Quaternary of the basin of the river Uruguay, Department of Concordia, Entre Ríos, Argentina]. *Rev. An. de Comunic. APA, Resúmenes, Diamante, APA*. Pp. 21-22.
- Norman, J., Ashley, M. (2000). Phylogenetics of Peryssodactyla and Tests of the Molecular Clock. *Journal of Molecular Evolution* 50: 11-21.
- Oden, N. (1984). Assessing the significance of a spatial correlogram. *Geographical Analysis* 16: 1-16.
- Orlando, L., Ginolhac, A., Zhang, G., et al., (2013). Recalibrating *Equus* evolution using the genome sequence of an early Middle Pleistocene horse. *Nature* 499: 74-81.
- Padilla, M., Dowler, R. C. (1994). *Tapirus terrestris*. *Mammalian species* 481: 1-8.
- Perini, A.F., Oliveira, J.A., Salles, L.O., Moraes Neto, C.R., Guedes, P.G., Oliveira, L.F.B., Weksler, M. (2011). New fossil records of *Tapirus* (Mammalia, Perissodactyla) from Brazil, with a critical analysis of intra-generic diversity assessments based on lower molar size variability. *Geobios* 44: 609-619.
- Pimm SL. (1991). *The balance of nature? Ecological issues in the conservation of species and communities*. University of Chicago Press, Chicago.
- Posada D, Crandall KA. (1998). MODELTEST: testing the model of DNA substitution. *Bioinformatics* 14: 817-818.
- Prance, G. T. (1982). *Biological diversification in the tropics*. New York: Columbia University.
- Prance, G. T. (1996). Islands in Amazonia. *Phil. Trans. R. Soc. Lond. B* 351: 823-833.
- Prance, G.T., Lovejoy, T.E. (1985). *Amazonia. Key environments*. Oxford: Pergamon Press.
- Rabinowitz, A., Zeller, K. A. 2010. A range-wide model of landscape connectivity and conservation for the jaguar, *Panthera onca*. *Biological Conservation* 143: 939-945.
- Radinsky, L. B. (1969). The early evolution of the Perissodactyla. *Evolution* 23: 308-328.
- Rambaut, A., Grassly, N. C. (1997). Seq-gen: an application for the monte carlo simulation of DNA sequence evolution along phylogenetic trees. *Computer Applications in the Biosciences* 13: 235-238.
- Ramos-Onsins, S.E., Rozas, J. (2002). Statistical Properties of New Neutrality Tests Against Population Growth. *Mol. Biol. Evol.* 19: 2092-2100.
- Ripley, B.D. (1981). *Spatial statistics*. John Wiley and Sons, New York.
- Rodbell, D. T. (1991). *Late Quaternary glaciation and climatic change in the northern Peruvian Andes*. PhD Thesis. University of Colorado, Boulder, Colorado, USA. Pp. 1-416.
- Rodbell, D. T., Seltzer, G. O. (2000). Rapid ice margin fluctuations during the Younger Dryas in the tropical Andes. *Quaternary Research* 54: 328-338.

- Rohlf, F.J. (2000). *NTSYS-pc, Numerical Taxonomy and Multivariate Analysis System, Version 2.1*. Exeter Publications, New York, USA.
- Rogers, A.R., Harpending, H.C. (1992). Population growth makes waves in the distribution of pairwise genetic differences. *Mol. Biol. Evol.* 9: 552-569.
- Rogers, A.R., Fraley, A.E., Bamshad, M.J., Watkins, W.S., Jorde, L.B. (1996). Mitochondrial mismatch analysis is insensitive to the mutational process. *Mol. Biol. Evol.* 13: 895-902.
- Romero, V. P., Ferrer A., Lew D. (2013). La danta de tierras bajas como principal fuente proteica en una comunidad agrícola al Sur del río Orinoco, Venezuela: Implicaciones para su conservación. *Libro de Resúmenes del I Congreso Latinoamericano de Tapires y II Congreso Ecuatoriano de Mastozoología*, [The lowland tapir as the main source of protein in an agricultural community south of the Orinoco River, Venezuela: Implications for its conservation. *Book of Abstracts of the 1st Latin American Congress of Tapires and II Ecuadorian Congress of Mastozoology*], Puyo, Ecuador, May 8-11, 2013. Pp. 103.
- Rothlisberger, F. (1987). *10,000 jahre gletschergeschichte der erde*. [10,000 years of glacial history of the earth]. Verlag Sauerlander, Aarau, Switzerland. Pp. 1-225.
- Rozas, J., Gullaud, M., Blandin. G., Aguadé, M. (2001). DNA Variation at the rp49 Gene Region of *Drosophila simulans*: Evolutionary Inferences from an Unusual Haplotype Structure. *Genetics* 158: 1147-1155.
- Ruiz-García, M. (2003). Molecular population genetic analysis of the spectacled bear (*Tremarctos ornatus*) in the Northern Andean Area. *Hereditas* 138: 81-93.
- Ruiz-García, M. (2013). The genetic demography history and phylogeography of the Andean bear (*Tremarctos ornatus*) by means of microsatellites and mtDNA markers. In *Molecular Population Genetics, Evolutionary Biology and Conservation of Neotropical Carnivores*. Ruiz-García, M., Shostell, J. (Eds.). Nova Science Publishers. New York. Pp. 129-158.
- Ruiz-García, M., Orozco-terWengel, P., Payán, E., Castellanos, A. (2003). Genética de Poblaciones molecular aplicada al estudio de dos grandes carnívoros (*Tremarctos ornatus* – Oso andino, *Panthera onca*- jaguar): lecciones de conservación. [Genetics of Molecular Populations applied to the study of two large carnivores (*Tremarctos ornatus* - Andean Bear, *Panthera onca*-jaguar): conservation lessons]. *Bol. Real Soc. Esp. Hist. Nat.* 98: (1-4): 135-158.
- Ruiz-García, M., Orozco-terWengel, P., Castellanos, A., Arias, L. (2005). Microsatellite analysis of the spectacled bear (*Tremarctos ornatus*) across its range distribution. *Genes and Genetics Systems* 80: 57-69.
- Ruiz-García, M., Vásquez, C., Pinedo-Castro, M., Sandoval, S., Kaston, F., Thoisy, B., Shostell, J. M. (2012). Phylogeography of the mountain tapir (*Tapirus pinchaque*) and the Central American tapir (*Tapirus bairdii*) and the molecular origins of the three South-American tapires. In Anamthawat-Jónsson, K. (Ed.). *Current Topics in Phylogenetics and Phylogeography of Terrestrial and Aquatic Systems*. InTech, Rijeka, Croatia. Pp. 83-116.
- Ruiz-García, M., Cossíos, D., Lucherini, M., Yañez, F., Pinedo-Castro, M., Angers, B. (2013). Population genetics and spatial structure in two Andean cats (the Pampas cat, *Leopardus pajeros* and the Andean mountain cat, *L. jacobita*) by means of nuclear and mitochondrial markers and some notes on skull biometrics. Ruiz-García, M., and Shostell, J. M., (Eds.) in *Molecular Population Genetics, Evolutionary Biology and Biological Conservation of Neotropical Carnivores*. Nova Science Publishers, Inc.. New York (USA). Pp. 187-246.
- Ruiz-García, M., Vásquez, C., Sandoval, S., Kaston, F., Luengas-Villamil, K., Shostell, J. M. (2015a). Phylogeography and spatial structure of the lowland tapir (*Tapirus terrestris*,

- Perissodactyla: Tapiridae) in South America. *Mitochondrial DNA* 26: 1-9, on line DOI: 10.3109/19401736.2015.1022766.
- Ruiz-García, M., Castellanos, A., Bernal, L. A., Pinedo-Castro, M., Kaston, F., Shostell, J.M. (2015b). Mitogenomics of the elusive mountain tapir (*Tapirus pinchaque*, Tapiridae, Perissodactyla, Mammalia) in Colombia and Ecuador: Phylogeography and insights into the origin and systematics of the South American tapirs. *J. Hered.* (in press).
- Ruiz-García, M., Pinedo-Castro, M., Shostell, J. M. (2015c). Is “*Tapirus kabomani*” a new Amazonian tapir species? The fulminant rise and fall of a “typological species” by means of mitogenomics and craniometrics. *J. Mammal.* (submitted).
- Saillard, J., Forster, P., Lynnerup, N., Bandelt, H-J., Norby, S. (2000). mtDNA variation among Greenland Eskimos: the edge of the Beringian expansion. *American Journal of Hum. Genet.* 67: 718-726.
- Salazar-Bravo, J., Tarifa T., Aguirre, L. F., Yensen, E., Yates, T. L. (2003). Revised checklist of Bolivian Mammals. *Occasional Papers, Museum of Texas Tech University*, 220: 1-27.
- Sambrook, J., Fritsch, E., Maniatis, T. (1989). *Molecular Cloning: A Laboratory manual*. 2nd edition. V1. Cold Spring Harbor Laboratory Press. New York.
- Sarriá, S. (1993). *Parque Nacional Natural Farallones de Cali*. Monografía. Corporación Autónoma Regional del Valle del Cauca-CVC. Fundación Protectora de las Cuencas-PROCUENCAS. [Natural Farallones de Cali National Park. Monograph. Regional Autonomous Corporation of Valle del Cauca-CVC. Protective Foundation of the Basins-PROCUENCAS]. Cali, Colombia. Pp. 1-312.
- Savage, R. J. G., Long, M. R. (1986). *Evolución de los Mamíferos: Una guía ilustrada*. [Evolution of Mammals: An illustrated guide]. Ediciones Akal, Torrejón de Ardoz, Madrid. Pp. 1-258.
- Schaller, G. B. (1983). Mammals and their biomass on a Brazilian ranch. *Arquivos de Zoologia, Sao Paulo* 31: 1-36.
- Seltzer, G. O. (1987). *Glacial history and climatic change in the central Peruvian Andes*. Msc Thesis, University of Minnesota, Minneapolis, Minnesota, USA. Pp. 1-276.
- Seltzer, G. O. (1990). Recent glacial history and paleoclimate of the Peruvian-Bolivian Andes. *Quaternary Science Reviews* 9: 137-152.
- Shimodaira, H., Hasegawa, H. (1999). Multiple comparisons of log-likelihoods with applications to phylogenetic inference. *Molecular Biology and Evolution* 16: 1114-1116.
- Simonsen, K., Churchill, G., Aquadro, C. (1995). Properties of Statistical Tests of Neutrality for DNA Polymorphism Data. *Genetics* 141: 413-429.
- Smouse, P.E., Long, J.C., Sokal, R.R. (1986). Multiple regression and correlation extension of the Mantel test of matrix corresponde. *Systematic Zoology* 35: 627-632.
- Sokal, R.R., Oden, N.L. (1978). Spatial autocorrelation in Biology. 1. Methodology. *Biological Journal of Linnean Society* 10: 199-228.
- Sombroek, W. G. (1966). *Amazon soils*. Wageningen: Centrum voor Landbouwpublikaties.
- Sphuler, J. N. (1972). Genetic, linguistic, and geographical distances in native North America. In *The assessment of population affinities in man*. Wiener, J. S., Huizinga, J (Eds.). Oxford University Press, Oxford.
- Swofford, D. L., Olsen, G. L., Wadell, P. J., Hillis, D. M. (1996). Phylogenetic inference. In Hillis, D.M., (Ed.) *Molecular Systematics*. Sunderland, Massachusetts: Sinauer Associates. Pp. 407-514.

- Tajima, F. (1989). Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics* 123: 585-595.
- Tamura K, Peterson D, Peterson N, Stecher G, Nei M, Kumar, S. (2011). MEGA5: Molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Mol Biol Evol* 28: 2731-2739.
- Terborgh, J. (1992). *Diversity and the tropical rain forest*. New York: Freeman.
- Thoisy, B., Goncalves da Silva, A., Ruiz-García, M., Tapia, A., Ramirez, O., Arana, M., Quse, V., Paz-y-Miño, C., Tobler, M., Pedraza, C., Lavergne, A. 2010. Population history, phylogeography, and conservation genetics of the last Neotropical mega-herbivore, the Lowland tapir (*Tapirus terrestris*). *BMC Evol. Biol.* 10: 278-295.
- Thompson, L. G., Mosley, E., Davies, M. E., Lin, P. N., Henderson, K. A., Coledal, J., Bolzan, J. F., Liu, K. B. (1995). Huascarán, Perú. *Science* 269: 46-50.
- Tirira, D. (2011). *Libro Rojo de los Mamíferos del Ecuador*. Publicación especial de los Mamíferos del Ecuador No 8. Pontificia Universidad Católica del Ecuador and Ministerio de Medio Ambiente, Quito, Ecuador. [Red Book of the Mammals of Ecuador. Special Publication of the Mammals of Ecuador No. 8. Pontifical Catholic University of Ecuador and Ministry of the Environment, Quito, Ecuador].
- Tirira, D., Castellanos, A. (2001). Tapir de montaña (*Tapirus pinchaque*). In Tirira D (Ed.). *Libro Rojo de los Mamíferos del Ecuador*. SIMBIOE/EcoCiencia/Ministerio del Ambiente/UICN. Serie Libros Rojos del Ecuador. Tomo 1. Publicación Especial sobre los Mamíferos del Ecuador 4, Quito. Pp. 98-100. [Mountain tapir (*Tapirus pinchaque*). In Tirira D (Ed.). *Red Book of the Mammals of Ecuador*. SIMBIOE / EcoCiencia / Ministry of the Environment / IUCN. Red Books Series of Ecuador. Volume 1. Special Publication on the Mammals of Ecuador 4, Quito. Pp. 98-100].
- Tonni, E. P. (1992). *Tapirus* Brisson, 1762 (Mammalia, Perissodactyla) en el Lujanense (Pleistoceno Superior-Holoceno Inferior) de la Provincia de Entre Ríos, Republica Argentina. *Ameghiniana* 29: 3-8. [*Tapirus* Brisson, 1762 (Mammalia, Perissodactyla) in the Lujanense (Lower Pleistocene-Lower Holocene) of the Province of Entre Ríos, Argentina. *Ameghiniana* 29: 3-8].
- Tougard, C., Delefosse, T., Hänni, F., Montgelard, C. (2001). Phylogenetics relationships of the five extant rhinoceros species (Rhinocerotidae, Perissodactyla) based on mitochondrial Cytochrome b and 12S rRNA genes. *Molecular Phylogenetics and Evolution* 19: 34-44.
- Upton G, Fingleton B. (1985). *Spatial data analysis by example*. Vol 1: Point pattern and quantitative data. John Wiley and Sons, Chichester.
- Van der Hammen, T. (1992). *Historia, ecología y vegetación*. [History, ecology and vegetation]. Editorial Corporación Colombiana para la Amazonía, Araracuara, Bogotá DC., Colombia. Pp. 1-411.
- Van der Hammen, T. (2001). Paleoeecology of Amazonia. In: Guimaraes Vieira, I. C., Silva, J. M. C., Oren, D. C. & D'Incao, M. A. D. (Eds). *Diversidade biológica e cultural da Amazonia*. [Biological and cultural diversity of Amazonia]. Museu Paraense Emilio Goeldi, Belem, Brazil. Pp. 19-44.
- Van der Hammen, T., Cleff, A. M. (1992). Holocene changes of rainfall and river discharge in northern South America and the El Niño phenomenon. *Erdkunde* 46: 252-256.
- Vanzolini, P. E. (1970). *Zoología sistemática, geografía e a origem das espécies*. [Systematic zoology, geography and the origin of species]. São Paulo: Instituto Geográfico de São Paulo. Pp. 1-56.

- Vanzolini, P. E. (1973). Paleoclimates, relief, and species multiplication in equatorial forests. In Meggers, B. J., Ayensu, E.S., Duckworth, W.D. (Eds.). *Tropical forest ecosystems in Africa and South America: A comparative review*. Washington: Smithsonian Institution. Pp. 255-258.
- Vanzolini, P. E. (1992). Paleoclimas e especiacao em animais da América do Sul tropical. [Paleoclimates and speciation in tropical South American animals]. São Paulo. *Estud. Avancados* 6: 41-65.
- Vanzolini, P.E., Williams, E. E. (1970). South American anoles: Geographic differentiation and evolution of the *Anolis chrysolepis* species group (Sauria, Iguanidae). *Arq. Zoolog* 19: 1-298.
- Verweij, P. A. (1995). *Spatial and temporal modelling of vegetation patterns*. International Institute for Aerospace Survey and Earth Sciences ITC. Enschede, The Netherlands.
- Voss, R.S., Helgen, K. M., Jansa, S. A. (2014). Extraordinary claims require extraordinary evidence: A comment on Cozzuol et al. (2013). *J Mammal* 95:893–8.
- Wallace, R. B., Gómez, H., Porcel, Z. R., Rumiz, D. I. (2010). *Distribución, Ecología y Conservación de los Mamíferos medianos y grandes de Bolivia*. [Distribution, Ecology and Conservation of the medium and large mammals of Bolivia]. Editorial Centro de Ecología Difusión Fundación Simón I. Patiño, Santa Cruz de la Sierra, Bolivia. Pp. 1-906.
- Walsh, P.S., Metzger, D. A., Higuchi, R. (1991). Chelex 100 as a medium for simple extraction of DNA for PCR-based typing from forensic material. *BioTechniques* 10: 506-513.
- White, M.J.D. (1968). Models of speciation. New concepts suggest that the classical sympatric and allopatric models are not the only alternatives. *Science* 159: 1065-1070.
- White, M.J.D. (1978). *Modes of speciation*. W. H. Freeman, San Francisco.
- Whithmore, R. T., Prance, G. T. (1987). *Biogeography and quaternary history in tropical America*. Oxford: Clarendon Press. (Monography on Biogeography, 3).
- Wright, S. (1951). The genetical structure of populations. *Annals of Eugenics* 15: 323–354.
- Xu, X. (1996). *Studies of mammalian mitochondrial genomes with special emphasis on the Perissodactyla*. Ph.D. dissertation, University of Lund, Sweden.
- Yamini, B., Schillhorn, T. W. (1988). Schistosomiasis and nutritional myopathy in a Brazilian tapir (*Tapirus terrestris*). *Journal of Wildlife Diseases* 24: 703-707.
- Zeller, K. A., Rabinowitz, A., Salom-Pérez, R., Quigley, H. (2013). The jaguar corridor initiative: A range-wide conservation strategy. In Ruiz-García, M., Shostell, J. M. (Eds.). *Molecular Population Genetics, Evolutionary Biology and Biological Conservation of Neotropical Carnivores*. Nova Science Publishers, Inc., New York, USA. Pp. 629-657.

Chapter 44

STRATEGIES FOR GENE PROSPECTING OF PLANTS IN RESPONSE TO DROUGHT AND SALINITY

***Deyvid Novaes Marques^{1,2}, Nicolle Louise Ferreira Barros^{1,3},
Diehgo Tuloza da Silva^{1,4}, Fabiano Melo de Brito^{1,5}
and Claudia Regina Batista de Souza^{1,*}***

¹Instituto de Ciências Biológicas, Universidade Federal do Pará,
Belém, PA, Brazil

²Programa de Pós-Graduação em Genética e Biologia Molecular,
Universidade Federal do Pará, Belém, PA, Brazil

³Bolsista de Iniciação Científica-PIBIC-UFPA/CNPq

⁴Programa de Pós-Graduação em Neurociências e Biologia Celular,
Universidade Federal do Pará, Belém, PA, Brazil

⁵Programa de Pós-Graduação em Agronomia, Universidade Federal Rural da Amazônia,
Belém, PA, Brazil

ABSTRACT

Plants are responsible for a significant part of food supply for the entire world, and through agriculture they play an extremely important socio-economic role for the mankind. Therefore, the development of genetically improved crops becomes even more relevant for it aims at an everlasting enhancement of agronomic traits of interest. For many years, plant genetic improvement program has been based in empiric selection of the target traits; however, significant advances were obtained in the last years. Many tools, allowing crops to be improved with greater optimization of the time needed to reach the necessary modifications, are currently available. Regarding the methods used in the genetic improvement, molecular studies have been essential to identify which genes are important for each specific agronomic trait, such as those related to tolerance to abiotic stress. Such studies contribute not only to a better understanding of the endogenous defense mechanisms of plants at molecular level by which these organisms adapt when facing hostile conditions, but also contribute to the generation of stress-tolerant crops by genetic

* Corresponding Author's Email: bsouza@ufpa.br.

engineering. These programs aim a significant productivity and sustainability that can be reached through soil preservation that is directly related to less necessity of farm inputs. Better adapted crop cultivars make it possible, as well as better use and decontamination of water resources. In this chapter we attempt at providing an overview regarding strategies that have been used for prospection of genes related to the response of plants to abiotic stress. The combination of biotechnological and bioinformatics tools used in the identification of stress-related genes and development of genetically engineered crops by silencing and/or over-expression of specific genes will be presented in this chapter. Emphasis will be given to the drought and salinity that represent a major part of abiotic stresses by which the plants are often exposed, leading to serious production losses in many important crops worldwide.

1. INTRODUCTION

The current human population of the entire world is estimated in seven billion people. Estimates also show an increasing to 7.5 billion of people by the year 2020 and to about 9 billion by 2050 (Lutz et al., 2001; Godfray et al., 2010). Therefore, to attend the nutritional requirements of human population in growing is necessary to enhance food productivity, including vegetable sources, which are also essential for animal feed. For example, estimates indicate that the world food grain production needs to be doubled by the year 2050 to attend the ever growing demands of the population (Tilman et al., 2002). Despite this need, some factors, such as climate changes and shrinking environmental resources, have significantly limited the agricultural production worldwide (Wang W et al., 2003). In addition, models of climate prediction indicate an increasing of 3–5°C of surface temperatures in the next 50–100 years, that will drastically affect the global agriculture (Solomon et al., 2007).

Adverse environmental conditions, which can potentially cause physiological limitations for plant growth and development, are known as stresses. Thus, plants are frequently exposed to a large range and number of environmental stresses, which when occurring simultaneously can cause severe consequences for these organisms.

Environmental stress, such as drought, salinity, cold and high temperature, are known as abiotic stresses. Among them, drought and salinity are common stress conditions that adversely affect crop production worldwide. For example, drought limits plant growth and development since can interfere in photosynthesis rates and soil nutrient availability. Likewise, salinity interferes with plant development, leading to physiological drought and ion toxicity (Krasensky and Jonak, 2012; Chaves et al., 2009). Otherwise, an improvement of abiotic stress tolerance might significantly increase the productivity of the most crops. Therefore, understanding the mechanisms of plant response to abiotic stress has been one of the most important priorities of studies in plant physiology, contributing for the development of stress-tolerant crops by genetic improvement programs that can be greatly advantageous in areas where the agriculture is limited by such stress.

For many years, the genetic improvement of crops has been performed by classical/conventional breeding programs that use the interbreeding of closely related individuals to produce enhanced crops with desirable agronomic traits. Classical breeding identifies traits of interest in parent lines and incorporates them into a new variety through crosses (or hybridization). This methodology is based on the genetic modification of wild species to create new altered cultivars (Doebley et al., 2006; Tang et al., 2010). Despite

conventional breeding has significantly contributed for development of new crop varieties, the pace to obtain new cultivars is relatively slow due to limitation of fertility barriers that allows only plants of the same, or closely related species for hybridization (Doebley et al., 2006). In addition, classical breeding requires a long period and several generations to select and evaluate useful genotypes, therefore, in some cases, this process can be limited to address global food security and attend the increasing requirements of food demands (Tester and Langridge, 2010). On the other hand, the advent of genetic engineering tools allowed the researchers to overcome some limitations found in classical breeding programs, where genes can only be exchanged between closely related species.

In 1970s, scientists were able to manipulate the genetic material at molecular level by DNA recombinant technology, also known as genetic engineering. This technology was the first step for isolation and manipulation of specific DNA sequences from any organism, allowing the transfer of genes within and across species boundaries (Cohen, 2013). With development of genetic engineering tools, significant advances in molecular biology were achieved in different research areas. Regarding investigations in plant molecular biology, one of the most important findings has been the understanding of molecular mechanisms by which plants recognize the stress signals from the environment to produce adaptive responses by expression of stress-inducible genes.

It is known that the complex process by which plant response to abiotic stress involves many genes and proteins playing specific functions in different biochemical and molecular pathways (Reis et al., 2012). Many drought-inducible genes are also induced by salt stress, suggesting the existence of similar mechanisms of stress responses. Stress-inducible genes coding for proteins can be classified into two main groups: (1) The first group comprises a large number of proteins with enzymatic or structural functions, such as enzymes involved in cellular detoxification process and synthesis of osmolytes, such as trehalose and proline, and other proteins with macromolecular protection, such as late embryogenesis abundant proteins, chaperones and mRNA binding proteins. (2) The second group involves a variety of regulatory proteins, such as transcription factors, protein kinases and receptor protein kinases, participating in diverse mechanisms of gene expression regulation (Agarwal et al., 2001).

Therefore, the identification of stress-induced genes contributes to understanding mechanism by which plants response and adapt themselves to abiotic stress, as well as molecular breeding of agriculturally important plants aiming the acquisition or increasing of stress tolerance. In this context, significant advances have been achieved by using the gene prospecting methodology, which covers the isolation and characterization of genes, including relevant *in vitro* and *in vivo* assays using molecular biology, bioinformatics and biotechnology tools. With the utilization of gene prospecting methodologies in many plant biology studies, several abiotic stress-inducible genes were isolated and their functions were precisely characterized in transgenic plants (Hirayama and Shinozaki, 2010).

In this chapter, we aimed to present some strategies that have used in prospecting of genes involved in response of plant to drought and salinity stresses. First, some genes coding for proteins playing roles in response to these stresses will be presented, taking into account their importance and potential application in molecular breeding programs. Later, some approaches for gene identification and genetic transformation of plants will be presented and discussed.

2. GENES CODING FOR PROTEINS INVOLVED IN RESPONSE TO DROUGHT AND SALINITY IN PLANTS

2.1. Basic Leucine Zipper Proteins and Absciscic Acid

Many physiological adaptations of plant to abiotic stress are under the control of the phytohormone abscisic acid (ABA) and involve specific activation of target genes that participate in the regulation of response to drought, salinity and low temperature (Cutler et al., 2010). It is known that cellular dehydration under water limited conditions induces an increase in endogenous ABA levels that trigger downstream target genes encoding signaling factors, metabolic enzymes, transcription factors, and others (Shinozaki and Shinozaki, 2006).

Transcription factors (TFs) are sequence-specific DNA-binding proteins able to activating and/or repressing transcription. TFs regulate transcription levels of target genes by mechanisms that often involve a whole cascade of signalling events determined by tissue type, developmental stage or environmental conditions (Wyrick and Young, 2002). Among environmental conditions, they represent a primary defense response to stimulus from abiotic stress (Riechmann et al., 2000). TFs can be classified in families according to their DNA-binding domain (Riechmann et al., 2000).

A single TF can control the expression of many target genes by specific binding to the *cis*-acting element within the promoters of target genes. This type of transcriptional regulatory system is termed *regulon* (Nakashima et al., 2009; Zahur et al., 2013). At least four different regulons can be identified, two ABA independent: (1) the CBF/DREB regulon, (2) the NAC (NAM, ATAF and CUC) and ZF-HD (zinc-finger homeodomain) regulon (Nakashima et al., 2009; Saibo et al., 2009); while two regulons are ABA dependent (3) the AREB/ABF (ABA-responsive element-binding protein/ ABA-binding factor), and (4) the MYC (myelocytomatosis oncogene)/MYB (myeloblastosis oncogene) (Saibo et al., 2009).

Many ABA-inducible genes contain a conserved *cis*-acting element named ABRE (ABA-responsive element; PyACGTGGC) in their promoter regions. Among TFs with ability to bind to ABRE, the basic leucine zipper (bZIP) proteins contain three functional regions involved in processes such as dimerization, DNA binding and transcriptional regulation (Bader and Vogt, 2006).

Studies have shown bZIP proteins involved in response to ABA, as well as response and tolerance to abiotic stress in many plant species. For example, Wang et al. (2011) reported increased transcript levels of bZIP in *Arabidopsis thaliana* plants under abiotic stresses, including drought and salinity. Other studies revealed bZIP proteins involved in the ABA signaling process in *Arabidopsis* that may, as a consequence, be classified as ABF or AREB (Choi et al., 2000; Uno et al., 2000). Also in *Arabidopsis*, an increased expression of bZIP genes promoted sensibility to ABA and changes in the expression of other genes related to stress, as well as improvement of the drought tolerance by transpiration reduction (Singh et al., 2002). In rice (*Oryza sativa*), the *OsbZIP23* gene conferred abscisic acid sensitivity and salinity and drought tolerance (Xiang et al., 2008). In another study also in rice, the *OsbZIP71* in presence of ABA significantly increased the tolerance to salinity and drought (Liu et al., 2014). In maize (*Zea mays*), two bZIPs with activity modulated by ABA and phosphorylation are involved in the expression of ABA inducible *rab28* gene, coding for a late embryogenesis abundant protein (Nieva et al., 2005). Since it is well known that late embryogenesis abundant

proteins are important in acquisition of tolerance to abiotic stress in many plants, they will be presented in next section. Another example of bZIP highly induced by ABA and abiotic stress comprises the soybean (*Glycine max*) *GmbZIP1*, which when over-expressed in transgenic *Arabidopsis* plants affected the expression of some ABA or stress-related genes involved in regulating stomatal closure in under ABA, drought and high salt stress conditions (Gao et al., 2011).

Therefore, the above cited studies confirm the importance of bZIP proteins in response to ABA and tolerance to abiotic stress, as well as their potential as candidate genes for molecular breeding programs aiming the development of genetically modified abiotic stress-tolerant crops.

2.2. Late Embryogenesis Abundant Proteins

Another class of proteins involved in abiotic stress response in plants comprises the late embryogenesis abundant (LEA) proteins. These proteins are ubiquitously distributed in the vegetal kingdom, and are found in vascular and nonvascular plants (Amara et al., 2014). LEA proteins are small molecular weight proteins ranging from 10 to 30 kDa (He and Fu, 1996) involved in acquisition of tolerance to drought, high temperature, salinity, cold, and freezing stress in many plants (Battaglia et al., 2008; Hundertmark and Hincha, 2008; Battaglia and Covarrubias, 2013).

Studies have shown that accumulation of LEA mRNAs is often observed in embryonic tissues during the final stages of the development of seeds exposed to desiccation (Hand et al., 2011). In addition, LEA gene expression is susceptible to considerable changes associated with the acquisition of tolerance to desiccation and development of seed germination capacity (Goldberg et al., 1989).

Several classifications have been used in order to group the LEA proteins, which in the most of case have their amino acid sequences deduced from cDNA and gene sequences. Among LEA classification methods, these proteins can be characterized according to the presence of natural motifs sequences (Dure et al., 1989), or a new version of the previous nomenclature proposed by Bateman et al. (2004), where is considered the presence of the Pfam motifs. It appears in contrast to the proposal of Wise (2003) which directs the analysis to the peptide profile of proteins (POPP analysis) by means of the current bioinformatics tools.

The most of LEA proteins are highly hydrophilic with no defined secondary structure in an hydrated stage (Amara et al., 2014), being consequently included into the group of intrinsically disordered proteins (Kushwaha et al., 2013). However, certain abiotic conditions, such as drought, can lead LEA proteins to a structured folding (Tunnacliffe and Wise, 2007). Wolkers et al. (2001) detected structural changes, such as conformations in α -helices or β -sheets, or both, in LEA proteins under dehydration. Likewise, Toller et al. (2007) observed structural changes in mitochondrial LEA of pea under drought conditions. Thus, this feature can be determinant to involvement of these proteins in such physiological processes (Olvera-Carrillo et al., 2011).

Due to the high hydrophilicity of LEA proteins, it has been suggested that they can act as hydration buffers by slowing down the rate of water loss (Garay-Arroyo et al., 2000), as proposed by Manfre et al. (2006) for AtEm6, an *Arabidopsis thaliana* LEA belonging to group 1. Also, it is known that these proteins can protect other proteins and/or component cellular

from aggregation or desiccation by retaining water, sequestration of ions, and acting as molecular chaperones (Kovacs et al., 2008).

The LEA roles in response to drought and salinity have been reported in many studies. For example, the accumulation of *MeLEA3* transcripts was increased in cassava (*Manihot esculenta*) leaves under *in vitro* salt stress (Costa et al., 2011). Wise (2003) reported that the tomato (*Solanum lycopersicum*) LE25 protein expressed in yeast favored growth of the transformants cells under salt stress in comparison to the control cells. On the other hand, bacteria cells expressing the recombinant LEA5 of rice were tolerant to drought in comparison to bacteria cells control (He et al., 2012). Also in bacteria cells, the recombinant PgLEA from *Pennisetum glaucum* conferred protection against salt stress (Reddy et al., 2012).

In addition to studies with heterologous expression of LEA proteins in yeast and bacteria, transgenic plants have also confirmed their roles in increasing of tolerance to abiotic stress. For example, the over-expression of a LEA gene in rice improved drought resistance under the field conditions (Xiao et al., 2007). Likewise, the *HVA1*, a LEA gene from barley (*Hordeum vulgare*) confers dehydration tolerance in transgenic rice via cell membrane protection (Babu et al., 2004).

Hence, with above cited studies, it is possible to conclude that the LEA genes are one of the most promising genes for molecular breeding programs aiming the production of crops tolerant to abiotic stresses.

2.3. Calcineurin B-like proteins

In plants, calcium is used as a second messenger in many signal transduction pathways, including responses to abiotic stresses (Sanders et al., 2002). Calcineurin B-like (CBL) proteins are calcium-binding proteins that function as signal sensor proteins able to detect changes in the concentration of cytosolic calcium and interact with target proteins (Luan et al., 2002; Sanders et al., 2002). CBL proteins can play roles in many calcium-dependent processes by interaction with a specific class of kinases named CBL-interacting protein kinases (CIPKs) (Luan et al., 2002; Batistic and Kudla, 2004). In addition, phosphorylation process by kinases is known as one of the major post-translational protein modifications playing important roles in many processes, including response against abiotic and biotic stress.

Studies have shown the involvement of CBL proteins and CIPKs in response and tolerance to drought and salinity. For example, Cheong et al. (2003) reported that CBL1 functions as a positive regulator of salt and drought responses and a negative regulator of cold response, since transgenic *Arabidopsis* plants overexpressing this gene enhanced tolerance to salt and drought but reduced tolerance to freezing. Later, the same group reported that the constitutive overexpression of the *CBL5* gene conferred osmotic or drought stress tolerance of *Arabidopsis* plants, that also showed highest resistance to salt during the primary development such as the seeds germination (Cheong et al., 2010). In poplar plants, the expression of *PeCBL6* and *PeCBL10* genes was induced by cold, drought, or high salinity, but not by ABA treatment. Likewise, transgenic *Arabidopsis* plants overexpressing *PeCBL6* or *PeCBL10* showed enhanced tolerance to high salinity, drought, and low temperature (Li D et al., 2013).

Mahajan et al. (2006) showed that the exposition of pea plants (*Pisum sativum*) to NaCl, cold and wounding co-ordinately up-regulated the expression of *PsCBL* and *PsCIPK* genes.

Also, these genes were coordinately stimulated in response to calcium and salicylic acid, however, drought and abscisic acid have no effect on their expression.

2.4. Trehalose-6-Phosphate Synthases and Phosphatases

Trehalose is a non-reducing disaccharide composed by two glucose molecules associated through anomeric carbons that plays an important role in metabolic regulation and abiotic stress tolerance in a variety of organisms, such as bacteria, fungi and some higher plants. The reaction of trehalose synthesis is catalyzed by two key enzymes: trehalose-6-phosphate synthase enzyme (TPS) and trehalose-6-phosphate phosphatase (TPP). TPS is codified by the *OtsA* gene in bacteria (Ström and Kaasen, 1993).

Trehalose naturally acts as a protective molecule against many types of stresses in different microorganisms (Eleutherio et al., 1993; Ström and Kassen, 1993). Likewise, many *in vitro* studies have confirmed the role of trehalose in the maintenance of biological structures (El-Bashiti et al., 2005) through stabilization of membrane or enzymes involved in dehydration defense in comparison to other carbohydrates (Paiva and Panek, 1996). Some mechanisms could explain the protective effects of trehalose, such as the replacement of water and chemical stability (Roser and Colaço, 1993). Therefore, the protection of structures on water removal can be accomplished by replacing water molecules in the hydration shell (Clegg, 1986).

Trehalose accumulation in plants can elevate the tolerance to salt and drought stress. Thus, due to above cited features and other peculiar characteristics, genes coding for enzymes involved in trehalose biosynthesis have been used in the genetic engineering of plants to improve abiotic stress tolerance (Serrano et al., 1999; Penna, 2003; Reis et al., 2012).

The initial studies that showed a yeast *TPS* gene promoting drought tolerance in plants, in this case tobacco plants (*Nicotiana tabacum*), were reported by Holmström et al. (1996). Since then, other studies have confirmed the importance of trehalose in tolerance to abiotic stress in many plants. Transgenic rice plants transformed with a bifunctional fusion enzyme (TPSP) of TPS and TPP from *Escherichia coli* showed increased trehalose accumulation and abiotic stress tolerance (Jang et al., 2003). Stiller et al. (2008) reported that potato plants (*Solanum tuberosum*) expressing the *TPS* gene of yeast showed an increased relative water content during the dry season. On the other hand, Suárez et al. (2008) reported the improvement of common bean (*Phaseolus vulgaris*) to drought tolerance and grain yield by over-expression of *TPS* gene in *Rhizobium* bacteria. In this study, common bean plants inoculated with *Rhizobium* bacteria over-expressing the *OtsA* gene showed increased biomass and water retention in comparison to control plants.

Regarding endogenous plant genes involved in trehalose synthesis, the over-expression of the *OsTPP1* gene conferred stress tolerance and caused the activation of other stress responsive genes in transgenic rice plants (Ge et al., 2008). Likewise, Li W et al. (2011) reported that transgenic rice plants over-expressing the *OsTPS1* gene showed improved tolerance to cold, high salinity and drought stress. These studies also revealed that in rice transgenic plants, trehalose and proline concentrations were increased and some stress-related genes were positively regulated by *OsTPS1* gene, including *WSI18* (water stress inducible protein), *RAB16C* (responsive to ABA), *HSP70* (Heat shock 70 KDa protein), and *ELIP* (early light inducible protein) (Li et al., 2011).

Besides trehalose, proline accumulation contributes also to increasing tolerance to abiotic stress in plants (Reis et al., 2012). Therefore, some studies about proline roles in protection against abiotic stress will be presented below.

2.5. Pyrroline-5-Carboxylate Synthases and Reductases

To adapt to drought conditions, plants need to reduce the water potential to ensure a positive flow of gradients from the soil to the roots and, therefore, plant cells need to reduce the osmotic potential through the accumulation of organic ions or solutes (Morgan, 1984). In this context, several studies have confirmed the importance of low-molecular weight metabolites and osmolytes, such as mannitol, trehalose, and proline, in tolerance to abiotic stress (Reis et al., 2012).

Proline participates in the osmotic adjustment of plants, that is characterized by the synthesis and accumulation of compatible osmolytes under drought (Kishor et al., 1995; Morgan, 1984). In plants, proline is synthesized in the cytosol and mitochondria from glutamate, which is converted to proline by two successive reductions catalyzed by pyrroline-5-carboxylate synthases (P5CS) and pyrroline-5-carboxylate reductases (P5CR) (Hare et al., 1999).

Studies show that the accumulations of products from synthesis and/or degradation of proline may increase the expression of several genes regulated by stresses in rice (Iyer and Caplan, 1998). As a result, high levels of cellular proline have been related to the prevention of protein denaturation, maintenance of structures and enzymatic activity (Samuel et al., 2000), as well as membrane protection against damage caused by the reactive oxygen species (ROS) during drought conditions and high light incidence (Hamilton and Heckathorn, 2001).

Many studies have confirmed the involvement of proline in response to abiotic stress. For example, the expression of *P5CS* gene in transgenic tobacco plants increased the production of proline and induced tolerance to osmotic stress in plants (Kishor et al., 1995). In cactus pear (*Opuntia streptacantha*), salt stress enhanced the expression of *P5CS* gene and induced proline accumulation (Silva-Ortega et al., 2008). In contrast, proline levels decreased when *P5CS* gene was silenced (Székely et al., 2008). Another study showed that *P5CS* gene was induced by drought stress, salt stress and ABA treatments, in contrast to *P5CR* gene (Yoshiba et al., 1995). Aroca et al. (2008) reported that in the absence of exogenous ABA, the *LsP5CS* gene was expressed only in plants under drought conditions. In soybean, the *GmP5CS* gene expression was increased in plants under drought conditions. In addition, the accumulation of *GmP5CS* transcripts was increased in plant nodes and proline levels were enhanced in plants submitted to conditions of low water potential (Porcel et al., 2004).

Wheat plants transformed with the *Vigna aconitifolia* *P5CS* gene showed tolerance to water deficit due to the mechanisms of proline protection against ROS produced by oxidative stress (Vendruscolo et al., 2007). In another study, the over-expression of moth bean *P5CS* gene in transgenic rice plants was driven separately with a constitutive and a stress-inducible promoter. This study revealed that stress-inducible expression of the *P5CS* transgene showed significant advantages in comparison to the constitutive expression, regarding the biomass production of transgenic rice plants cultivated under stress conditions (Su and Wu, 2004). Therefore, this study indicated that the use of an stress-inducible promoter driving the over-expression of *P5CS* gene can be a better option to development of transgenic plants with tolerance to abiotic stress.

2.6. Aquaporins

Drought and salinity interfere negatively in the plant-water relationship, since they cause absorption difficulties or substantial loss of H₂O molecules (Zhu et al., 2005). The rate of water flow across cellular membranes can be modulated by aquaporins, which are expressed in many cells and tissues. Aquaporins (AQPs) are proteins that enhance the water permeability of the membranes through differences between water potentials (Tyerman et al., 2002).

AQPs play a primary role in the formation of water channels that mediate and regulate the passive flux of molecules during plant growth and development processes, such as cell elongation, stomatal movement and responses against stress (Eisenbarth et al., 2005). AQPs have highly hydrophobic nature, with six membrane-spanning domains and molecular weights ranging from 26 to 34 kDa. Belonging to the family of major intrinsic proteins (MIPs), these membrane proteins have been localized in several kingdoms (Tyerman et al., 2002) and have 35 representatives isolated from *Arabidopsis* (Quigley et al., 2001), 31 genes located in corn (Chaumont et al., 2001), at least 14 obtained from *Mesembryanthemum crystallinum* (Tyerman et al., 1999), and 36 genes in wheat, *Triticum aestivum* (Forrest and Bhawe, 2008).

Regarding the AQPs classification, they are divided into four families considering the homology between amino acid sequences and subcellular protein location (Johanson et al., 2001). They are presented as follows: Plasma Membrane Intrinsic Proteins (PIPs) (Kammerloher et al., 1994), Tonoplast Intrinsic Proteins (TIPs) (Karlsson et al., 2000), NOD26-like Intrinsic Proteins (NIPs) (Weaver et al., 1991) and Small basic Intrinsic Proteins (SIPs) (Chaumont et al., 2001).

Since aquaporins are not only regulated in transcript levels, but also in regard to their activity (Martre et al., 2002). As reported by Gao et al. (2010), in wheat mutants to tolerance and sensitive to salinity submitted to salt stress, the *TaNIP* gene was regulated positively, however, the highest levels of expression were observed in mutant salt tolerant species. Also, studies by Mahdiah et al. (2008) supported the hypothesis that these genes are regulated by drought stress, despite a decreasing in *NtPIP1* and *NtPIP2* transcripts production in tobacco roots under exposure to drought, during the recovery period, the abundance of these transcripts was replenished. It shows the regulatory role of the aforesaid abiotic stress on gene expression of aquaporins. Moreover, the same study suggested that the transport of water to cell in roots is strongly inhibited in response to water stress. Thus, it can be inferred that the PIP water channels have participation in this process.

Plants under optimal water conditions showed a high *GmPIP2* gene expression. In opposition, plants submitted to stressful conditions showed significant reductions in gene expression (Porcel et al., 2006). Thus, inhibition of expression of certain genes must be understood as a strategy for reducing water molecules losses or restriction to water absorption into cells. It may also prevent water flow decrease in the tissues (Zhu et al., 2005), as observed in desert plants and beech seeds, in which aquaporins are inhibited possibly due to some of the reasons mentioned above (North et al., 2004). Also, *PIP* transcript levels were reduced in *Arabidopsis* plants under gradual lack of water on leaves (Alexandersson et al., 2005).

Results similar to those cited above have also been found by other researchers, such as Cui et al. (2008), who determined that the effects of bean *PIP1* expression in *Arabidopsis*, such as more elongated roots or abundant lateral roots, may secondarily assist in the acquisition of drought tolerance. On the other hand, when expressed in *Arabidopsis*, *PIP1* and *PIP2* genes of rice prevented the inhibition of roots growth, and were also involved in the improvement of salt

stress tolerance (Guo et al., 2006). Therefore, aquaporin genes showing both positive and negative regulation by drought and high salinity have been identified in *Arabidopsis*, rice and corn (Kawasaki et al., 2001; Seki et al., 2002; Wang H et al., 2003).

3. APPROACHES TO IDENTIFICATION OF GENES IN RESPONSE TO DROUGHT AND SALINITY

The above presented data ratify the significant advances in gene prospection in the search for better understanding about endogenous defense mechanisms of plants. However, there is still much information to be elucidated on this research field. In this sense, strategies related to the identification of genes in response to abiotic stresses constitute the first step for prospection of genes, so that the other steps can be implemented successfully. Among them, the subsequent *in vitro* functional analyzes at the transcript level or at the protein level by heterologous expression; and *in vivo* genetic transformation studies, as will be discussed later.

The high genomic complexity of plants requires effective tools and molecular biology strategies combined with bioinformatics to identification of genes. In addition, other eukaryotes characteristics, such as the presence of multiple introns, the alternative splicing, multiple copies of some genes and the presence of non-coding DNA regions among the genes, contribute also for requirement of multiple tools for identification of genes from plant genome. Such strategies are associated with the large number of sequences with available information in public databases, as well as computational programs, which can search these databases towards the identification of genes. It is worth noting that the nucleotide sequence analysis of a gene can be used to predict the amino acid sequence of the protein, contributing therefore to deduction of protein sequences aiming to find evidence that a potential gene is being expressed (transcription and subsequent translation of the corresponding mRNA molecule).

A computational procedure which is part of initial steps to determine the gene function consists of conducting a homology search, which is based on the comparison of DNA and/or proteins sequences from a determinated organism with other sequences from different organisms. The genes evolutionarily correlated are called *homologous*. In general, the identification of a particular gene (or gene prediction) is generated by automatic annotation, which identifies all biochemically active portions of the genome by sequence processing algorithms. The homologous genes found in different organisms that diverged from the common ancestral gene due to speciation are called *orthologous*. Similarly, the computational programs can search in a genome for the presence of *paralogs*, which are genes derived from an ancestral gene via gene duplication. The homologous genes, in general, have the same function or related functions. The sequence of a newly identified gene can be compared with a database seeking to find known domains. Whether the gene sequence encodes one or more domains whose functions were previously determined, the domain function can provide important information about the possible function of a new gene. So, after a function has been assigned to a given gene, it may provide clues, such as the function of another homologous gene (Cohen, 2004; Rhee et al., 2006).

A search program commonly used by the scientific community is the BLAST (basic local alignment search tool) (Altschul et al., 1990) that finds regions with local similarity among sequences. This program compares nucleotide or protein sequences obtained by sequencing

with the database sequences, and calculates the significance of the statistical combination. BLAST can be used to infer about evolutionary and functional relationships among sequences, as well as to help with the identification of gene family members (Cohen, 2004; Rhee et al., 2006).

The functions indicated by computational methods, such as searches for homologies, phylogenetic profiles, fusion proteins and neighborhood analysis do not define completely the function of a protein or a transcript. These methods provide inferences about possible functions, which can be confirmed by detailed analyzes of biochemistry and molecular biology, such as the generation of cDNA (complementary DNA) libraries for identification of differentially expressed genes. In this case, each obtained clone is called *cDNA clone*, and the entire collection of clones derived from an mRNA population constitutes a *cDNA library*. There are several methodologies of molecular biology that can be used to generate libraries with cDNA molecules. By action of a reverse transcriptase, such molecules are synthesized from mRNA molecules extracted from a tissue. After the synthesis of cDNA molecules by reverse transcription (RT), they are converted to double-stranded molecules by DNA polymerase and inserted into a plasmid or other vector by DNA ligase, followed by bacteria transformation, and they can therefore be stored for a long period with stability, unlike mRNA molecules, which are highly unstable (Birren et al., 1999; Sambrook and Russel, 2001; Alberts et al., 2008). Because it contains only sequences corresponding to the mature mRNA molecules (without the presence of introns), it is presumably that a cDNA library can contain only molecules having the identified genes sequences of interest that are differentially expressed under a certain condition. Since a cDNA library containing genes involved in response to stress was successfully synthesized, the screening of gene of interest can be performed by using specific primers or probes for isolation of the corresponding clone.

The isolation of a DNA molecule, followed by its insertion into a cloning vector and subsequent transformation of bacterial cell for its multiplication, constitutes an important stage in the gene prospection process, aiming its molecular characterization. Such studies have been conducted to genes in response to abiotic stress, such as drought and salinity, for which there are already pre-established studies; however, for certain species of socio-economic interest, more detailed studies are needed, for instance, genes that encode LEA proteins, RING-type zinc finger proteins and translationally controlled tumor proteins (Costa et al., 2011; Reis et al., 2012; Santa Brígida et al., 2014).

Next, will be presented some methodologies for prospecting genes regarding the identification of differentially expressed genes in response to abiotic stress in plants, such as those with up and down regulation in response to drought and salinity.

3.1. Differential Display Reverse Transcription-Polymerase Chain Reaction

As aforementioned, the study of gene detection can be carried out by analysis of correspondent RNA molecules. Such approach was initially performed using techniques related to the visualization of transcripts, where the intensity of cDNA bands detected by autoradiography allows inference about the expression level of a gene relative to a sample submitted to a determined condition.

In this context, the Differential Display Reverse Transcription-Polymerase Chain Reaction (DDRT-PCR) is a sensitive methodology for identification of genes, whose expression is

changed under a given situation. This methodology allows the identification of genes differentially expressed, with up or down regulated expression and the comparison of more than two mRNA populations (Liang and Pardee, 1992; Zhang et al., 1998), including plant genes involved in plant stress responses and physiological events (Cushman and Bohnert, 2000; Yamazaki and Saito, 2002). The laboratorial steps required for development of DDRT-PCR assays include: extraction and purification of total RNA, synthesis of cDNAs from mRNAs by using a reverse transcriptase enzyme, amplification of cDNAs by PCR, visualization of PCR products, reamplification and cloning of the differentially expressed PCR products, sequencing of the differential clones, and the obtention of full-length cDNAs of interest by synthesis of full-length cDNA libraries or RACE (Rapid Amplification of cDNA Ends) assays.

Several studies aiming to identify differentially expressed genes in response to abiotic stress in plants by using DDRT-PCR were successfully performed. For example, Liu and Baird (2003) detected thirteen cDNAs in response to drought and salinity in sunflower (*Helianthus annuus*), among them, some genes were up or down regulated according to each stress condition. Using DDRT-PCR, Li and Chen (2000) evaluated the differential accumulation of the S-adenosylmethionine decarboxylase (SAMDC) transcript in rice seedlings in response to salt and drought stresses. SAMDC is an enzyme related to synthesis of polyamines, which are involved in plant response to abiotic stress (Alcázar et al., 2006). Also in rice, Kong et al. (2003) identified a gene in response to salinity, with role in the alternative oxidase (AOX) pathway, which is related to several abiotic stresses in plants (Smith et al., 2009; Li C et al., 2013).

Torres et al. (2006) used the DDRT-qPCR (quantitative PCR in real time) and identified 16 cDNAs in beans, generating the identification of 4 genes in response to the water deficit related to several functions at the physiological level. In soybean, Martins et al. (2008) identified a differential cDNA coding for a phosphatidylinositol transfer protein, possibly related to response to drought and salinity by interaction with Ca^{2+} in the cytoplasm (Knight et al., 1997; Mueller-Roeber and Pical, 2002).

3.2. cDNA-Amplified Fragment Length Polymorphism

The cDNA-amplified fragment length polymorphism (cDNA-AFLP) is a methodology for gene identification, where a double-stranded cDNA is synthesized from mRNA molecules by a transcriptase reverse enzyme, followed by digestion with restriction enzymes, separation/visualization of DNA fragments on gels or autoradiography, generating unique patterns of gene expression and allowing the transcripts identification expressed in contrasting conditions (differential bands) (Bachem et al., 1996).

Since 1996, when the cDNA-AFLP was first reported, this methodology has been improved, comprising a useful tool for quantitative transcript profiling, as reported by Breyne et al. (2003), who proposed a adapted cDNA-AFLP methodology to perform genome-wide expression studies.

Therefore, in the last years, the cDNA-AFLP methodology has been used to the identification of differentially expressed genes in response to abiotic stress in many plant species with socio-economic importance. For example, using cDNA-AFLP, Campalans et al. (2001) identified genes that were highly expressed by dehydration in almond (*Prunus dulcis*), among them, genes encoding cysteine proteases with known roles in response to drought and salinity (Groten et al., 2006). In rice roots subjected to drought, Yang et al. (2003) detected in 28 genes

with unknown functions, 18 new genes in response to water deficit, and 60 genes with known functions, such as those related to signal transduction and the cell wall biogenesis. Gupta et al. (2013) also used the cDNA-AFLP to identify genes in response to drought in *Camellia sinensis*, an important medicinal and ornamental species. In this study, a functional ontology analysis ratified the detection of genes related to metabolism of carbohydrates and others in response to stress.

Using cDNA-AFLP, Umezawa et al. (2002) isolated genes in response to the high salinity in soybean. In addition, in combination with the RNA blot analysis, these authors observed that the differential expression of soybean genes occurred according to effects closely related to the salt stress in plants: osmotic effects (related to the limited water absorption by the rhizosphere due to the salinity) and ionic effects (unbalance or intracellular toxicity due to the ions excess).

Hmida-Sayari et al. (2005) used cDNA-AFLP to evaluated potato plants stressed by salinity, finding about 5000 bands, in which 154 were up-regulated and 120 were repressed by salt stress. Among them, these authors found putative genes for protein involved in cell wall structure and proline-rich proteins, amino acids commonly accumulated in plants subjected to salt stress and water deficit (Ashraf and Foolad, 2007).

3.3. Microarray

Methodological advances achieved by studies of gene expression include tools based on nucleic acid hybridization, such as the microarray. This methodology has been used in the analysis of genome of many plants and allows to analyze thousands of genes expression in a single assay. Also, microarrays enable integration of large datasets from several experiments (Lockhart and Winzeler, 2000; Gregory et al., 2008).

Microarray analysis requires a pre-determined numbers of oligonucleotide synthesized *in situ* or cDNA affixed (probes) in a solid surface (chip), which act as molecular detectors (Holloway et al., 2002). After RNA extraction from cells or tissues to be investigated, the transcripts are labeled with fluorescent dyes and hybridized with molecules affixed in the chip. Through the base pairing described by Watson and Crick between probe and target, is possible to evaluate a quantitative abundance measure of a particular mRNA sequence in the target population. The probes which correspond to a determined transcript hybridize to their respective complementary target. Thus, since the transcripts are labeled with fluorescent dyes, the light intensity can be used as a measure of gene expression. The comparison of hybridization patterns allows the identification of mRNA molecules which differ in abundance in two or more target samples (Murphy, 2002; Malone and Oliver, 2011). From the digital information captured, several analyses can be performed to the obtention of biological information of interest.

In *Arabidopsis thaliana*, Seki et al. (2001) evaluated the expression pattern of 1300 genes under drought and cold stresses using microarrays of full-length cDNAs. By RNA gel blot analysis, these authors detected 44 cDNAs induced by drought, being 6 new genes for DREAB1A/CBF3, a transcription factor which controls the gene expression induced by stress (Liu et al., 1998). Later, the same group reported the evaluation of expression of 7000 genes of *Arabidopsis* under drought, cold and high-salinity stresses using microarray. They evaluated the differential expression of 277 and 194 genes in response to drought and salinity, respectively, among them, genes that encode LEA proteins, heat shock proteins and enzymes

involved in the osmoprotectors synthesis (Seki et al., 2002). Also in *Arabidopsis*, the use of microarray in combination with qPCR, proved to be a tool quite useful to identify up or down regulation of gene expression under salt stress and genes simultaneously regulated by such stress and by ABA (Liu Y et al., 2013).

In rice, microarray assays were performed by Rabbani et al. (2003) aiming to evaluate the expression of genes in response to drought, high-salinity, ABA and cold. They used around 1700 independent cDNAs derived from three libraries related to these stresses previously prepared, and found 36, 43, 57 and 62 sequences induced by cold, high-salinity, drought and ABA, respectively. Also in rice, a microarray covering 36926 genes was used to analyze the gene expression profile to drought and salinity in three tissues, where it was detected strong specificity related to the expression profile detection in each tissue (Zhou et al., 2007).

The microarray allowed the identification of about 168 genes up-regulated under drought stress simultaneously in three cassava genotypes (Utsumi et al., 2012), as well as more than 1811 genes in wheat under salinity (Kawaura et al., 2006). The detection of transcripts expressed in chickpea (*Cicer arietinum*) for which the quantity depended on the kind and severity of stress, where more transcripts were expressed under high-salinity rather than by drought (Mantri et al., 2007). Curiously, among transcripts differentially expressed, these authors detected more repressed transcripts than induced (Mantri et al., 2007). In potato leaves submitted to salinity, Legay et al. (2009) verified that induction of various factors related to osmotic stress response occurs by ABA dependent or independent pathways and that there is a cross-talk between response to biotic and abiotic stresses, that is associated to several adaptation mechanisms involving proteins related to pathogenesis, heat shock proteins and late embryogenesis abundant proteins.

Chen et al. (2014) identified soybean genes encoding transcription factors HD-Zip (homeodomain-leucine-zipper), which were predicted to be related to drought and salinity. Likewise, in soybean plants subjected to drought, thousands of up and down-regulated genes were identified, among homologous encoding factors of transcription in *Arabidopsis thaliana*, DREB, NAC, AREB, and ZAT/STZ (Le et al., 2012).

3.4. Suppression Subtractive Hybridization

Another tool based on the nucleic acid hybridization is the subtractive hybridization. Its objective is to generate a cDNA library, which represents transcripts found in a specific tissue or organism, in determined moment or situation. In short, the first step is the cDNA synthesis from mRNA molecules extracted from cell populations or tissues to be compared. For example, a sample under abiotic stress conditions, and another sample under normal conditions. The cDNA population in which specific transcripts are found (i.e., the samples with cDNAs subjected to determined treatment) is called *Tester*, and the cDNA population (untreated) is called *Driver*. Then, tester and driver samples are hybridized and the resulting hybrids are removed. The unhybridized cDNAs, thereby, represent genes found in Tester and absent in Driver. Then, the differentially expressed cDNAs can be cloned, generating a subtractive cDNA library (Diatchenko et al., 1996; Lukyanov et al., 2007). Since subtractive cDNA libraries are constituted by partial cDNA sequences, methodologies aiming the isolation of full-length cDNAs, such as full-length cDNA library and RACE assays, are required for further characterization of differentially expressed genes.

Along the years, several subtractive hybridization techniques were developed, aiming a better optimization and effectiveness of this methodology, among them, the suppression subtractive hybridization (SSH), based on PCR suppression, in order to avoid the amplification of undesired DNA fragments. One of the main SSH advantages is the normalization of cDNAs abundance, allowing the identification of cDNAs encoded by genes which are rarely expressed (Diatchenko et al., 1996; Lukyanov et al., 2007).

In wheat, differentially expressed genes in response to drought were isolated by SSH generating a total of 300 expressed sequence tags (ESTs), that in combination with the microarray analysis, it was detected that 30% of the genes were up-regulated and 18% were significantly down-regulated under water deficit (Way et al., 2005).

The microarray and SSH approaches were also combined in study reported by Daldoul et al. (2010), where it was detected 7 vine (*Vitis vinifera*) cDNAs up-regulated by salt stress. With the same methodological combination, 201 non-redundant genes related to the high-salinity tolerance were isolated from tomato roots, among them, transcription factors and genes involved in the SOS (salt overly sensitive) pathway (Ouyang et al., 2007).

The SSH technique was also used to generate a cDNA library enriched with soybean transcripts differentially expressed in response to drought and, by Northern hybridization, 56 ESTs were validated as up-regulated genes in response to dehydration stress (Clement et al., 2008). Rodrigues et al. (2012) evaluated soybean cultivars with tolerance and susceptibility to drought by SSH, identifying differentially expressed genes during the initial response to water deficit in roots and leaves. In another study, Vidal et al. (2012) evaluated two soybean contrasting cultivars in terms of tolerance to drought and detected about two thousands genes up-regulated genes in both cultivars.

Montalvo-Hernández et al. (2008), in combination with the microarray analysis and Northern Blot, detected 18 beans cDNAs, among which, related to aquaporins, which, as highlighted above, are important water channel proteins and correlated to plants drought tolerance (Tyerman et al., 2002). Also in beans, Recchia et al. (2013) obtained 1120 ESTs, detecting sequences correspondent to proteins related to molecular aquaporins and chaperones, being verified the greatest expression of genes identified in the cultivar drought tolerant.

In drought analysis, the SSH was used to analyze the differential expression in sugarcane (*Saccharum officinarum*), being identified genes responsible for the sucrose-phosphate synthesis (Almeida et al., 2013). Ding et al. (2014) isolated and characterized genes in response to drought by SSH in peanut (*Arachis hypogaea*) involved mainly in the cell structure and metabolism, as well as some of them also associated to other stresses, such as high-salinity, cold, and stress to high temperatures. In addition, the analysis of seven genes by RT-PCR confirmed the genes differential expression in response to drought in the root (Ding et al., 2014).

In a study reported by Wu et al. (2005), SSH was used for evaluation of differentially expressed genes of a salt tolerant rice cultivar under NaCl treated, in comparison to non-treated plants. After a BLAST search in GenBank Databases, 31 cDNAs sequences were identified, among them, those involved in oxidative stress response, which is related to high salinity and other abiotic stresses (Wu et al., 2005).

3.5. Next Generation Sequencing

For a comprehensive understanding of genetic mechanisms involved development and response of plants to environmental stresses, a global gene expression analysis, associated with *in silico* approaches for sequencing analysis, has been fundamental. The sequencing process include a number of methods, which are broadly grouped from preparation of template to images supply and data analysis. The combination of specific protocols distinguish a technology from another and the kind of data produced by each platform.

By the wide variety of approaches adopted aiming large-scale studies of gene expression, bioinformatics tools associated to next generation sequencing (NGS) have significantly collaborated to studies of transcriptomes.. Although the first generation sequencing (Sanger) produces results to the biological researches, including the gene prospection related to plants resistance (Varshney et al., 2009; Recchia et al., 2013), more and more NGS platforms tend to be adopted, due to the higher generation of information combined with the higher results accuracy, covering the transcripts sequencing, as well as the genome resequencing or sequencing yet unknown (*de novo*).

The NGS platforms, available since 2005, promote the DNA sequencing providing information about millions of base pairs in a single run. Among them: the 454 System (Roche) based on the pyrosequencing; Solexa (Illumina) which, like the Sanger sequencing, is performed using DNA polymerase and nucleotide terminators labeled with different fluorophores and SOLiD System (Applied Biosystems), where the sequencing reaction is catalyzed by a DNA ligase, instead of a DNA polymerase (Mardis, 2008; Shendure and Ji, 2008).

The combination of NGS approaches with other methodologies has also been used as a strategy to reach the biological results of interest. The SAGE technology, serial analysis of gene expression, is based on high scale counting of specific regions (*tags*), obtained from a transcripts population. A latter connection (concatenation) of *tags* allows the efficient analysis of transcripts in a serial mode, by sequencing of multiple tags contained in a single clone (Velculescu et al., 1995; Chen et al., 2000). The methodological adaptations performed by Matsumura et al. (2005) to this method (*SuperSAGE*), in combination with the NGS, can contribute to the extent of gene identification involved in abiotic stress (Kido et al., 2013).

To identify genes in response to abiotic stresses, a sequencing focus has being the comparison of contrasting transcriptional profiles, as for instance, genotypes and distinct physiological conditions. In this context, the RNA-sequencing analysis (RNA-Seq) is the direct sequencing of transcripts by high-throughput sequencing technologies, where the mRNAs molecules are randomly fragmented and converted to cDNA fragments, which are subject to in-depth sequencing. The sequences are assembled and annotated using genome sequences as reference or *de novo* assemble. This methodology is very valuable, once it provides complete transcriptomes sequences, which can be used for various purposes, among them, the mRNAs quantitative characterization present in the transcriptome (Wang et al., 2009; Malone and Oliver, 2011).

Many studies aiming a global gene expression analysis of plants under abiotic stress have been reported. In maize, Xu et al. (2014) aimed the identification of candidate genes to drought tolerance and nsSNPs (non-synonymous single nucleotide polymorphisms), that interfere in the amino acid sequence, and can cause a modification in the structure and/or function of the protein. Then, these authors performed the resequencing of maize lines genome using the

Illumina Hiseq 2000 platform and further analysis of RNA-Seq, in combination with cluster grouping and CV (Common variants) analysis to identify nsSNPs and their associations with genes in response to drought. The authors identified a total of 524 nsSNPs associated with 271 candidate genes with different functions related to abiotic stresses, such as those related to the oxidative-reductive balance, commonly altered by the water stress.

In tomato, Wang et al. (2013) used the RNA-Seq to evaluate the gene expression in leaves sprayed with ABA, and detected various genes in response to drought and salinity. In soybean, the transcriptional profile of seedlings under salt and drought stress was analyzed by Fan et al. (2013) using the Illumina platform that detected 874 and 535 up-regulated genes under salt and drought stress in leaves, respectively, and 1822 and 1690 genes up-regulated under salt and drought stress in roots, respectively.

After treatment of mustard seedlings with high temperature, drought and salinity, Bhardwaj et al. (2014) identified 126 micro RNAs (miRNAs) by using the Illumina GA IIX sequencer. Li H et al. (2011) by using the Solexa sequencing technology, found 71 and 50 miRNAs differentially expressed in soybeans under drought and salinity, respectively. Kulcheski et al. (2011) compared soybean seedlings susceptible and tolerant to drought and detected several families of new miRNAs. In addition, RT-qPCR assays revealed that the majority of these miRNAs was up-regulated under stress in susceptible plants, and down-regulated in tolerant plants (Kulcheski et al., 2011), confirming the important roles of these molecules on the regulation of genes related to drought stress.

In common beans, His et al. (2014) detected 441 putative sequences of transcription factors related to the protection against high salinity by using the NGS. Wu et al. (2014) performed a large scale expression analysis in response to drought in common beans using RNA-Seq and validation by RT-PCR, where genes coding for transcription factors, such as WRKY, NAC and DREB, and proteins involved in pathways of phytohormones, such as auxin, gibberellins, and ethylene, were detected.

4. PLANT GENETIC TRANSFORMATION

After the identification of genes related to abiotic stress by using molecular approaches and *in silico* analysis, a more detailed functional characterization of these genes is required aiming an efficient prospection of genes of interest. In this context, the utilization of genetic transformation of plants is one of the most powerful approaches for functional studies of genes with potential for agriculture.

The plant genetic transformation allows the evaluation of the plant phenotype of transgenic plants achieved by the selection of successfully transformed cells (with the exogenous DNA integrated into the plant genome) and the plant regeneration (performed by tissue culture procedures aiming the formation of a whole plant from a single plant cell), requiring that the introduced gene into the plant genome (transgene) is properly expressed and transmitted across generations.

Aiming to understand how pre-identified genes are important at functional level, a useful strategy has been their over-expression in transgenic plants. Analysis based on gene silencing constitutes also an important tool to elucidate the functional relevance of certain genes that respond to abiotic stresses. Therefore, functional studies can be performed by over-expression

and/or silencing of genes with potential roles in response to abiotic stress, that were pre-identified by molecular and *in silico* methodologies.

In over-expression strategy, the coding region of gene to be expressed (transgene) is cloned into a vector under the control of a very strong promoter, which can be constitutive, inducible or tissue-specific, followed by introduction into the plant genome by genetic transformation methods. Despite constitutive promoters, such as the CaMV35S, can be very efficient in studies involving the over-expression of genes related to tolerance to abiotic stress, in some cases, stress-inducible promoters seem to comprise a better option in such studies. For example, Su and Wu (2004) reported that the over-expression of the *P5CS* gene, involved in synthesis of proline, in transgenic rice was more efficient when driven by a stress-inducible promoter than a constitutive promoter.

For silencing of genes, cassettes containing fragments of the gene of interest in sense and antisense orientations separated by an intron have been used (Zalewski et al., 2012). After plant transformation with cassettes, a double-stranded hairpin RNA (hpRNA) is formed and induces a silencing signal, which is short interfering RNA (siRNA). The formation of the hybrid molecule siRNA-mRNA of gene to be silenced induces the degradation process and post-transcriptional gene silencing (PTGS) (Chuang and Meyerowitz, 2000; Watson et al., 2005; Zalewski et al., 2012). This allows the understanding how the partial or total inhibition of a gene product can affect the plant phenotype.

The strategies for over-expression and silencing of genes can be performed by indirect and direct methods of plant genetic transformation, such as the *Agrobacterium* and biolistic that have been the most frequently used methods in molecular breeding programs of many important crops.

4.1. *Agrobacterium*-Mediated Genetic Transformation

Agrobacterium tumefaciens is a widespread naturally occurring soil bacterium, which shows the ability to introduce exogenous DNA into the plant cell. Since the first study reporting the use of *A. tumefaciens* in genetic transformation of tobacco plants by Herrera-Estrella (1983), the plant transformation mediated by *Agrobacterium* has become the most used method for the introduction of exogenous DNA into plant cells, followed by regeneration of transgenic plants. Therefore, along of years, the *Agrobacterium* method has allowed the introduction of genes conferring traits of interest for agriculture into many plants, including genes related to tolerance to abiotic stress (Cheong et al., 2003; Gao et al., 2011; Li D et al., 2013).

Agrobacterium genus has five species with the ability to penetrate into the plants, especially in dicotyledonous. Therefore, initially it was believed that *Agrobacterium* was able to infect only dicots, however, it was later established that some monocotyledonous plants can also be transformed using *Agrobacterium* method.

Among *Agrobacterium* species, the *A. tumefaciens* is the most used in studies of genetic transformation of plants. However, the *Agrobacterium rhizogenes*, a soil bacterium which stimulates in the host the proliferation of secondary roots from the infection point, can also be used for this purpose. Currently, the species *A. tumefaciens* and *A. rhizogenes* concentrate more than 600 host plant species. These bacteria contain a large plasmid, involved in development of tumor, named Ti (Tumor inducing) plasmid for *A. tumefaciens*, or Ri (Root inducing) plasmid, in the case of *A. rhizogenes*. During infection, the T-DNA (Transferred-DNA), a

mobile segment of Ti or Ri plasmid, is transferred to the plant cell nucleus and integrated into the plant chromosome. The T-DNA contains two types of genes: the oncogenic genes, encoding for enzymes involved in the synthesis of auxins and cytokinins and responsible for tumor formation; and the genes encoding for the synthesis of opines (de la Riva et al., 1998).

Along the years, many specific vectors for the *Agrobacterium*-mediated genetic transformation have been developed, taking into account that any exogenous DNA placed between the T-DNA borders can be transferred to plant cells. For example, the pCAMBIA vectors from the CAMBIA Institute (<http://www.cambia.org/>) comprised one of the most currently used vectors in studies of plant genetic transformation. In brief, in the transformation process using the *Agrobacterium*, a strain bacterium transformed with the vector containing the gene of interest is placed in contact with any explants, which have potential for plant regeneration, such as cotyledon, zygotic embryos, segments of young leaves, internodes and others. After contact with plant tissue, the *Agrobacterium* begins the stage of infection by transferring DNA and transforming the host genome. After the infection, the tissue culture techniques are required in order to regenerate a whole transformed plant containing the transgene. Selection agents and antibiotics are used to elimination of *Agrobacterium* bacteria and selection of transformed cells. Finally, molecular procedures are used to evaluate the transformed plants, such as DNA and RNA blot hybridization and PCR assays.

Advantages of using *Agrobacterium*-mediated transformation in comparison to other methods of transformation include a reduced copy number of transgene, and intact and stable integration of the transgene into the plant genome (Jones et al., 2005). In studies aiming the acquisition or increasing tolerance of plants to abiotic stresses, *Agrobacterium* method showed high efficiency and wide applicability, with a high potential for integration of genes in plants to monitor the phenotype and perform functional studies, in both dicots and monocots plants. For example, Xiang et al. (2008) obtained transgenic rice plants with tolerance to drought and high-salinity stresses by over-expression of the *OsbZIP23* gene using the *Agrobacterium* method. Likewise, Zhou et al. (2012) produced transgenic tobacco plants over-expressing the *TaAQP7* gene (aquaporin) with increased resistance to drought. In millet plants (*Setaria italica*), the over-expression of the gene *SiLEA14* by *Agrobacterium*-mediated genetic transformation provided a high tolerance to drought and salinity (Wang et al., 2014).

4.2. Biolistic-Mediated Genetic Transformation

The genetic transformation of plants may also be performed by direct methods, among them, the biolistic has shown high efficiency and broad applicability in plant molecular breeding aiming the resistance to abiotic stresses, our goal in this chapter.

The biolistic, also known as microparticle bombardment, was firstly reported by Sanford et al. (1987), who invented an air pistol modified to fire dense tungsten particles used in delivering of substances into cells and tissues. Since then, the ‘biolistic apparatus’ has been modified and improved, and currently many models of microparticle bombardment apparatus are commercially available for scientific community. Many of them use a helium pulse, at high pressure, to accelerate gold or tungsten micro-particles containing the DNA molecule to be incorporated into the target tissue or cells. Then, the accelerated particles penetrate both the cell wall and membranes, and the DNA separates from the metal and can be integrated into the genetic material inside the nucleus.

Among advantages of plant genetic transformation mediated by biolistic, it is useful to practically all species of plants, in special the monocots. Then, with the biolistic is possible to maximize the gene transfer to a variety of cells, tissues and plant species. Also, the biolistic is considered a rapid and very simple procedure, being very useful in the direct transformation of totipotent tissues, such as pollen, embryos, meristems and morphogenic cell cultures. In addition, the biolistic process is suitable for organelle transformation (Sanford, 1990), that is particularly very appreciable, since the genetic transformation of organelles has emerged as a powerful tool in genetic engineering of plants in the last years. The expression of exogenous genes in organelles, such as chloroplasts, offers several advantages over expression in the nucleus, such as high-level expression and no transfer of genes expressed in the plastome through pollination to weedy or wild relatives of the transgenic crop (Wani et al., 2010). On the other hand, the use of biolistic requires the acquisition of special instrumentation (biolistic apparatus) (Sanford, 1990). Also, for performing of studies using biolistic, frequently, is necessary the development or optimization of specific protocols for genetic transformation, taking into account various physical and biological factors for the success of the method, such as particle size to be used, the speed employed on the particle, the type of apparatus to be used, precipitation method, plant species, explant type, among other issues (Sanford, 1990).

Many authors have reported the efficient use of the biolistic in functional studies of genes potentially related to plant response to abiotic stress. For example, Xu et al. (1996) reported the generation of transgenic rice plants with resistance to drought and salinity by expression of LEA gene via biolistic-mediated genetic transformation. Likewise, Ganguly et al. (2012) reported the over-expression of *Rab16A* gene (group 2 LEA protein) in transgenic rice plants with increased salt tolerance.

A valuable strategy for studies involving the plant genetic transformation is regarding the comparison of efficiency between different methods of transformation. Then, a gene of interest is transformed by using different transformation methods and their results are comparatively evaluated. In this context, the suitability of the *Agrobacterium* method over the biolistic has been shown by some studies. For example, Zalewski et al. (2012) reported that the *Agrobacterium* method was more efficient in silencing of the developmentally regulated *HvCKX2* gene in barley in comparison to the biolistic, which showed low productivity and disturbances in plant development.

Regarding to the use of such comparative study using genes involved in tolerance to abiotic stress, Shou et al. (2004a) detected that the expression of the *Nicotiana* protein kinase (*NPK1*) gene enhanced drought tolerance in transgenic maize. Also, the same authors reported that maize transgenic plants over-expressing the *NPK1* gene by *Agrobacterium*-mediated genetic transformation showed lower transgene copies, and higher and more stable gene expression than transgenic plants generated by biolistic. It is known that the introduction of a reduced copy number of transgene comprises one of the advantages of *Agrobacterium* method, since the integration of high numbers of transgene copies often leads to transgene silencing (Shou et al., 2004b; Travella et al., 2005).

CONCLUSION

The several topics discussed in this chapter confirm the importance of gene prospection studies, deeply related to the scientific efforts in the search for crop genetic improvement to overcome the abiotic stresses. Strategies for molecular breeding of agriculturally important plants aiming the acquisition or increasing of tolerance to abiotic stress include the over-expression and/or silencing of specific genes. Despite the identification of several genes coding for proteins related to response to abiotic stress, such as drought and salinity, this prospection still has a long way to go, not only to continue identifying more genes, which have their expression regulated by stress, but also to unravel with more detail the physiological and molecular roles of genes and proteins already identified. Additional studies involving functional analysis and identification of genes essential for abiotic resistance, seeking to better understand the optimal level of expression of various genes so that there is optimal interaction of their gene products compatible with the generation of well adapted plants, thereby overcoming the adverse conditions at the field level.

REFERENCES

- Agarwal, P. K., Agarwal, P., Reddy, M. K. & Sopory, S. K. (2001). Role of DREB transcription factors in abiotic and biotic stress tolerance in plants. *Plant Cell Rep.*, 25, 1263–1274.
- Alberts, B., Johnson, A., Lewis, J., Raff, M., Roberts, K. & Walter, P. (2008). Molecular biology of the cell. 5^a ed. New York, USA: *Garland Science*.
- Alcázar, R., Marco, F., Cuevas, J. C., Patron, M., Ferrando, A., Carrasco, P., Tiburcio, A. F. & Altabella, T. (2006). Involvement of polyamines in plant response to abiotic stress. *Biotechnol. Lett.*, 28, 1867–1876.
- Alexandersson, E., Fraysse, L., Sjøvall-Larsen, S., Gustavsson, S., Fellert, M., Karlsson, M., Johanson, U. & Kjellbom, P. (2005). Whole gene family expression and drought stress regulation of aquaporins. *Plant Mol. Biol.*, 59, 469–484.
- Almeida, C. M. A., Donato, V. M. T. S., Amaral, D. O. J., Lima, G. S. A., Brito, G. G., Lima, M. M. A., Correia, M. T. S. & Silva, M. V. (2013). Differential gene expression in sugarcane induced by salicylic acid and under water deficit conditions. *Agric. Sci. Res. J.*, 3 (1), 38–44.
- Altschul, S. F., Gish, W., Miller, W., Myers, E. W. & Lipman, D. J. (1990). Basic local alignment search tool. *J. Mol. Biol.*, 215, 403–410.
- Amara, I., Zaidi, I., Masmoudi, K., Ludevid, M. D., Pagès, M., Goday, A. & Brini, F. (2014). Insights into Late Embryogenesis Abundant (LEA) proteins in plants: from structure to the functions. *Am. J. Plant Sci.*, 5, 3440–3455.
- Aroca, R., Vernieri, P. & Ruiz-Lozano, J. M. (2008). Mycorrhizal and non-mycorrhizal *Lactuca sativa* plants exhibit contrasting responses to exogenous ABA during drought stress and recovery. *J. Exp. Bot.*, 59, 2029–2041.
- Ashraf, M. & Foolad, M. R. (2007). Roles of glycine betaine and proline in improving plant abiotic stress resistance. *Environ. Exp. Bot.*, 59, 206–216.

- Babu, R. C., Zhang, J., Blum, A., Ho, T. H. D., Wu, R. & Nguyen, H. T. (2004). *HVA1*, a LEA gene from barley confers dehydration tolerance in transgenic rice (*Oryza sativa* L.) via cell membrane protection. *Plant Sci.*, 166, 855–862.
- Bachem, C. W., Hoeven, R. S., de Bruijn, S. M., Vreugdenhil, D., Zabeau, M. & Visser, R. G. (1996). Visualization of differential gene expression using a novel method of RNA fingerprinting based on AFLP: analysis of gene expression during potato tuber development. *Plant J.*, 9, 745–53.
- Bader, A. G. & Vogt, P. K. Leucine zipper transcription factors: Bzip proteins, In: Ganten, D., Ruckpaul, K. (eds.). *Encyclopedic reference of genomics and proteomics in molecular medicine*. Heidelberg, Germany: Springer; 2006, 964–967.
- Bateman, A., Coin, L., Durbin, R., Finn, R. D., Hollich, V., Griffiths-Jones, S., Khanna, A., Marshall, M., Moxon, S., Sonnhammer, E. L. L., Studholme, D. J., Yeats, C. & Eddy, S. R. (2004). The Pfam protein families database. *Nucleic Acids Res.*, 32, 138–141.
- Batistic, O. & Kudla, J. (2004). Integration and channeling of calcium signaling through the CBL calcium sensor/CIPK protein kinase network. *Planta*, 219, 915–924.
- Battaglia, M. & Covarrubias, A. A. (2013). Late embryogenesis abundant (LEA) proteins in legumes. *Front. Plant Sci.*, 25, 4, 190.
- Battaglia, M., Olvera-Carrillo, Y., Garcíarrubio, A., Campos, F. & Covarrubias, A. A. (2008). The enigmatic LEA proteins and other hydrophilins. *Plant Physiol.*, 148, 6–24.
- Bhardwaj, A. R., Joshi, G., Pandey, R., Kukreja, B., Goel, S., Jagannath, A. K., Katiyar-Agarwal, S. & Agarwal, M. A. (2014). Genome-wide perspective of miRNAome in response to high temperature, salinity and drought stresses in *Brassica juncea* (Czern) L. *PLoS ONE*, 9 (3).
- Birren, B., Green, E. D., Klapholz, S., Myers, R. M. & Roskams, J. (1999). Genome analysis laboratory manual. Vol. 2: Detecting genes. New York, USA: Cold Spring Harbor Laboratory Press.
- Breyne, P., Dreesen, R., Cannoot, B., Rombaut, D., Vandepoele, K., Rombauts, S., Vanderhaeghen, R., Inzé, D. & Zabeau, M. (2003). Quantitative cDNA-AFLP analysis for genome-wide expression studies. *Mol. Gen. Genomics*, 269, 173–179.
- Campalans, A., Pages, M. & Messegueir, R. (2001). Identification of differentially expressed genes by the cDNA-AFLP technique during dehydration of almond (*Prunus amygdalus*). *Tree Physiology*, 21, 633–643.
- Chaumont, F., Barrieu, F., Wojcik, E., Chrispeels, M. J. & Jung, R. (2001). Aquaporins constitute a large and highly divergent protein family in maize. *Plant Physiol.*, 12, 1206–1215.
- Chaves, M. M., Flexas, J. & Pinheiro, C. (2009). Photosynthesis under drought and salt stress: regulation mechanisms from whole plant to cell. *Ann. Botany*, 103, 551–560.
- Chen J. J., Rowley, J. D. & Wang, S. M. (2000). Generation of longer cDNA fragments from serial analysis of gene expression tags for gene identification. *Proc. Natl. Acad. Sci.*, 97, 349–353.
- Chen, X., Chen, Z., Zhao, H., Zhao, Y., Cheng, B. & Xiang, Y. (2014). Genome-wide analysis of soybean HD-Zip gene Family and expression profiling under salinity and drought treatments. *PLoS ONE*, 9 (2).
- Cheong, Y. H., Kim, K. N., Pandey, G. K., Gupta, R., Grant, J. J. & Luan, S. (2003). CBL1, a calcium sensor that differentially regulates salt, drought, and cold Responses in *Arabidopsis*. *Plant Cell*, 15, 1833–45.

- Cheong, Y. H., Sung, S. J., Kim, B. G., Pandey, G. K., Cho, J. S., Kim, K. N. & Luan, S. (2010). Constitutive overexpression of the calcium sensor CBL5 confers osmotic or drought stress tolerance in *Arabidopsis*. *Mol. Cells*, 29, 159-65.
- Choi, H., Hong, J., Ha, J., Kang, J. & Kim, S. Y. (2000). ABFs, a family of ABA-responsive element binding factors. *J. Biol. Chem.*, 275, 1723–1730.
- Chuang, C. F. & Meyerowitz, E. M. (2000). Specific and heritable genetic interference by double-stranded RNA in *Arabidopsis thaliana*. *Proc. Nat. Acad. Sci.*, 9, 4985-4990.
- Clegg, J. S. The physical properties and metabolic status of *Artemia* cysts at low water contents: “The water replacement hypothesis”. In: & Leopold, A. C. (ed.). *Membranes, metabolism and dry organisms*. Ithaca, USA: Cornell University Press, 1986, 169-187.
- Clement, M., Lambert, A., Herouart, D. & Boncompagni, E. (2008). Identification of new up-regulated genes under drought stress in soybean nodules. *Gene*, 426, 15-22.
- Cohen, J. Bioinformatics – An introduction for computer scientists. (2004). *ACM Computing Surveys*, 36, (2), 2004, 122–158.
- Cohen, S. N. (2013). DNA cloning: A personal view after 40 years. *Proceedings of the National Academy of Sciences*, 110 (39), 15521–15529.
- Costa, C. N. M., Santa Brígida, A. B., Borges, B. N., Menezes Neto, M. A., Carvalho, L. J. C. B. & de Souza, C. R. B. (2011). Levels of MeLEA3, a cDNA sequence coding for na atypical late embryogenesis abundant protein in cassava, increase under *in vitro* salt stress treatment. *Plant Mol. Biol. Rep.*, 29, 997–1005.
- Cui, X. H., Hao, F. S., Chen, H., Chen, J. & Wang, X. C. (2008). Expression of the *Vicia faba* VfPIP1 gene in *Arabidopsis thaliana* plants improves their drought resistance. *J. Plant Res.*, 121, 207–214.
- Cushman, J. C. & Bohnert, H. J. (2000). Genomic approaches to plant analysis of RNA accumulation in arbuscular mycorrhiza of pea stress tolerance. *Curr. Opin. Plant Biol.*, 3, 117–124.
- Cutler, S. R., Pedro L. Rodriguez, P., Ruth, R., Finkelstein, R. R. & Abrams, S. R. (2010). Absciscic acid: Emergence of a core signaling network. *Annu. Rev. Plant Biol.*, 61, 651–79.
- Daldoul, S., Guillaumie, S., Reustle, G. M., Krczal, G., Ghorbel, A., Delrot, S., Mliki, A. & Hofer, M. U. (2010). Isolation and expression analysis of salt induced genes from contrasting grapevine (*Vitis vinifera* L.) cultivars. *Plant Sci.*, 179, 489-498.
- de la Riva, G. A., González-Cabrera, J., Vázquez-Padrón, R. & Ayra-Pardo, C. (1998). *Agrobacterium tumefaciens*: a natural tool for plant transformation. *Electron. J. Biotechnol.*, 1 (3), 118-133.
- Diatchenko, L., Lau, Y. F., Campbeel, A. P., Chenchik, A., Moqadam, F., Huang, B., Lukyanov, S., Lukyanov, K., Gurskaya, N., Sverdlo, E. D. & Siebert, P. D. (1996). Suppression subtractive hybridization: A method for generating differentially regulated or tissue-specific cDNA probes and libraries. *Proc. Natl. Acad. Sci. U. S. A.*, 93, 6025–6030.
- Ding, H., Zhang, Z.M., Qin, F. F., Dai, L. X., Li, C. J., Ci, D.W. & Song, W. W. (2014). Isolation and characterization of drought-responsive genes from peanut roots by suppression subtractive hybridization. *Electron. J. Biotechnol.*, 17, 304-310.
- Doebley, J. F., Gaut, B. S. & Smith, B. D. (2006). The molecular genetics of crop domestication. *Cell*, 127, 1309-1321.
- Dure, L., Crouch, M., Harada, J., Ho, T-HD., Mundy, J., Quatrano, R., Thomas, T. & Sung, Z. R. (1989). Common amino acid sequence domains among the LEA proteins of higher plants. *Plant Mol. Biol.*, 12, 475–486.

- Eisenbarth, D. A. & Weig, A. R. (2005). Dynamics of aquaporins and water relations during hypocotyl elongation in *Ricinus communis* L. seedlings. *J. Exp. Bot.*, 56, 1831-1842.
- El-Bashiti, T., Hamamc, H., Öktem, H. A. & Yücel, M. (2005). Biochemical analysis of trehalose and its metabolizing enzymes in wheat under abiotic stress conditions. *Plant Sci.*, 169, 47-54.
- Eleutherio, E. C. A., Araujo, P. S. & Panek, A. D. (1993). Protective role of trehalose during heat stress in *Saccharomyces cerevisiae*. *Cryobiology*, 30, 591-596.
- Fan, X., W, J., Yang, N., Dong, Y., Liu, L., Wang, F., Wang, N., Chen, H., Liu, W., Sun, Y., Wu, J. & Li, H. (2013). Gene expression profiling of soybean leaves and roots under salt, saline-alkali and drought stress by high-throughput Illumina sequencing. *Gene*, 512 (2), 392-402.
- Forrest, K. L. & Bhav, M. (2008). The PIP and TIP aquaporins in wheat form a large and diverse family with unique gene structures and functionally important features. *Funct. Integr. Genomics*, 8, 115-133.
- Ganguly, M., Datta, K., Roychoudhury, A., Gayen, D., Sengupta, D. N. & Datta, S. K. (2012). Overexpression of *Rab16A* gene in indica rice variety for generating enhanced salt tolerance. *Plant Signal Behav.*, 7 (4), 502-509.
- Gao, S. Q., Chen, M., Xu, Z. S., Zhao, C. P., Li, L., Xu, H. J., Tang, Y. M., Zhao, X. & Ma, Y. Z. (2011). The soybean GmbZIP1 transcription factor enhances multiple abiotic stress tolerances in transgenic plants. *Plant Mol. Biol.*, 75 (6), 537-53.
- Gao, Z., He, X., Zhao, B., Zhou, C., Liang, Y., Ge, R., Shen, Y. & Huang, Z. (2010). Overexpressing a putative aquaporin gene from wheat, *TaNIP*, enhances salt tolerance in transgenic *Arabidopsis*. *Plant Cell Physiol.*, 51, 767-75.
- Garay-Arroyo, A., Colmenero-Flores, J., Garcarrubio, A. & Covarrubias, A. A. (2000). Highly hydrophilic proteins in prokaryotes and eukaryotes are common during conditions of water deficit. *J. Biol. Chem.*, 275, 5668-5674.
- Ge, L. F., Chao, D. Y., Shi, M., Zhu, M. Z., Gao, J. P. & Lin, H. X. (2008). Overexpression of the trehalose-6-phosphate phosphatase gene *OsTPP1* confers stress tolerance in rice and results in the activation of stress responsive genes. *Planta*, 228, 191-201.
- Godfray, H. C., Beddington, J. R., Crute, I. R., Haddad, L., Lawrence, D., Muir, J. F., Pretty, J., Robinson, S., Thomas, S. M. & Toulmin, C. (2010). Food security: The challenge of feeding 9 billion people. *Science*, 327, 812-818.
- Goldberg, R. B., Barker, S. J. & Perez-Grau, L. (1989). Regulation of gene expression during plant embryogenesis. *Cell*, 56, 149-160.
- Golldack, D., Li, C., Mohan, H. & Probst, N. (2014). Tolerance to drought and salt stress in plants: unraveling the signaling networks. *Front. Plant Sci.*, DOI: 10.3389/fpls.2014.00151.
- Gregory, B. D., Yazaki, J. & Ecker, J. R. (2008). Utilizing tiling microarrays for whole-genome analysis in plants. *Plant J.*, 53 (4), 636-644.
- Groten, K., Dutilleul, C., van Heerden, P. D. R., Vanacker, H., Bernard, S., Finkemeier, I., Dietz, K.-J. & Foyer, C. H. (2006). Redox regulation of peroxiredoxin and proteinases by ascorbate and thiols during pea root nodule senescence. *FEBS Lett.*, 580, 1269-1276.
- Guo, L., Wang, Z., Lin, H., Cui, W., Chen, J., Liu, M., Chen, Z. L., Qu, L. J. & Gu, H. (2006). Expression and functional analysis of the rice plasma-membrane intrinsic protein gene family. *Cell Res.*, 16, 277-286.

- Gupta, S., Bharalee, R., Bhorali, P., Das, S. K., Bhagawati, P., Bandyopadhyay, T., Gohain, B., Agarwal, N., Ahmed, P., Borchetia, S., Kalita, M. C., Handique, A. K. & Das, S. (2013). Molecular analysis of drought tolerance in tea by cDNA-AFLP based transcript profiling. *Mol. Biotechnol.*, 53 (3), 237-248.
- Hamilton, E. W. & Heckathorn, S. A. (2001). Mitochondrial adaptations to NaCl. Complex I is protected by anti-oxidants and small heat shock proteins, whereas complex II is protected by a proline and betaine. *Plant Physiol.*, 126, 1266-74.
- Hand, S.C., Menze, M. A., Toner, M., Boswell, L. & Moore, D. (2011). LEA proteins during water stress: Not just for plants anymore. *Ann. Rev. Physiol.*, 73, 115-134.
- Hare, P. D., Cress, W. A. & van Staden, J. (1999). Proline synthesis and degradation: a model system for elucidating stress-related signal transduction. *J. Exp. Bot.*, 50, 413-434.
- He, J. X. & Fu, J. R. (1996). Research progress in Lea proteins of seeds. *Plant Physiol. Commun.*, 32, 241-246.
- He, S., Tan, L., Hu, Z., Chen, G., Wang, G. & Hu, T. (2012). Molecular characterization and functional analysis by heterologous expression in *E. coli* under diverse abiotic stresses for OsLEA5, the atypical hydrophobic LEA protein from *Oryza sativa* L. *Mol. Genet. Genomics*, 287, 39-54.
- Herrera-Estrella, L. (1983). *Transfer and expression of foreign genes in plants*. PhD thesis. Laboratory of Genetics, Gent University, Belgium.
- Hirayama, T. & Shinozaki, K. (2010). Research on plant abiotic stress responses in the post-genome era: past, present and future. *Plant J.*, 61, 1041-1052.
- His, M. C., Canher, B., Niron, H. & Turet, M. (2014). Transcriptome analysis of salt tolerant common bean (*Phaseolus vulgaris* L.) under saline conditions. *PLoS ONE*, 9 (3).
- Hmida-Sayari, A., Costa, A., Leone, A., Jaoua, S. & Gargouri-Bouزيد, R. (2005). Identification of sal stress-induced transcripts in potato leaves by cDNA-ALFP. *Mol. Biotechnol.*, 30 (1), 31-40.
- Holloway, A. J., Van Laar, R. K., Tothill, R. W. & Bowtell, D. D. L. (2002). Options available – from start to finish – for obtaining data from DNA microarrays II. *Nature Genetics*, 32, 481-489.
- Holmström, K. O., Mäntylä, E., Welin, B., Mandal, A., Palva, E. T., Tunnela, O. E. & Londesborough, J. (1996). Drought tolerance in tobacco. *Nature*, 379, 683-684.
- Hundertmark, M. & Hincha, D. K. (2008). LEA (late embryogenesis abundant) proteins and their encoding genes in *Arabidopsis thaliana*. *BMC Genomics*, 9, 1-22.
- Iyer, S. & Caplan, A. (1998). Products of proline catabolism can induce osmotically regulated genes in rice. *Plant Physiol.*, 116, 203-211.
- Jang, I. C., Oh, S. J., Seo, J. S., Choi, W. B., Song, S. I., Kim, C. H., Kim, Y. S., Seo, H. S., Choi, Y. D., Nahm, B. H. & Kim, J. K. (2003). Expression of a bifunctional fusion of the *Escherichia coli* genes for trehalose-6-phosphate synthase and trehalose-6-phosphate phosphatase in transgenic rice plants increases trehalose accumulation and abiotic stress tolerance without stunting growth. *Plant Physiol.*, 131, 516-524.
- Johanson, U., Karlsson, M., Johansson, I., Gustavsson, S., SJövall, S., Fraysse, F., Weig, A. R. & Kjellbom, P. (2001). The complete set of genes encoding major intrinsic proteins in *Arabidopsis* provides a framework for a new nomenclature for major intrinsic proteins in plants. *Plant Physiol.*, 126, 1358-1369.

- Jones, D. H., Doherty, A. & Wu, H. (2005). Review of methodologies and a protocol for the Agrobacterium-mediated transformation of wheat. *Plant Methods*, DOI: 10.1186/1746-4811-1-5.
- Kammerloher, W., Fisher, U., Piechottka, G. P. & Schäffner, A. R. (1994). Water channels in the plant plasma membrane cloned by immunoselection from a mammalian expression system. *Plant J.*, 6, 187-199.
- Karlsson, M., Johansson, L. & Bush, M. (2000). An abundant TIP expressed in mature highly vacuolated cells. *Plant J.*, 21, 83-90.
- Kawasaki, S., Borchert, C., Deyholos, M., Wang, H., Brazille, S., Kawai, K., Galbraith, D. & Bohnert, H. J. (2001). Gene expression profiles during the initial phase of salt stress in rice. *Plant Cell*, 13, 889-905.
- Kawaura, K., Mochida, K., Yamazaki, Y. & Ogihara, Y. (2006). Transcriptome analysis of salinity stress responses in common wheat using a 22k oligo-DNA microarray. *Funct. Integr. Genomics*, 6 (2), 132-142.
- Kido, E. A., Ferreira Neto, J. R. C., Silva, R. L. O., Belarmino, L. C., Bezerra Neto, J. P., Soares-Cavalcanti, N. M., Pandolfi, V., Silva, M. D., Nepomuceno, A. L. & Benko-Iseppon, A. M. B. (2013). Expression dynamics and genome distribution of osmoprotectants in soybean: identifying important components to face abiotic stress. *BMC Bioinformatics*, 14.
- Kishor, P. B. K., Hong, Z., Miao, G. H., Hu, C. A. A. & Verma, D. P. S. (1995). Overexpression of pyrroline-5-carboxylate synthetase increases proline production and confers osmotolerance in transgenic plants. *Plant Physiol.*, 108, 1387-1394.
- Knight, H., Trewavas, A. J. & Knight, M. R. (1997). Calcium signalling in *Arabidopsis thaliana* responding to drought and salinity. *Plant J.*, 12 (5), 1067-1078.
- Kong, J., Gong, J. M., Zhang, J. G., Zhang, J. S. & Chen, S. Y. (2003). A new AOX homologous gene *OsIM1* from rice (*Oryza sativa* L.) with an alternative splicing mechanism under salt stress. *Theor. Appl. Genet.*, 107, 326-331.
- Kovacs, D., Agoston, B. & Tompa, P. (2008). Disordered plant LEA proteins as molecular chaperones. *Plant Signal. Behav.*, 3, 710-713.
- Krasensky, J. & Jonak, C. (2012). Drought, salt, and temperature stress-induced metabolic rearrangements and regulatory networks. *J. Exp. Bot.*, DOI:10.1093/jxb/err460.
- Kulcheski, F. R., Oliveira, L. F. V., Molina, L. G., Almerão, M. P., Rodrigues, F. A., Marcolino, J., Barbosa, J. F., Stolf-Moreira, R., Nepomuceno, A. L., Marcelino-Guimarães, F. C., Abdelnoor, R. V., Nascimento, L.C., Carazzole, M. F., Pereira, G. A. G. & Margis, R. (2011). Identification of novel soybean microRNAs involved in abiotic and biotic stresses. *BMC Genomics*, 12, 307.
- Kushwaha, R., Downie, A. B. & Payne, C. M. (2013). Uses of phage display in agriculture: Sequence analysis and comparative modeling of late embryogenesis abundant client proteins suggest protein-nucleic acid binding functionality. *Comput. Math. Methods Med.*, 1-11.
- Le, D. T., Nishiyama, R., Watanabe, Y., Tanaka, M., Seki, M., Ham, L. H., Yamaguchi-Shinozaki, K., Shinozaki, K. & Tran, L. S. P. (2012). Differential gene expression in soybean leaf tissues at late developmental stages under drought stress revealed by genome-wide transcriptome analysis. *PLoS ONE*, 7 (11).

- Legay, S., Lamoureux, D., Hausman, J. F., Hoffmann, L. & Evers, D. (2009). Monitoring gene expression of potato under salinity using cDNA microarrays. *Plant Cell. Rep.*, 28 (12), 1799-1816.
- Li, C., Liang, D. D., Li, J., Duan, Y. B., Li, H., Yang, Y. C., Qin, R. Y., Li, L., Wei, P. C. & Yang, J. B. (2013). Unravelling mitochondrial retrograde regulation in the abiotic stress induction of rice Alternative Oxidase 1 genes. *Plant Cell. Environ.*, 36, 775–788.
- Li, D., Xia, X., Yin, W. & Zhang, H. (2013). Two poplar calcineurin B-like proteins confer enhanced tolerance to abiotic stresses in transgenic *Arabidopsis thaliana*. *Biol. Plantarum*, 57 (1), 70-78.
- Li, H., Dong, Y., Yin, H., Wang, N., Yang, J., Liu, X., Wang, Y., Wu, J. & Li, X. (2011). Characterization of the stress associated microRNAs in Glycine max by deep sequencing. *BMC Plant Biology*, 1, 170.
- Li, W., Zang, B., Deng, X. & Wang, X. Overexpression of the trehalose-6-phosphate synthase gene *OsTPS1* enhances abiotic stress tolerance in rice. (2011). *Planta*, 234, 1007-1018.
- Li, Z. Y. & Chen, S. Y. (2000). Differential accumulation of the S-adenosylmethionine decarboxylase transcript in rice seedlings in response to salt and drought stresses. *Theor. Appl. Genet.*, 100, 782-788.
- Liang, P. & Pardee, A. B. (1992). Differential display of eucaryotic messenger RNA by means of polymerase chain reaction. *Science*, 257, 967–971.
- Liu, C., Mao, B., Ou, S., Wang, W., Liu, L., Wu, Y., Chu, C. & Wang, X. (2014). OsbZIP71, a bZIP transcription factor, confers salinity and drought tolerance in rice. *Plant Mol. Biol.*, 84, 19-36.
- Liu, Q., Kasuga, M., Sakuma, Y., Abe, H., Miura, S., Yamaguchi-Shinozaki, K. & Shinozaki, K. (1998). The transcription factors, DREB1 and DREB2, with an EREBP/AP2 DNA binding domain separate two cellular signal transduction pathways in drought- and low-temperature-responsive gene expression, respectively, in *Arabidopsis*. *Plant Cell.*, 10, 1391–1406.
- Liu, X. & Baird, W. M. (2003). Differential expression of genes regulated in response to drought or salinity stress in sunflower, *Crop Scienc.*, 43, 678-687.
- Liu, Y., Ji, X., Zheng, L., Xianguang, N. & Wang, Y. (2013). Microarray analysis of transcriptional responses to abscisic acid and salt stress in *Arabidopsis thaliana*. *Int. J. Mol. Sci.*, 14, 9979-9998.
- Lockhart, D. J., & Winzeler, E. A. Genomics, gene expression and DNA arrays. (2000). *Nature*, 405, 827-836.
- Luan, S., Kudla, J., Rodriguez-Concepcion, M., Yalovsky, S. & Gruissem, W. (2002). Calmodulins and calcineurin B-like proteins: calcium sensors for specific signal response coupling in plants. *Plant Cell*, 14, 389-400.
- Lukyanov, S. A., Rebrikov, D. & Buzdin, A. A. (2007). Supression subtractive hybridization. In: Buzdin, A., Lukyanov, S. (eds.). *Nucleic acids hybridization*. Heidelberg, Germany: Springer, 53-84.
- Lutz, W., Warren Sanderson, W. & Scherbov, S. (2001). The end of world population growth. *Nature*, 412, 543-545.
- Mahajan, S., Sopory, S.K. & Tuteja, N. (2006). CBL-CIPK paradigm: role in calcium and stress signaling in plants. *Trends Plant Sci.*, 14, 37-42.

- Mahdieh, M., Mostajeran, A., Horie, T. & Katsuhara, M. (2008). Drought stress alters water relations and expression of *PIP*-Type aquaporin genes in *Nicotiana tabacum* plants. *Plant Cell Physiol.*, 49, 801-13.
- Malone, J. H. & Oliver, B. (2011). Microarrays, deep sequencing and the true measure the transcriptome. *BMC Biology*, 9, 34.
- Manfre, A., Lanni, L. & Marcotte, W. (2006). The *Arabidopsis* Group 1 Late-Embryogenesis Abundant Protein ATEM6 is required for normal seed development. *Plant Physiol.*, 140, 140-149.
- Mantri, N. L., Ford, R., Coram, T. E. & Pang, E. C. K. (2007). Transcriptional profiling of chickpea genes differentially regulated in response to high-salinity, cold and drought. *BMC Genomics*, 8, 303.
- Mardis, E. R. (2008). Next-generation DNA sequencing methods. *Annu. Rev. Genomics Hum. Genet.*, 9, 3087-402.
- Martins, P.K., Jordão, B.Q., Yamanaka, N., Farias, J. R. B., Beneventi, M. A., Binneck, E., Fuganti, R., Stolf, R. & Nepomuceno, A. L. (2008). Differential gene expression and mitotic cell analysis of the drought tolerant soybean (*Glycine max* L. Merrill Fabales, Fabaceae) cultivar MG/BR46 (Conquista) under two water deficit induction systems. *Genet. Mol. Biol.*, 31(2), 512-521.
- Martre, P., Morillon, R., Barrieu, F., North, G. B., Nobel, P. S. & Chrispeels, M. J. (2002). Plasma membrane aquaporins play a significant role during recovery from water deficit. *Plant Physiol.*, 130, 2101-2110.
- Matsumura, H., Ito, A., Saitoh, H., Winter, P., Kahl, G., Reuter, M., Kruger, D. H. & Terauchi, R. (2005). SuperSAGE. *Cell. Microbiol.*, 7 (1), 11–18.
- Montalvo-Hernández, L., Piedra-Ibarra, E., Gómez-Silva, L., Lira-Carmona, R., Acosta-Gallegos, J. A. A., Vazquez-Medrano, J., Xoconostle-Cázares, B. & Ruíz-Medrano, R. (2008). Differential accumulation of mRNAs in drought-tolerant and susceptible common bean cultivars in response to water deficit. *New Phytol.*, 177, 102-113.
- Morgan, J. M. (1984). Osmoregulation and water stress in higher plants. *Ann. Rev. Plant Physiol.*, 35, 299-319.
- Mueller-Roeber, B. & Pical, C. (2002). Inositol phospholipid metabolism in *Arabidopsis*. Characterized and putative isoforms of inositol phospholipids kinase and phosphoinositide-specific phospholipase C. *Plant Physiol.*, 130, 22-46.
- Murphy, D. (2002). Gene expression studies using microarrays: principles, problems, and prospects. *Advan. Physiol. Edu.*, 26, 256-270.
- Nakashima, K., Ito, Y. & Shinozaki, K. Y. (2009). Transcriptional regulatory networks in response to abiotic stresses in *Arabidopsis* and *Grasses*. *Plant Physiol.*, 149, 88–95.
- Nieva, C., Busk, P., Domínguez-Puigjaner, E., Lumbreras, V., Testillano, P., Risueno, M. C. & Pagés, M. (2005). Isolation and functional characterisation of two new bZIP maize regulators of the ABA responsive gene *rab28*. *Plant Mol. Biol.*, 58, 899-914.
- North, G. B., Martre, P. & Nobel, P. S. (2004). Aquaporins account for variations in hydraulic conductance for metabolically active root regions of *Agave deserti* in wet, dry and rewetted soil. *Plant Cell Environ.*, 27, 219–228.
- Olvera-Carrillo, Y., Reyes, J. L. & Covarrubias, A. A. (2011). Late embryogenesis abundant proteins: Versatile players in the plant adaptation to water limiting environments. *Plant Signal. Behav.*, 6, 586-589.

- Ouyang, B., Yang, T., Li, H., Zhang, L., Zhang, Y., Zhang, J., Fei, Z. & Ye, Z. (2007). Identification of early salt stress response genes in tomato root by suppression subtractive hybridization and microarray analysis. *J. Exp. Bot.*, 58, 507–520.
- Paiva, C. L. A. & Panek, A. D. (1996). Biotechnological applications of the disaccharide trehalose. *Biotechnol. Annu. Rev.*, 2, 293–314.
- Penna, S. (2003). Building stress tolerance through over-producing trehalose in transgenic plants. *Trends Plant Sci.*, 8, 355–7.
- Porcel, R., Aroca, R., Azcón, R. & Ruiz-Lozano, J. M. (2006). PIP aquaporin gene expression in arbuscular mycorrhizal *Glycine max* and *Lactuca sativa* plants in relation to drought stress tolerance. *Plant Mol. Biol.*, 60, 389–404.
- Porcel, R., Azcón, R. & Ruiz-Lozano, J. M. (2004). Evaluation of the role of genes encoding for Δ^1 -pyrroline-5-carboxylate synthetase (P5CS) during drought stress in arbuscular mycorrhizal *Glycine max* and *Lactuca sativa* plants. *Physiol. Mol. Plant Pathol.*, 65, 211–221.
- Quigley, F., Rosenberg, J.M., Shachar-Hill, Y. & Bohnert, H. J. (2001). From genome to function: the *Arabidopsis* aquaporins. *Genome Biol.*, 3,1-17.
- Rabbani, M. A., Maruyama, K., Abe, H., Khan, M. A., Katsura, K., Ito, Y., Yoshiwara K., Seki, M., Shinozaki, K. & Yamaguchi-Shinozaki, K. (2003). Monitoring expression profiles of rice genes under cold, drought, and high-salinity stresses and abscisic acid application using cDNA microarray and RNA gel-blot analyses. *Plant Physiol.*, 133, 1755–1767.
- Recchia, G. H., Caldas, D. G. G., Beraldo, A. L. A., Silva, M. J. & Tsai, S. M. (2013). Transcriptional analysis of drought-induced genes in the roots of a tolerant genotype of the common bean (*Phaseolus vulgaris* L.). *Int. J. Mol. Sci.*, 14, 7155–7179.
- Reddy, P. S., Reddy, G. M., Pandey, P., Chandrasekhar, K. & Reddy, M. K. (2012). Cloning and molecular characterization of a gene encoding late embryogenesis abundant protein from *Pennisetum glaucum*: protection against abiotic stresses. *Mol. Biol. Rep.*, 39, 7163–7174.
- Reis, S. P., Tavares, L. de S. C., Costa, C. de N. M., Santa Brígida, A. B. & de Souza, C. R. B. (2012). Molecular cloning and characterization of a novel RING zinc-finger protein gene up-regulated under in vitro salt stress in cassava. *Mol. Biol. Rep.*, 39, 6513–6519.
- Rhee, S.Y., Dickerson, J. & Xu, D. (2006). Bioinformatics and its applications in plant biology. *Annu. Ver. Plant. Biol.*, 57, 355–360.
- Riechmann, J. L., Heard, J., Martin, G., Reuber, L., Jiang, C., Keddie, J., Adam, L., Pineda, O., Ratcliffe, O. J., Samaha, R. R., Creelman, R., Pilgrim, M., Broun, P., Zhang, J. Z., Ghandehari, D., Sherman, B, K. & Yu, G. (2000). *Arabidopsis* transcription factors: genome-wide comparative analysis among eukaryotes. *Science*, 290, 2105–2110.
- Rodrigues, F. A., Marcolino-Gomes, J., Carvalho, J. F. C., Nascimento, L. C., Neumaier, N., Farias, J. R. B., Carazzolle, M. F., Marcelino, F. C. & Nepomuceno, A. L. (2012). Subtractive libraries for prospecting differentially expressed genes in the soybean under water deficit. *Genet. Mol. Biol.*, 35 (1), 304–314.
- Roser, B. & Colaço, C. (1993). A sweeter way to fresh food. *New Sci.*, 138, 25–28.
- Saibo, N. J. M., Lourenço, T. & Oliveira, M. M. (2009). Transcription factors and regulation of photosynthetic and related metabolism under environmental stresses. *Ann. Bot.*, 103, 609–623.
- Sambrook, J., Russel, D.W. *Molecular Cloning: A laboratory manual*. 3^a ed. New York, USA: Cold Spring Harbor Laboratory Press; 2001.

- Samuel, D., Kumar, T. K., Ganesh, G., Jayaraman, G., Yang, P. W., Chang, M. M., Trivedi, V. D., Wang, S. L., Hwang, K. C., Chang, D. K. & Yu, C. (2000). Proline inhibits aggregation during proteins refolding. *Protein Sci.*, 9, 344-352.
- Sanders, D., Pelloux, J., Brownlee, C. & Harper, J. F. (2002). Calcium at the crossroads of signaling. *Plant Cell*, 14, 401-417.
- Sanford, J. C. (1990). Biolistic plant transformation. *Physiol. Planta.*, 79 (1), 206-209.
- Sanford, J. C., Klein, T. M., Wolf, E. D. & Allen, N. (1987). Delivery of substances into cells and tissues using a particle bombardment process. *Particul. Sci. Technol.*, 5 (1). DOI: 10.1080/02726358708904533.
- Santa Brígida, A. B., dos Reis, S. P., Costa, C. de N., Cardoso, C. M., Lima, A. M. & de Souza, C.R.B. (2014). Molecular cloning and characterization of a cassava translationally controlled tumor protein gene potentially related to salt stress response. *Mol. Biol. Rep.*, 41, 1787-1797.
- Seki, M., Narusaka, M., Abe, H., Kasuga, M., Yamaguchi-Shinozaki, K., Carninci, P., Hayashizaki, Y. & Shinozaki, K. (2001). Monitoring the expression pattern of 1300 *Arabidopsis* genes under drought and cold stresses by using a full-length cDNA microarray. *Plant Cell*, 13, 61-72.
- Seki, M., Narusaka, M., Ishida, J., Nanjo, T., Fujita, M., Oono, Y., Kamiya, A., Nakajima, M., Enju, A., Sakurai, T., Satou, M., Akiyama, K., Taji, T., Yamaguchi-Shinozaki, K., Carninci, P., Kawai, J., Hayashizaki, Y. & Shinozaki, K. (2002). Monitoring the expression profiles of 7000 *Arabidopsis* genes under drought, cold and high-salinity stresses using a full-length cDNA microarray. *Plant J.*, 31, 279-292.
- Serrano, R., Cullianz-Macia, F. A. & Moreno, V. (1999). Genetic engineering of salt and drought tolerance with yeast regulatory genes. *Sci. Hort.*, 78, 261-269.
- Shendure, J. & Ji, H. (2008). Next-generation DNA sequencing. *Nat. Biotechnol.*, 1135-1145.
- Shinozaki, K. Y. & Shinozaki, K. (2006). Transcriptional regulatory networks in cellular responses and tolerance to dehydration and cold stresses. *Annu. Rev. Plant Biol.*, 57 (7), 81-803.
- Shou, H., Bordallo, P. & Wang, K. (2004a). Expression of the *Nicotiana* protein kinase (*NPK1*) enhanced drought tolerance in transgenic maize. *J. Exp. Bot.*, 55 (399), 1013-1019.
- Shou, H., Frame, B. R., Whitham, S. A. & Wang, K. (2004 b). Assessment of transgenic maize events produced by particle bombardment or *Agrobacterium*-mediated transformation. *Molecular Breeding*, 13, 201-208.
- Silva-Ortega, C. O., Ochoa-Alfaro, A. E., Reyes-Agüero, J. A., Aguado-Santacruz, G. A. & Jiménez-Bremont, J. F. (2008). Salt stress increases the expression of *p5cs* gene and induces proline accumulation in cactus pear. *Plant Physiol. Biochem.*, 46, 82-92.
- Singh, K. B., Foley, R. C. & Oñate-Sánchez, L. (2002). Transcription factors in plant defense and stress responses. *Curr. Opin. Plant Biol.*, 5, 430-436.
- Smith, C. A., Melino, V. J., Sweetman, C. & Soole, K. L. (2009). Manipulation of alternative oxidase can influence salt tolerance in *Arabidopsis thaliana*. *Physiol. Plant*, 137 (4), 459-472.
- Solomon, S., Qin, D., Manning, M., Chen, Z., Marquis, M., Averyt, K. B., Tignor, M. & Miller, H. L. (2007). *Climate change 2007: The physical science basis*. Contribution of working group I to the fourth assessment report of the intergovernmental panel on climate change. Geneva, Switzerland: IPCC Secretariat.

- Stiller, I., Dulai, S., Kondrák, M., Tarnai, R., Szabó, L., Toldi, O. & Bánfalvi, Z. (2008). Effects of drought on water content and photosynthetic parameters in potato plants expressing the trehalose-6-phosphate synthase gene of *Saccharomyces cerevisiae*. *Planta*, 227, 299–308.
- Ström, A. R. & Kaasen, I. (1993). Trehalose metabolism in *Escherichia coli*: Stress protection and stress regulation of gene expression. *Mol. Microbiol.*, 8, 205–210.
- Su, J. & Wu, R. (2004). Stress-inducible synthesis of proline in transgenic rice confers faster growth under stress conditions than that with constitutive synthesis. *Plant Sci.*, 166, 941–948.
- Suárez, R., Wong, A., Ramírez, M., Barraza, A., Orozco, M., del, C., Cevallos, M. A., Lara, M., Hernández, G. & Iturriaga, G. (2008). Improvement of drought tolerance and grain yield in common bean by overexpressing trehalose-6-phosphate synthase in rhizobia. *Mol. Plant Microbe Interact.*, 21, 958–66.
- Székel, G., Ábrahám, E., Cséplő, A., Rigó, G., Zsigmond, L., Csiszár, J., Ayaydin, F., Strizhov, N., Jásik, J., Schmelzer, E., Koncz, C. & Szabados, L. (2008). Duplicated *P5CS* genes of *Arabidopsis* play distinct roles in stress regulation and developmental control of proline biosynthesis. *Plant J.*, 53, 11–28.
- Tang, H., Sezen, U. & Paterson, A. H. (2010). Domestication and plant genomes. *Curr. Opin. Plant Biol.*, 13, 160–166.
- Tester, M. & Langridge, P. (2010). Breeding technologies to increase crop production in a changing world. *Science*, 327, 818–822.
- Tilman, D., Cassman, K. G., Matson, P. A., Naylor, R. & Polasky, S. (2002). Agricultural sustainability and intensive production practices. *Nature*, 418, 671–677.
- Tolletier, D., Jaquinod, M., Mangavel, C., Passirani, C., Saulnier, P., Manon, S., Teyssier, E., Payet, N., Avelange-Macharel, M-H. & Macharel, D. (2007). Structure and function of a mitochondrial late embryogenesis abundant protein revealed by desiccation. *Plant Cell*, 19, 1580–1589.
- Torres, G. A. M., Pflieger, S., Corre-Menguy, F., Mazubert, C., Hartmann, C. & Lelandais-Brière, C. (2006). Identification of novel drought-related mRNAs in common bean roots by differential display RT-PCR. *Plant Sci.* 171, 300–307.
- Travella, S., Ross, S., Harden, J., Everett, C., Snape, J. & Harwood, W. (2005). A comparison of transgenic barley lines produced by particle bombardment and *Agrobacterium*-mediated techniques. *Plant Cell Rep.* 23, 780–789.
- Tunnacliffe, A. & Wise, M. (2007). The continuing conundrum of the LEA proteins. *Naturwissenschaften*, 94, 791–812.
- Tyerman, S. D., Bohnert, H. J., Maurel, C., Steudle, E. & Smith, J. A. C. (1999). Plant aquaporins: their molecular biology, biophysics and significance for plant water relations. *J. Exp. Bot.*, 50, 1055–1071.
- Tyerman, S. D., Niemietz, C. M. & Bramley, H. (2002). Plant aquaporins: multifunctional water and solute channels with expanding roles. *Plant Cell. Environ.*, 25, 173–194.
- Umezawa, T., Mizuno, K. & Fujimura, T. (2002). Discrimination of genes expressed in response to the ionic or osmotic effect of salt stress in soybean with cDNA-AFLP. *Plant Cell. Environ.*, 25, 1617–1625.
- Uno, Y., Furihata, T., Abe, H., Yoshida, R., Shinozaki, K. & Yamaguchi-Shinozaki, K. (2000). *Arabidopsis* basic leucine zipper transcription factors involved in an abscisic acid-dependent signal transduction pathway under drought and high-salinity conditions. *Proc. Natl. Acad. Sci. USA*, 97, 11632–11637.

- Utsumi, Y., Tanaka, M., Morosawa, T., Kurotani, A., Yoshida, T., Mochida, K., Matsui, A., Umemura, Y., Ishitani, M., Shinozaki, K., Sakurai, T. & Seki, M. (2012). Transcriptome analysis using a highdensity oligomicroarray under drought stress in various genotypes of cassava: an important tropical crop. *DNA Res.*, 19, 335–451.
- Varshney, R. K., Hiremath, P. J., Lekha, P., Kashiwagi, J., Balaji, J., Deokar, A. A., Vadez, V., Xiao, Y., Srinivasan, R., Gaur, P. M., Siddique, K. H. M., Town, C. D. & Hoisington, D. A. (2009). A comprehensive resource of drought- and salinity responsive ESTs for gene discovery and marker development in chickpea (*Cicer arietinum* L.). *BMC Genomics*, 10, 523.
- Velculescu, V. E., Zhang, L., Vogelstein, B. & Kinzler, K. W. (1995). Serial analysis of gene expression. *Science*, 270, 484–487.
- Vendruscolo, E. C. G., Schusterb, I., Pileggic, M., Scapimd, C. A., Molinarie, H. B. C., Marure, C. J. & Vieira, L. G. E. (2007). Stress-induced synthesis of proline confers tolerance to water deficit in transgenic wheat. *J. Plant Physiol.*, 164, 1367–1376.
- Vidal, R. O., Nascimento, L. C., Mondego, J. M. C., Pereira, G. A. G. & Carazzolle, M. F. (2012). Identification of SNPs in RNA-Seq data of two cultivars of *Glycine max* (soybean) differing in drought resistance. *Genet. Mol. Biol.*, 35 (1), 331–334.
- Wang, D., Pan, Y., Zhao, X., Zhu, L., Fu, B. & Li, Z. (2011). Genome-wide temporal-spatial gene expression profiling of drought responsiveness in rice. *BMC Genomics*, 12, 149.
- Wang, H., Miyazaki, S., Kawai, K., Deyholos, M., Galbraith, D. & Bohnert, H. (2003). Temporal progression of gene expression responses to salt shock in maize roots. *Plant Mol. Biol.*, 52, 873–891.
- Wang, M., Li, P., Li, C., Pan, Y., Jiang, X., Zhu, D., Zhao, Q. & Yu, J. (2014). SiLEA14, a novel atypical LEA protein, confers abiotic stress resistance in foxtail millet. *BMC Plant Biol.*, 14, 290. DOI: 10.1186/s12870-014-0290-7.
- Wang, W., Vinocur, B. & Altman, A. (2003). Plant responses to drought, salinity and extreme temperatures: towards genetic engineering for stress tolerance. *Planta*, 218, 1–14.
- Wang, Y., Tao, X., Tang, X., Xiao, L., Sun, J., Yan, X., Li, D., Deng, H. & Ma, X. (2013). Comparative transcriptome analysis of tomato (*Solanum lycopersicum*) in response to exogenous abscisic acid. *BMC Genomics*, 14, 841.
- Wang, Z., Gerstein, M. & Snyder, M. (2009). RNA-Seq: a revolutionary tool for transcriptomics. *Nature Rev. Genet.*, 10, 57–63.
- Wani, S. H., Haider, N., Kumar, H. & Singh, N. B. (2010). Plant plastid engineering. *Current Genomics*, 11, 500–512.
- Watson, J. M., Fusario, A. F., Wang, M. & Waterhouse, P. M. (2005). RNA silencing platforms in plants. *FEBS Lett.*, 578, 5982–5987.
- Way, H., Chapman, S., McIntyre, L., Casu, R., Xue, G. P., Manners, J. & Shorter, R. (2005). Identification of differentially expressed genes in wheat undergoing gradual water deficit stress using a subtractive hybridisation approach. *Plant Sci.*, 168 (3), 661–670.
- Weaver, C. D., Crombie, B. & Stacey, G. (1991). Calcium-dependent phosphorylation of symbiosome membrane proteins from nitrogen-fixing soybean nodules. *Plant Physiol.*, 95, 222–227.
- Wise, M. (2003). LEAping to conclusions: A computational reanalysis of late embryogenesis abundant proteins and their possible roles. *BMC Bioinformatics*, 4, 52.

- Wolkers, W., McCready, S., Brandt, W., Lindsey, G. & Hoekstra, F. (2001). Isolation and characterization of a D-7 LEA protein from pollen that stabilizes glasses in vitro. *Biochim. Biophys. Acta*, 1544, 196-206.
- Wu, F., Wang, L., Li, L. & Wang, S. (2014). De novo assembly of the common bean transcriptome using shor reads for the discovery of drought-responsive genes. *PLoS ONE*, 9 (10).
- Wu, Y., Wang, Q., Ma, Y. & Chu, C. (2005). Isolation and expression analysis of salt up-regulated ESTs in upland rice using PCR-based subtractive suppression hybridization method. *Plant Sci.*, 168, 847-853.
- Wyrick, J. J. & Young, R. A. (2002) Deciphering gene expression regulatory Networks. *Curr. Opin. Genet. Dev.*, 12, 130-6.
- Xiang, Y., Tang, N., Du, H., Ye, H. & Xiong, L. (2008). Characterization of *OsbZIP23* as a key player of the basic leucine zipper transcription factor family for conferring abscisic acid sensitivity and salinity and drought tolerance in rice. *Plant Physiol.*, 148, 1938–1952.
- Xiao, B., Huang, Y., Tang, N. & Xiong, L. (2007). Over-expression of a LEA gene in rice improves drought resistance under the field conditions. *Theor. Appl. Genet.*, 115, 35–46.
- Xu, D., Duan, B., Wang, B., Hong, T. H. D. & Wu, R. (1996). Expression of a late embryogenesis abundant protein gene, *HVA1*, from barley confers tolerance to water deficit and salt stress in transgenic rice. *Plant Physiol.*, 110 (1), 249-257.
- Xu, J., Yibing, Y., Yunbi, X., Gengyun, Z., Xiaosen, G., Fenkai, W., Wang, Q., Rong, T., Pan, G., Cao, M., Tang, Q., Gao, S., Liu, Y., Wang, J., Lan, H. & Lu, Y. (2014). Identification of candidate genes for drought tolerance by whole-genome resequencing in maize. *BMC Plant Biology*, 14, 83.
- Yamazaki, M. & Saito, K. (2002). Differential display analysis of gene expression in plants. *Cell Mol. Life Sci.*, 59 (8), 1246-1255.
- Yang, L., Zheng, B., Mao, C., Yi, K., Liu, F., Wu, Y. & Tato, Q. (2003). cDNA-AFLP analysis of inducible gene expression in rice seminal root tips under a water deficit. *Gene*, 314, 141-148.
- Yoshiba, Y., Kiyosue, T., Katagiri, T., Ueda, H., Mizoguchi, T., Yamaguchi-Shinozaki, K., Wada, K., Harada, Y. & Shinozaki, K. (1995). Correlation between the induction of a gene for delta 1-pyrroline-5-carboxylate synthetase and the accumulation of proline in *Arabidopsis thaliana* under osmotic stress. *Plant J.*, 7,751-760.
- Zahur, M., Asif, M. A., Zeeshan, N., Mehmood, S., Malik, M. F. & Asif, A. R. (2013). Homeobox leucine zipper proteins and cotton improvement. *Adv. Biosci. Biotechnol.*, 4, 15-20.
- Zalewski, W., Orczyk, W., Gasparis & Nadolska-Orczyk, S. (2012). *HvCKX2* gene silencing by biolistic or *Agrobacterium*-mediated transformation in barley leads to different phenotypes. *BMC Plant Biology*, 12, 206. DOI: 10.1186/1471-2229-12-206.
- Zhang, J. S., Duncan, E. L., Chang, A. C. & Reddel, R. R. (1998). Differential display of mRNA. *Mol. Biotech.*, 10, 155-165.
- Zhou, J., Wang, X., Jiao, Y., Qin, Y., Liu, X., He, K., Chen, C., Ma, L., Wang, J., Xiong, L., Zhang, Q., Fan, L. & Deng, X. W. (2007). Global genome expression analysis of rice in response to drought and high-salinity stresses in shoot, flag leaf and panicle. *Plant Mol. Biol.*, 63, 591–608.

- Zhou, S., Hu, W., Deng, X., Ma, Z., Chen, L., Huang, C., Wang, C., Wang, J., He, Y., Yang, G. & He, G. (2012). Overexpression of the wheat aquaporin gene, *TaAQP7*, enhances drought tolerance in transgenic tobacco. *PLoS ONE*, 7 (12).
- Zhu, C., Schraut, D., Hartung, W. & Schaffner, A. R. J. (2005). Differential responses of maize MIP genes to salt stress and ABA. *J. Exp. Bot.*, 56, 2971-2981.

Chapter 45

CLINICAL EVIDENCE AND THE GENETIC EFFECTS OF TRADITIONAL CHINESE MEDICINE FOR THE MANAGEMENT OF PROTEINURIA IN PATIENTS WITH DIABETIC NEPHROPATHY

***Xiang Tu¹, XueFeng Ye², Quan Hu³, ChunGuang Xie⁴,
ChengShi He¹, Ming Chen², James B. Jordan¹ and Sen Zhong^{1,*}***

¹National Traditional Chinese Medicine Clinical Research Base for Diabetes Mellitus/Teaching Hospital of Chengdu University of Traditional Chinese Medicine, Chengdu, Sichuan Province, China

²Department of Nephrology, Teaching Hospital of Chengdu University of Traditional Chinese Medicine, Chengdu, Sichuan Province, China

³Department of Gerontology, Teaching Hospital of Chengdu University of Traditional Chinese Medicine, Chengdu, Sichuan Province, China

⁴Department of Endocrinology, Teaching Hospital of Chengdu University of Traditional Chinese Medicine, Chengdu, Sichuan Province, China

ABSTRACT

Proteinuria is the hallmark of diabetic nephropathy (DN) and vastly increases the incidence of cardio-vascular disease and mortality. Traditional Chinese Medicine (TCM) has been used for diabetes and its complications for thousands of years and appears to be promising in the treatment of proteinuria in DN patients. Clinical trials evaluating TCM for proteinuria, either used as a monotherapy or in combination with western medicine, has produced positive results. Although a large number of clinical studies have been conducted, the clinical evidence with regard to TCM for proteinuria in patients with DN remains inconclusive. The recent progression of evidence will be introduced in this chapter. Current basic research has disclosed that TCM might affect a variety of genes regarding DN

* Corresponding Author's Email: zhongsen6606@163.com (Professor Sen Zhong, Teaching Hospital of Chengdu University of Traditional Chinese Medicine, Chengdu 610072, Sichuan Province, China. Tel: 86-18980086606).

etiology. This chapter will analyze recent achievements in this area and address the issue of the association between clinical evidence and the genetic effects of TCM.

INTRODUCTION

Diabetic nephropathy (DN) is one of the micro-vascular complications of diabetes mellitus. It is characterized by proteinuria and glomerulosclerosis [1]. Proteinuria not only marks the occurrence of DN, but also vastly increases the incidence of cardio-vascular disease and mortality. Many factors contribute to the etiology of DN, such as glomerular hyperfiltration, involvement of cytokines, increase of glycosylated end products, activation of sorbitol and protein kinase C pathways, and susceptibility of genes, etc. In the past few decades, Chinese medical experts have disclosed a number of genetic variants which may be associated with diabetic nephropathy in the Chinese population. Some of the recent findings are shown in Table 1.

Table 1. Genetic variants which have been investigated in Chinese ethnic populations with diabetic nephropathy

Genetic variants	Results associated with DN	Studies	Cases/ Controls (N)	Ethnic population	Region
REN (ACAG)n-STRP	No	Zheng1997[2]	85/42	Unclear	China
ACE I/D	Yes	Zhang1999[3]	92/57	Han	Hebei Province, China
RAGE Gly82Ser	No	Liu2000[4]	91/65	Han	Shanghai, China
PAI-1 4G/4G	Yes	Wong2000[5]	95/46	Unclear	China
RAS ACE and AGT-M235T	Yes	Wu2000[6]	71/41	Han	Shanghai, China
ACE I/D	Yes	Qu2001[7]	957/1061	Han	China
HSPG	No	Yang2001[8]	136/190	Han	Jiangsu Province, China
ACE	Yes	Chen2002[9]	155/85	Unclear	China
ACE I/D	Yes	Fu2002[10]	44/47	Hmong	Western Hunan Province, China
ACE I/D	Yes	Liao2002[11]	53/67	Han	Guangxi Province, China
5' -ALR2 Dinucleotide repeat	Yes	Liu2002[12]	213/77	Unclear	Southern China
Apo E	No	Shen2002[13]	159/106	Han	Shanghai, China
ANP C/T	No	Xu2002[14]	83/50	Han	Guangzhou, Guangdong Province, China
HSPG BamHI	Yes	Liu2003[15]	290/77	Unclear	China
HSPG T and ApoE E2	Yes	Liu2003[16]	218/80	Unclear	China
MTHFR C677T	Yes	Xu2003[17]	69/54	Han	Hebei Province,

Genetic variants	Results associated with DN	Studies	Cases/ Controls (N)	Ethnic population	Region
					China
CA(n) z-2 allele and ALR2 T allele	Yes	Wang2003[18]	159/392	Unclear	Hong Kong, China
TGF-beta T869C	Yes	Wong2003[19]	58/65	Unclear	China
MTHFR C6677T	Yes	Chen2004[20]	41/50	Han	Gansu Province, China
PAI-1 4G/5G	Yes	Liu2004[21]	77/70	Han	Guangdong Province, China
ArE/PON1 Q192R	No	Qian2004[22]	123/121	Han	Zhengzhou, Henan Province, China
HNF-1 β exon2	No	Su2004[23]	56/62	Han	Henan Province, China
eNOS 4A/B	Yes	Sun2004[24]	188/114	Han	Anhui Province, China
HLA-DQA1*0302	Yes	Yang2004[25]	56/52	Han	Yunnan Province, China
HLT+	Yes	Baum2005[26]	374/392	Unclear	Hong Kong, China
CCR5 promoter region 59029G/A	Yes	Li2005[27]	64/53	Han	Kunming, Yunnan Province, China
Vitamin D receptor	Yes	Li2005[28]	39/55	Han	China
MnSOD V16A	Yes	Yang2005[29]	137/50	Han	Kunming, Yunnan Province, China
APOE epsilon3/epsilon3 and APOC3 CC	Yes	Ng2006[30]	374/392	Unclear	Hong Kong, China
ACE I/D	No	Xie2006[31]	53/47	Han	Baotou, Neimenggu Province, China
RANTES promoter- 28C/G	No	Zhang2006[32]	90/53	Han	Guizhou Province, China
eEOS intron 4	No	Dong2007[33]	130/107	Unclear	Shandong Province, China/ Singapore
AT ₁ R A1166C	No	Luo2007[34]	110/74	Han	Kunming, Yunnan Province, China
Apo E e2 allele	Yes	Zhang2007[35]	40/38	Unclear	China
AGT M235 and AGT T174M	No	Zhong2007[36]	53/54	Han/ Zhuang	Guangxi Province, China
Apo E	Yes	Li2008[37]	788/516	Han	China
ACE I/D	Yes	Wen2008[38]	110/74	Han	Kunming, Yunnan Province, China
Apo E	Yes	Xian2008[39]	26/30	Uygur/ Han	Hetian, Xinjiang Uygur Autonomous Region, China
MnSOD-V16A	Yes	Li2009[40]	154/103	Unclear	China
MCP-I promoter region A~2518G	Yes	Tang2009[41]	99/51	Zhuang	China
AA and GA					
Apo E	Yes	Wan2009[42]	1240/945	Unclear	China
ICAM-1 E allele	Yes	Chen2010 [43]	94/95	Zhuang	Northern Guilin, Guangxi Province, China

Table 1. (Continued)

Genetic variants	Results associated with DN	Studies	Cases/ Controls (N)	Ethnic population	Region
MnSOD	No	Liu2010[44]	90/70	Han	Kunming, Yunnan Province, China
ACACB SNP rs2268388	Yes	Tang2010[45]	295/300	Unclear	China
ACE gene intron 16 I/D	Yes	Wang2010[46]	26/30	Han	Neimenggu Province, China
PAI-1	Yes	Xue2010[47]	70/50	Han	Baotou, Neimenggu Province, China
NADPH oxidase subunit p22phox C242T	No	Jin2011[48]	186/139	Korean	Yanbian, Jilin Province, China
CD14 promoter region C-159T	No	Jin2011[49]	152/139	Korean	Yanbian, Jilin Province, China
A poE	Yes	Wang2011[50]	1528/1202	Unclear	China
ApoC3 promoter region 482T/T	Yes	Chen2012[51]	121/58	Han	Kunming, Yunnan Province, China
ACE I/D	Yes	Jia2012[52]	2295/2197	Han	China
MTHFS rs6495446	No	Li2012[53]	180/178	Unclear	Taiwan, China
ACE I/D	Yes	Liu2012[54]	2884/2722	Han	China
APN SNP+45 and SNP+276	No	Peng2012[55]	42/40	Han	Zhejiang Province, China
ACE I/D	Yes	Peng2012[56]	1536/1670	Han	China
eNOS	Yes	Rao2012[57]	49/39	Han	Wenling, Zhejiang Province, China
TCF7L2, CDKAL1, HHEX, SLC30A8	Yes	Li2013[58]	100/100	Han	Guangxi Province, China
ACE I/D	Yes	Long2013[59]	2189/2160	Han	China
TOX, SMAD3	Yes	Lv2013[60]	615/475	Han	Tianjin, China
ELMO1	Yes	Wu2013[61]	123/77	Unclear	China
AGTR1 A1166C	Yes	Yin2013[62]	152/141	Unclear	China
MBL2 rs1800450 and rs11003125 SNPs	Yes	Zhang2013[63]	415/260	Han	Northern China
MTHFR C6677T	Yes	Liu2014[64]	82/81	Unclear	Gaungxi Province, China
SNPs and haplotypes located at 16q22.1 region	Yes	Liao2014[65]	217/357	Han	China
TLR4	No	Peng2014[66]	622/833	Unclear	China
RAGE 2245G/A	Yes	Yao2014[67]	176/168	Han	Guangdong, Jiangxi, Hunan, Fujian, Guangxi Province, China
PRKCB1 rs3760106	Yes	Zhao2014[68]	174/228	Han	Shanghai, China
Let-7a	Yes	Zhou2014[69]	108/104	Han	Chongqing, China

This chapter primarily focuses on the association between the clinical evidence of Traditional Chinese Medicine (TCM) for the treatment of proteinuria in patients with DN and the genetic effects of TCM on DN. The chapter examines the following research issues: 1) Clinical evidence of TCM for proteinuria in DN patients; 2) The effects of TCM on DN at the genetic level; 3) The association between the clinical evidence and the genetic effects of TCM. This chapter's discussion will analyze the research and answer the following questions:

Clinical evidence of TCM for proteinuria in DN patients

1. What are the results of randomized controlled trials (RCTs) studying the anti-proteinuric effects of TCM in patients with diabetic nephropathy?
2. What do the systematic reviews and meta-analyses report in this area?
3. What are the limitations with respect to this clinical evidence?

The effects of TCM on DN at the genetic level

1. What are the effects of TCM on DN at the genetic level in animal experiments?
2. What are the effects of TCM on DN at the genetic level in humans?

The association between the clinical evidence and the genetic effects of TCM

1. What is the association?
2. How to relate them?

CLINICAL EVIDENCE OF TCM FOR PROTEINURIA IN DN PATIENTS

What Are the Results of Randomized Controlled Trials Studying the Anti-Proteinuric Effects of TCM in Patients with DN?

Conventional western medicine has developed common medications for the treatment of proteinuria, i.e., angiotensin-converting enzyme inhibitor (ACEI) and angiotensin receptor blocker (ARB). Unlike conventional western medicine, the TCM community can use not only ACEIs or ARBs but also TCM for the management of DN in China. The body of research literature for RCTs evaluating TCM for DN is rapidly increasing. Most of these RCTs measured the changes of urinary albumin excretion (UAE) levels and evaluated the anti-proteinuric effects of TCM in DN patients; consequently, this has become a topical area of research [70,71]. A large number of TCM has been assessed in this area and numerous clinical trials have been reported [72]. It is beyond the scope of the present analysis to elaborate on the multitude of individual research. Instead, our method will be to utilize the PubMed international medical research engine to present the RCT citations.

TCM as a Monotherapy

TCM Extract

Tripterygium Glycoside

Studies where TCM was used as a monotherapy usually produced positive results [73]. Tripterygium Glycoside is extracted from the Chinese Medicinal herb, *Radix Tripterygii Wilfordii*. It has immunosuppressive and anti-inflammatory effects [74,75] and some animal experiments demonstrated that it could reduce UAE by protecting podocytes [76,77]. Chinese military medical experts conducted clinical trials to assess its efficacy and safety in the treatment of DN. The Chinese academician, Liu ZH and his team, published an article in 2010, reporting a six month RCT where Tripterygium Glycoside was used as a monotherapy in the treatment of DN [78]. 65 patients whose $\text{UAE} \geq 2.5 \text{ g/24h}$ and $\text{SCr} < 265.2 \text{ mol/L}$ were included and they were randomly assigned into two groups: Tripterygium Glycoside group (34 S) and valsartan group (31 S). Both groups had four patients lost at follow-up and one patient withdrew. The reduction of UAE was EXTREMELY significant in the Tripterygium Glycoside group: compared with the baseline values, there was a 32.9% reduction at month 1; a 38.8% reduction at month 3; and a 34.3% reduction at month 6. However, the reduction in the valsartan group was not significant: there was a 1.05% reduction at month 1; 10.1% at month 3; and -11.7% at month 6. The authors analyzed the lack of efficacy in the valsartan group, stating it might be associated with the dose of valsartan (160 mg/d), a high value of inclusion criteria on proteinuria ($>2.5 \text{ g/24h}$) and the short trial period (six months). However, it is obvious that the efficacy of Tripterygium Glycoside is superior to valsartan. The reported adverse events of Tripterygium Glycoside were as follows: gastrointestinal adverse reaction (1 S, 2.94%); abnormal liver function (3 S, 8.82%); reduction of white blood cell (1 S, 2.94%); hyperkalemia (8 S, 23.53%); $\text{K}^+ > 6.0 \text{ mmol/L}$ (3 S, 8.82%). The reported adverse events of valsartan were: photosensitive dermatitis (1S, 3.22%); hyperkalemia (10 S, 32.2%); $\text{K}^+ > 6.0 \text{ mmol/L}$ (2 S, 6.45%); and serum creatinine doubling (1 S, 3.22%). The three subjects who had abnormal liver function recovered when the dose of Tripterygium Glycoside was reduced to 60 mg/d; they did not withdraw from the trial. The incidence of hyperkalemia was high because many subjects had renal insufficiency, but symptomatic treatment could successfully manage it. The authors concluded that Tripterygium Glycoside is a safe drug whose efficacy is superior to ARB in the treatment of DN.

Formulae

Tangjiang Shenkang (TJSK)

Although clinical trials of poor methodology continue to devitalize legitimate TCM research, RCTs of higher methodology are emerging. Ma et al. (2011) carried out a randomized placebo controlled double blind trial to evaluate the efficacy and safety of Tangjiang Shenkang (TJSK) granule in the treatment of DN [79]. The methodology of using a well-designed RCT was able to correctly identify the exact efficacy of TCM because it minimizes the risk of bias. This multi-center trial included 194 patients with DN at Mogensen stage III and stage IV and the intervention course was eight weeks. The primary outcome measure of this RCT was the change in UAE. There were three arms, i.e., a placebo group (60 S), low dose group (64 S), and

high dose group (62 S). The baseline values of UAE among the three arms were similar, and the values of UAE after eight weeks significantly changed: both the high dose and the low dose group were lower than the placebo group ($P < 0.01$), but there was no significant difference between the low dose and the high dose group ($P > 0.05$). Two subjects in the high dose group and one S in the low dose group had slight abdominal pain and diarrhea. The adverse events disappeared when TCM was terminated. The authors concluded that TJSK is effective for the management of proteinuria in DN patients. TCM is usually seen as a medicine based on experience, but a well-designed RCT could make it evidence-based.

Qiyao Xiaoke (QYXK)

Ni Qing and his co-workers (2013) carried out a multi-center randomized double blind placebo-controlled trial to assess the clinical effects of Qiyao Xiaoke (QYXK) capsule in the treatment of early DN [80]. The trial included DN patients whose 24-h UAE between 30 mg/24h to 300 mg/24h and 24-h proteinuria < 0.5 g. The trial lasted 12 weeks and the 24-h UAE and proteinuria were measured. 150 patients were included and 101 S were assigned to the TCM treatment group and 45 to the control group. Four S did not complete the trial. The statistical analysis based on the per-protocol set showed that the QYXK capsule could significantly improve proteinuria. The 24-h UAE decreased from 219.74 ± 68.36 to 109.52 ± 56.43 in the TCM treatment group; meanwhile, the values also decreased from 223.95 ± 70.67 to 142.67 ± 73.45 in the placebo control group. The authors reported the efficacy of TCM was superior to that of the placebo based on a paired t test. In general, this trial has optimal methodological quality, but it is a pity that it did not use intention-to-treat (ITT) analysis because the statistical analysis was based on the per-protocol set.

Combination Therapy with TCM and Western Medicine

Astragalus

The clinical trials where combination therapy with TCM and western medicine were used for the management of DN also reported positive results in reducing UAE. Astragalus is a famous Chinese medicinal herb in the treatment of DN and in China is usually used in injection form. Liu YH et al. (2005) carried out a RCT to assess the clinical efficacy of combination therapy with astragalus and captopril in the treatment of early DN [81]. 69 S were randomly assigned to three groups: a captopril group (23 S); astragalus group (23 S); and a combination group (23 S). The UAER value of the captopril group decreased from 140 ± 37.8 $\mu\text{g}/\text{min}$ to 111.2 ± 28.6 $\mu\text{g}/\text{min}$; that of the astragalus group decreased from 135.0 ± 33.4 $\mu\text{g}/\text{min}$ to 110.9 ± 21.5 $\mu\text{g}/\text{min}$; and that of the combination therapy group decreased from 138.8 ± 27.8 $\mu\text{g}/\text{min}$ to 96.2 ± 18.1 $\mu\text{g}/\text{min}$. The adverse events were cough (captopril group: 5 S and combination therapy group: 4 S); elevation in serum creatinine (captopril group: 1 S); and hyperkalemia (captopril group: 1 S). There was no adverse event in the astragalus injection group. The authors concluded that the combination therapy with astragalus injection and captopril could significantly reduce the UAER in patients with early DN.

Other TCM

Tu Xiang et al. (2014) reviewed the current literature concerning combination therapy with TCM and ACEI/ARB in the treatment of DN [82]. The authors searched three major Chinese

databases and PubMed. Rigorous inclusion criteria was applied and eventually only eight RCTs were included. The present chapter summarizes some interesting findings: three observational RCTs reported that TCM could provide benefits on proteinuria and renal function in addition to conventional western medicine (CWM); three observational RCTs reported that TCM provided added benefits on proteinuria to CWM; and two observational RCTs reported that TCM does not provide additional benefits on proteinuria or renal function to CWM. The authors fully discussed current progress in this area and stated – “Although the present review could provide an overall impression that TCM has some clinical effect, especially, when combined with ACEI/ARBs, it is really difficult to provide a clear picture about individual TCM or TCM products”.

What Do the Systematic Reviews and Meta-Analyses Report in This Area?

In sharp contrast to the positive results reported by RCTs, all systematic reviews and meta-analyses could not draw convincing conclusions. A closer look reveals some peculiar results.

TCM Extract

Breviscapine

The Chinese Journal of Evidence-based Medicine (CJEBM) published a series of systematic reviews to evaluate the efficacy and safety of TCM for the management of DN. Shi JY et al. (2009) published a systematic review which aimed to assess the efficacy and safety of breviscapine for DN, where the UAE levels were one of the primary outcome measures [83]. 33 original studies met the inclusion criteria and 2322 DN patients were eventually included (1214 S in the breviscapine treatment group and 1108 S in the control group). Breviscapine was administered intravenously. The results of meta-analyses demonstrated that breviscapine could significantly reduce the UAE level, compared with the controls, but the methodological quality of the 33 observational studies were classified as C grade (indicating poor methodology). The funnel plot was asymmetric, which indicates that negative results might not be published. The authors stated “no significant adverse events were reported” and they concluded “breviscapine shows some effects and is relatively safe on diabetic nephropathy”. Most importantly, the authors realized that “the evidence is not strong enough”.

Tripterygium Glycoside

Wu WH et al. (2010) carried out a systematic review to evaluate the efficacy and safety of Tripterygium Glycoside for DN [84]. 12 Chinese RCTs were included and none reported the allocation concealment, blindness, drop-out or attrition at follow-up, or ITT analysis. Obviously, methodology of most included studies was poor. The results of meta-analysis showed that Tripterygium Glycoside was superior to controls in reducing 24-h proteinuria [WMD=-0.49, 95%CI (-0.63, -0.34)] and 24-h UAE [WMD=-148.75, 95%CI (-238.01, -59.48)]. However, there was no sound evidence, and no reliable conclusions.

Ligustrazine

Wang B et al. (2012) published a meta-analysis to assess the clinical effect of ligustrazine, extracted from Chuanxiong (*Ligusticum chuanxiong* Hort), on diabetic nephropathy [85]. 25

RCTs involving 1645 patients were included. Although the authors concluded that “Ligustrazine injection has a significant therapeutic effect on ... reducing urine protein (24 h urine protein, urine micro albumin and urinary albumin excretion rate [UAER]) in DN patients”, the quality assessment from the Center for Reviews and Dissemination of the University of York arrived at a more cautionary statement – “Due to potential limitations in the included evidence and in the reporting and conduct of the review, this conclusion may not be reliable” [86].

Ginkgo Biloba Extract

Zhang L and his colleagues (2013) reported a systematic review and meta-analysis to evaluate the effectiveness and safety of a Ginkgo biloba extract for patients with early DN [87]. UAER was the primary outcome measure. 16 RCTs involving 1099 participants were included. Six trials elaborated randomization methods and none used blinding in the design. Although the methodological quality of observational studies were sub-optimal, the results of meta-analyses were in favor of Ginkgo biloba extract. Comparing the Ginkgo biloba extract plus conventional treatment versus conventional treatment only: 1) for the patients whose baseline UAER >150 µg/min, UAER decreased with an overall effect size of 74.52 µg/min (95% CI, from 63.89 to 85.15, $P < 0.00001$); 2) for the patients whose baseline UAER <100 µg/min, UAER decreased with an overall effect size of 23.88 µg/min (95% CI, from 22.58 to 25.18, $P < 0.00001$). Comparing the Ginkgo biloba extract combined with ACEI/ARB versus ACEI/ARB alone, UAER decreased with an overall effect size of 27.95 µg/min (95% CI, from 22.06 to 33.84, $P < 0.00001$). The authors therefore concluded that Ginkgo biloba extract is a “valuable drug” for DN, especially for those have high baseline UAER levels.

Puerarin

Wu Wei and his co-workers (2013) conducted a meta-analysis and reported that puerarin could significantly reduce UAE levels in patients with micro-albuminuria, no matter the dose low, middle or high [88]. Seven included studies scored 1 point and two studies scored two points, which means the results of meta-analysis were not sound – being based on poor methodological clinical trials.

Individual Traditional Chinese Medicinal Herb

Astragalus

There was only one individualized study. Li M et al. (2011) published a meta-analysis to evaluate the efficacy of astragalus in the treatment of DN [89]. 25 clinical trials involving 1804 patients were included. The results of meta-analysis showed that astragalus could significantly improve proteinuria compared with the controls. The authors believe that the results were encouraging. Few individualized herb meta-analysis has been published and astragalus is obviously the most extensively studied individual TCM herb.

Formulae

Xuezhikang

A systematic review was published in CJEBM (2009) which aimed to evaluate the efficacy and safety of Xuezhikang for DN [90]. Nine Chinese RCTs were included and they were classified as B grade in methodology indicating moderate bias. Xuezhikang was superior to routine western medicine in reducing 24-h proteinuria [WMD=-0.87, 95%CI (-1.34, -0.41)] and UAER [WMD=-65.46, 95%CI (-68.87, -62.12)]. It is also interesting that the results of meta-analysis showed Xuezhikang was similar to routine western medicine in improving serum creatinine. These results are in line with reports (pending publication) by Tu Xiang et al. [82] – the benefits of TCM on renal function parameters were not always consistent.

Danhong

Zhang MX et al. (2009) published a systematic review in the CJEBM to evaluate Danhong injection for DN [91]. Ten original studies involving 736 subjects were included. The dose of Danhong injection ranged from 20 ml/d – 40 ml/d and the treatment course ranged from two weeks to six weeks. There was significant difference between the Danhong injection treatment group and the control group in reducing UAER [MD=-27.08, 95%CI (-30.40, -24.02)]. Although the authors concluded “Danhong can improve UAER of DN”, they had to admit that “conclusive results cannot be made about the effectiveness and safety of Danhong for DN” (p. 1087) because all of the included studies were classified as grade C in methodology (the possibility of high risk of bias).

Tongxinluo

A systematic review (2010) assessed the clinical efficacy and safety of Tongxinluo capsule in the treatment of DN [92]. 11 Chinese RCTs were included and all of them used ITT analysis. The authors classified the 11 RCTs as B grade in methodology (having moderate bias). Compared with no treatment, Tongxinluo capsule could significantly reduce UAER [WMD=-38.88, 95%CI (-60.24, -17.52)]. Interestingly, Tongxinluo capsule was similar to no treatment in improving serum creatinine.

Other TCM

Zhang et al. (2009) conducted a meta-analysis to evaluate the anti-proteinuric effects of TCM clearing heat and activating blood in DN [93]. Eight RCTs were included and the primary outcome measure was the UAE level. The results of meta-analysis demonstrated that the efficacy of TCM clearing heat and activating blood might be similar to ACEI or ARB in reducing 24h proteinuria level (WMD=-1.81, 95%CI: -3.59 to -0.04). Compared with western medicine alone, combination therapy was superior in reducing 24h proteinuria and UAE. Due to the poor methodology of all eight observational RCTs, results of the meta-analysis were not valid. A systematic review (2012) was published to assess the combination therapy with TCM and losartan in the treatment of DN [94]. 28 studies were included and all were classified as C grade (probability of high risk of bias). The results of meta-analysis demonstrated that the combination therapy was superior to losartan alone in reducing UAER [WMD=-47.05, 95%CI (-64.99,-29.11)] and 24-h proteinuria [WMD=-0.56, 95%CI (-0.75,-0.37)]. However, it is almost impossible to arrive at firm conclusions because the observational studies were not sound.

A systematic review and meta-analysis (2013) published in Evidence-based Complementary and Alternative Medicine (eCAM) evaluated the effect of Chinese herbal medicine (CHM) on albuminuria in DN patients [73]. It included 29 trials involving 2440 participants and the results were positive for CHM. Compared with placebo, CHM was more effective in reducing UAER (mean difference [MD] -82.95 μ g/min, [-138.64, -27.26]) and proteinuria (MD -565.99 mg/24 h, [-892.41, -239.57]). CHM was even superior to ACEI or ARB in reducing UAER (MD -13.41 μ g/min, [-20.63, -6.19]) and proteinuria (MD -87.48 mg/24 h, [-142.90, -32.06]). Compared with ACEI/ARB alone, combination therapy with CHM and ACEI/ARB significantly improved UAER (MD -28.18 μ g/min, [-44.4, -11.97]) and proteinuria (MD -26.60 mg/24 h, [-26.73, -26.47]). 11 observational trials were “of superior quality” and “no serious adverse events were reported”. The authors concluded “CHM seems to be an effective and safe therapy option to treat proteinuric patients with DN”, but they also suggested future well designed, large sample size RCT is warranted.

What Are the Limitations with Respect to this Clinical Evidence?

As mentioned above, few RCTs describe the details of randomization, used allocation concealment, blindness or ITT analysis. Most Chinese RCTs reported positive results, but the funnel plots in meta-analyses were usually asymmetric suggesting there was publication bias. “Negative reports are still rarely reported in Chinese journals and this phenomenon should be corrected in the future. [82]” Tu Xiang et al. have elaborated exclusively on the shortcomings of current literature in this area and hence, readers are hereby referred to their recent article on *Curr Vasc Pharmacol* [82]. Although most systematic reviews and meta-analyses arrived at conclusions which favor TCM, their conclusions should be interpreted cautiously because the included studies were usually of sub-optimal methodology. In a word, the limitation of current evidence is methodological shortcoming. It is obvious we cannot draw any firm conclusions if the evidence is not sound.

THE EFFECTS OF TCM ON DN AT THE GENETIC LEVEL

What Are the Effects of TCM on DN at the Genetic Level in Animal Experiments?

Current reports studying the genetic effects of TCM indicate that TCM might exert benefits on DN by regulating some genes in animal experiments.

TCM Extract

Gypenosides

Tao LH et al. (2007) compared the difference of gene expression profile between the DN rats and the gypenosides treated DN rats [95]. This animal experiment reported gypenosides could improve blood glucose, proteinuria and SCr, compared with DN rats. The major finding

was the alteration of the gene’s expression: 135 genes altered between the DN model rats and the gypenosides treated DN rats; 31 genes up-regulated; and 104 genes down-regulated. Some altered genes are listed in Table 2. Now that NM-053336 and NM-147205 genes are related to advanced glycosylation end products (AGEs), the authors concluded gypenosides conferred renal benefits by inhibiting the formation of AGEs.

Table 2. Altered genes between the DN model rats and the gypenosides treated DN rats in the article published by Tao LH et al

Gene bank number
NM-053336
NM-147205
NM-021657
XM-134042
NM-021842
NM-031507

Breviscapine

Only one article (2008) explored the genetic effect of breviscapine in DN animal models [96]. In the kidney tissues of diabetic rats induced by streptozotocin (STZ), the mRNA expression level of the p47phox gene significantly increased, which indicated the oxidative stress was enhanced. The diabetic rats given breviscapine showed not only improved serum creatinine (SCr) and blood urea nitrogen (BUN), but also reduced mRNA expression levels of the p47phox gene. The authors concluded that their experiment indicated that breviscapine could ameliorate renal function of diabetic rats by inhibiting the oxidative stress in the kidney.

Artemisinin

Zhou FJ et al. (2014) studied the effects of artemisinin on the genetic expression of c-jun and c-fos in diabetic rats induced by STZ injection [97]. The histomorphology in the artemisinin treated group were significantly improved, compared with the diabetic rats. Using immunohistochemistry and reverse transcription-polymerase chain reaction (RT-PCR) methods, the authors detected the changes of c-jun gene and c-fos gene in the kidney tissues: their mRNA and protein expression significantly increased in diabetic rats and the elevation was significantly inhibited in artemisinin treated rats. The conclusion was as follows: artemisinin could ameliorate renal function injury and pathological changes in diabetic rats by inhibiting the expression of c-jun gene and c-fos gene in kidney tissues.

Individual Traditional Chinese Medicinal Herb

Astragalus

Astragalus not only reduced the levels of proteinuria, SCr and BUN in humans, but also in animals. Applying gene chip technology produced encouraging results. An analysis of 88 genes altered between DN mice induced by STZ and DN mice treated by astragalus resulted in 81 genes being changed in the opposite direction and the remaining 7 genes being changed in the

same direction [98]. The altered genes were associated with metabolism, inflammation and immunity, signal transduction, etc. Some altered genes are listed in Table 3.

Table 3. Altered genes between the DN model mice and the astragalus treated DN mice in the article published by Hong XP et al.

Gene bank number
NM-133808
AB042432
BC015248
L07096
V00711
NM-028024
NM-010382
NM-008363
XM-129176
NM-018738
NM-009071
XM-127336
XM-125952
AF176524
BC006714
BC025573
NM-009174
BC020286

Formulae

Shenxiao Kang (SXX)

During 2006 – 2012, Jiang DY and his team published a six-part series of articles to study the expression difference of various genes in DN rats induced by STZ and in the rats given SXX [99-104]. An array of technology was applied in their publication, such as electron microscopy, gene chip technology, RT-PCR, etc. In conclusion, all of their publications reported that SXX could improve the levels of proteinuria, SCr, BUN and the histopathological changes of DN rats. The authors reported that Gna12, Eif3s8, myosin Va were down-regulated and that S100A9, Cxcl12 and CcL5mRNA were up-regulated in the kidney tissues. SXX could up-regulate Gna12 [99], Eif3s8 [100] and myosin Va [101] and down-regulate S100A9 [102], Cxcl12 [103] and CcL5mRNA [104]. The authors concluded that SXX exerted benefits in DN rats by regulating these genes.

What Are the Effects of TCM on DN at the Genetic Level in Humans?

Articles studying the genetic effects of TCM on DN patients were scarce. Two identified articles explored the genetic effects of TCM formulae.

Xu Xun et al. (2011) carried out a clinical study to assess the effects of Tangshen Formula (TSF) on three hemodynamic susceptibility genes [105]. 31 DN patients given TSF for six months and 23 healthy subjects were the controls. Glomerular filtration rate (GFR), AGT gene, AGTR₂ gene, ADR β ₃ gene were measured at month 3 and month 6. The results showed that the GFR significantly improved at month 3 and then, continued to improve at month 6. The expression of the three genes gradually increased within six months and approached to normal levels. The amelioration of the three genes at month 6 was better than that at month 3. The authors postulated that TSF could inhibit angiotensin-converting enzyme (ACE) in the rennin-angiotensin system (RAS) and play a role like an ACE inhibitor, by regulating the AGT gene and AGTR₂ gene. Their second hypothesis was that TSF improved GFR by improving insulin resistance via ADR β ₃. The expression trend chart of the three genes were highly consistent, which suggested “the mechanism of the clinical efficacy of TSF has multi-pathways”. Wang HW and his team (2013) published an article to study the genetic effects of TCM supplementing *Qi* and nourishing *Yin* on patients with DN [106]. 65 S were included; for three months they took Tangwei Kang capsule, which is a patented TCM - supplementing *Qi* and nourishing *Yin*. The serum homocysteine (Hcy) and methylene-tetrahydrofolate reductase (MTHFR) gene polymorphism were measured by enzyme-linked immuno sorbent assay (ELISA) and the polymerase chain reaction - restriction enzyme fragment length polymorphism (PCR-RFLP) technique, respectively. The values of the 65 S were measured at the beginning and the end of the study. The results showed that TCM could significantly decrease the Hcy levels in DN patients. 44 S had MTHFR C677T gene mutation and the MTHFR gene polymorphism did not change in patients receiving TCM. The authors explained the results as follows: Tangwei Kang capsule cannot reverse gene mutation, but could obviously decrease the serum Hcy level and this might be because Tangwei Kang improved the activity decline of the MTHFR caused by the MTHFR C677T gene mutation.

THE ASSOCIATION BETWEEN THE CLINICAL EVIDENCE AND THE GENETIC EFFECTS OF TCM

What Is the Association?

Articles exploring the association between the clinical efficacy and the genetic effects of TCM were scarce (Reviewed 1979-2014). Only one published article addressed this issue [107]. 60 DN patients were randomly allocated to the Xiao SiWu Tang (XSWT) group and the captopril group. Proteinuria decreased in both groups and there was no significant difference between the two groups. The XSWT could significantly improve hemodynamic parameters, but the captopril failed. The ACE gene polymorphism was also tested. As a result, the authors detected that various genotypes have obvious differences in efficacy; DD genotype and D allele were associated with clinical efficacy of XSWT.

The clinical efficacy is the fundamental basis for complementary and alternative medicine. Therefore, the clinical evidence of TCM in the treatment of DN is the key to many questions. Current literature discloses that there is impoverished evidence, but new persuasive evidence is emerging. At least, for proteinuria, TCM may be promising. On the other hand, it is notable that current literature reveals that the genetic effects of TCM were primarily established on DN

animal models. For example, SXK may be effective in altering the expression of some rat genes, but there is not any sound evidence to support its application in proteinuric DN human patients. Those TCM which have sound clinical anti-proteinuric evidence have not been studied at the genetic level, such as QYXK capsule. However, the anti- proteinuric effect of TCM may be conferred by regulating gene expression – breviscapine and astragalus are good examples. In short, the association between the clinical evidence and the genetic effects of TCM is currently lacking.

How to Relate Them?

Although it is not easy to relate clinical evidence and the genetic effects of TCM, if the TCM community intends to push on with TCM-gene research, the relationship has to be taken seriously. It is obvious that TCM needs the translational conception to bridge the gap between achievements from the research lab and hands-on clinical practice. From a holistic perspective, the balance of the *Tao* in medicine is clinical TCM evidence, observation and documentation of genetic effect, comparison/inclusion of western medicine, and creating a new east/west synthesis model. Current clinical evidence may indicate which TCM deserves investigation. Genes disclosed by animal experiments may be good targets to study. When a well-designed large sample size RCT is carried out, it may be the best chance for TCM to explore the solid genetic effects. When sound clinical evidence concerning the genetic evidence is established, it will be invaluable to guide TCM clinical practice and research.

SUMMARIZATION OF CURRENT LITERATURE

Clinical Evidence

1. A large number of RCTs produced positive results with respect to TCM for proteinuria in patients with DN.
2. RCTs of high methodological quality in this area remain scarce.
3. Systematic reviews or meta-analyses cannot arrive at firm conclusions because the included studies are usually of sub-optimal methodology.

Genetic Effects

1. Current literature reports that TCM may affect some particular genes on DN etiology.
2. TCM may alter the expression of the genetic profile in patients with DN.
3. TCM which has been reported to possess genetic effects are not supported by strong clinical evidence.

CONCLUSION

The current literature demonstrates a disconnect between clinical evidence and demonstrable TCM genetic effects; stronger correlations are needed.

Conflict of Interest

None declared.

ACKNOWLEDGMENTS

The present study is funded by the Sichuan Province 2014: Support Project in Science & Technology (CKY2014007) and supported by Demonstration Project Chengdu City - 2014: Improvement of Livelihood through Science & Technology (2014-HM02-00014-SF).

REFERENCES

- [1] Tervaert TW, Mooyaart AL, Amann K, et al. Pathologic classification of diabetic nephropathy. *J Am Soc Nephrol*. 2010; 21: 556-63.
- [2] Zheng TS, Xiang KS, Sun DQ. Association of RENIN gene with type 2 diabetes mellitus, and its microangiopathy, coronary heart disease and hypertension in chinese. *Chin J Endocrinol Metab*. 1997; 13(4): 206-209.
- [3] Zhang YL, Li Y, Gu LF, et al. Relationship between angiotensin I converting enzyme gene I/D polymorphism and diabetic nephropathy in NIDDM in the Hans of Hebei province. *Chin J Nephrol*. 1999; 15(1): 40-42.
- [4] Liu LM, Xiang SS. A study for RAGE Gly82Ser polymorphism in Chinese type 2 diabetic microangiopathy. *Chinese Journal of Diabetes*. 2000; 8(3): 135-138.
- [5] Wong TY, Poon P, Szeto CC, et al. Association of plasminogen activator inhibitor-1 4G/4G genotype and type 2 diabetic nephropathy in Chinese patients. *Kidney Int*. 2000; 57(2): 632-638.
- [6] Wu S, Xiang K, Zheng T, et al. Relationship between the renin-angiotensin system genes and diabetic nephropathy in the Chinese. *Chin Med J (Engl)*. 2000; 113(5): 437-441.
- [7] Qu HQ, Lin S, Qiu MC. Meta-analysis on the association between the insertion/ deletion polymorphism of ACE gene and diabetic nephropathy among Chinese population. *Chinese Journal of Evidence-Based Medicine*. 2001; 1(4): 218-225.
- [8] Yang T, Chen JW, Shen J. A study for HSPG gene polymorphism in Chinese type 2 diabetic nephropathy. *Chinese Journal of Diabetes*. 2001; 9(2): 82-86.
- [9] Cheng YB, Luo R, Xue YM, et al. Research of the Association of the ACE Gene with the Kidney-Asthenia Type of DN. *Chinese Journal of Basic Medicine in Traditional Chinese Medicine*. 2002; 8(5): 29-30, 33.

- [10] Fu HM, Wu HW, Yang JP, et al. Research of the crowd of genetic epidemiology of diabetic nephropathy and Angiotensin-Converting Enzyme gene polymorphism in Western Hunan population of Hmong nationality. *Chin J Nephrol*. 2002; 18(3): 228-229.
- [11] Liao YH, Xian S, Luo ZJ, et al. Relationship between the polymorphism of angiotensin enzyme gene and vascular complications in type 2 diabetes in Guangxi region. *Chinese Journal of Practical Internal Medicine*. 2002; 22(5): 270-272.
- [12] Liu YF, Wat NM, Chung SS, et al. Diabetic nephropathy is associated with the 5'-end dinucleotide repeat polymorphism of the aldose reductase gene in Chinese Subjects with Type 2 diabetes. *Diabet Med*. 2002; 19(2): 113-118.
- [13] Shen HL, Liu LM, Xiang KS, et al. Relationship between ApoE gene polymorphism and type 2 diabetes mellitus with its nephropathy in Chinese. *Chinese Journal of Diabetes*. 2002; 10(1): 4-6.
- [14] Xu SJ, Yang RZ, Fu ZZ, et al. A study on the association between atrial natriuretic peptide (ANP) gene polymorphism and proteinuria in type 2 diabetes mellitus. *Journal of new medicine*. 2002; 33(4): 201-202.
- [15] Liu L, Xiang K, Zheng T, et al. The heparan sulfate proteoglycan gene polymorphism: association with type 2 diabetic nephropathy in Chinese. *Mol Cell Biochem*. 2003; 245(1-2): 121-126.
- [16] Liu L, Xiang K, Zheng T, et al. Co-inheritance of specific genotypes of HSPG and ApoE gene increases risk of type 2 diabetic nephropathy. *Mol Cell Biochem*. 2003; 254 (1-2):353-358.
- [17] Xu JS, Zhang J, Shan BE, et al. Relationship between methylenetetrahydrofolate reductase gene polymorphism and diabetic nephropathy in type 2 diabetes mellitus in the Hans of Hebei Province. *Clinical Focus*. 2003; 18(14): 787-789.
- [18] Wang Y, Ng MC, Lee SC, et al. Phenotypic heterogeneity and associations of two aldose reductase gene polymorphisms with nephropathy and retinopathy in type 2 diabetes. *Diabetes Care*. 2003; 26(8): 2410-2415.
- [19] Wong TY, Poon P, Chow KM, et al. Association of transforming growth factor-beta (TGF-beta) T869C (Leu 10Pro) gene polymorphisms with type 2 diabetic nephropathy in Chinese. *Kidney Int*. 2003; 63(5): 1831-1835.
- [20] Chen AR, Ning YY, Zhu XZ, et al. Study on the relationship between gene polymorphisms of N5,10-methylenetetrahydrofolate reductase and nephropathy in type 2 diabetes mellitus in Gansu Han Chinese of China. *Chin J Prev Contr Non-commun Dis*. 2004; 12(5): 195-197.
- [21] Liu SQ, Xue YM, Yang GC, et al. Relationship between plasminogen activator inhibitor-1 gene 4G/5G polymorphism and type 2 diabetic nephropathy in Chinese Han patients in Guangdong Province. *J First Mil Med Univ*. 2004; 24(8): 904-907.
- [22] Qian SH, Han QW. Relationship between arylesterase/paraoxonase activity and ArE/PON1 Q192R gene polymorphism and patients with 2 type diabetic renal failure in Zhengzhou population of Han nationality. *Clinical Focus*. 2004; 19(24): 1384 -1387.
- [23] Su Y, Chang WW, Niu HY, et al. Detection of hepatocyte nuclear factor-1beta gene exon2 in diabetes mellitus nephron patients of the Han nationality in Henan. *Journal of Zhengzhou University (Medical Sciences)*. 2004; 39(6): 1063-1065.
- [24] Sun HY, Yang MG, Liu SQ, et al. Study on the correlation of the polymorphisms of endothelial nitric oxide synthase gene with type 2 diabetes mellitus and diabetic nephropathy. *Chin J Prev Contr Chron Non-commun Dis*. 2004; 12(3): 101-103.

- [25] Yang HY, Tai WL, Yuan HY, et al. Study on the association between the polymorphism of HLA-DQA1 alleles and type 2 diabetes in Yunnan Han nationality. *Chin J Med Genet.* 2004; 21(3): 291-293.
- [26] Baum L, Ng MC, So WY, et al. Effect of hepatic lipase -514C->T polymorphism and its interactions with apolipoprotein C3 -482C->T and apolipoprotein E exon 4 polymorphisms on the risk of nephropathy in Chinese type 2 diabetic patients. *Diabetes Care.* 2005; 28(7): 1704-1709.
- [27] Li HF, Song DP, Zhang M, et al. The polymorphism of promoter region 59029G/A of CCR5 gene is related to diabetic nephropathy. *Chin J Diabetes.* 2005; 13(3): 206-208, 212.
- [28] Li HM, Miao H, Lu YB, et al. Association between the polymorphism of human vitamin D receptor gene and the susceptibility of diabetic nephropathy in Chinese Han population. *Chinese Journal of Clinical Rehabilitation.* 2005; 9(47): 1-4.
- [29] Yang Y, Song DP, Liu H, et al. The Ethnic Variation in the V (16) A Polymorphism of MnSOD gene and the relationship between this polymorphism and diabetic nephropathy. *Academic Journal of Kunming Medical College.* 2005; (2): 10-14.
- [30] Ng MC, Baum L, So WY, et al. Association of lipoprotein lipase S447X, apolipoprotein E exon 4, and apoC3-455T>C polymorphisms on the susceptibility to diabetic nephropathy. *Clin Genet.* 2006; 70(1): 20-28.
- [31] Xie JM, Wang YZ, Yue XL, et al. Relationship between angiotensin I converting enzyme gene I/D polymorphism and diabetic. *Acta Acad Med Nei Mongol.* 2006; 14-18.
- [32] Zhang S, Shi LX, Wang R, et al. An analysis of relationship between polymorphism of RANTES gene promoter and type 2 diabetes mellitus with nephropathy in the Hans of Guizhou Province. *Guizhou Medical Journal.* 2006; 30(11): 970-973.
- [33] Dong JJ, Zhao JJ, Li GA, et al. Distribution of polymorphism in endothelial nitric oxide synthase gene in mainland Chinese and Singapore Chinese and its association with diabetic nephropathy. *Natl Med J China.* 2007; 87(48): 3415-3417.
- [34] Luo P, Li BY, Wen L, et al. Association between AT1R gene A1166C polymorphism and type 2 diabetic nephropathy in the Hans of Kunming. *Journal of Practical Diabetology.* 2007; 3(6): 32-34.
- [35] Zhang GW, Zhang AZ, Xu ZF. An analysis of the relationship between apolipoprotein E(Apo E) gene polymorphism and Chinese patients with type 2 diabetes mellitus and diabetic nephropathy. *Zhejiang clinical medicine.* 2007; 9(6): 735-736.
- [36] Zhong WG, Wei YL, Liu J, et al. Associations of AGT gene polymorphisms with type2 diabetic nephropathy in Guangxi people. *Journal of Guangxi Traditional Chinese Medical University.* 2007; 10(4): 17-20.
- [37] Li SB, Yang GY, Li L, et al. Meta-analysis of the association of ApoE gene polymorphisms with type 2 diabetic nephropathy in Chinese Han population. *Chin J Diabetes.* 2008; 16(10): 580-583.
- [38] Wen L, Li BY, Luo p, et al. Association between ACE and AT1R gene polymorphism and type 2 diabetic nephropathy in the Hans of Kunming. *Journal of Practical Diabetology.* 2008; 4(6): 34-36.
- [39] Xianmusiya R, Fan Y, Halida M, et al. The correlation between genetic polymorphism of Apo-E and diabetes in uygur and Han populations. *Journal of Xinjiang Medical University.* 2008; 31(9): 1178-1180.

- [40] Liu L, Zheng T, Wang N, et al. The manganese superoxide dismutase Val16Ala polymorphism is associated with decreased risk of diabetic nephropathy in Chinese Patients with type 2 diabetes. *Mol Cell Biochem.* 2009; 322(1-2): 87-91.
- [41] Tang HW, Hu DL, Liu TJ, et al. A study on the association of MCP-I gene polymorphism with type 2 diabetics and diabetic nephropathy in Zhuang minority population. *Chin J Diabetes.* 2009; 17(6): 430-432,436.
- [42] Wan G, Feng J, Sun L, et al. Meta analysis on relationship between ApoE polymorphism and type 2 diabetic nephropathy in Chinese population. *Chin J Diabetes.* 2009; 17(6): 433-6.
- [43] Chen JN, Li XL, Wei JZ, et al. The correlation of the ICAM-1 gene and type II diabetic nephropathy in Zhuang populations in the northern regions of Guilin. *Acta Medicinæ Sinica.* 2010; 23(1): 10-14.
- [44] Liu WJ, Wang YM, Liu H, et al. PCR-DGGE scanning of the manganese superoxide dismutase gene in Chinese patients with type 2 diabetes mellitus. *Journal of Kunming Medical University.* 2010; (11):80-83.
- [45] Tang SC, Leng VT, Chan LY, et al. The acetyl-coenzyme A carboxylase beta (ACACB) gene is associated with nephropathy in Chinese patients with type 2 diabetes. *Nephrol Dial Transplant.* 2010; 25(12): 393-4.
- [46] Wang QL, Ma YZ, Bai LM, et al. Associations of ACE gene polymorphisms with type 2 diabetic and nephropathy merger in Neimenggu Province. *Journal of Inner Mongolia Normal University (Natural Science Edition).* 2010; 39(5): 505-9.
- [47] Xue J, Tian GS, Shi FH, et al. Polymorphisms of plasminogen activator inhibitor-1 genes in type 2 diabetes with nephropathy in Han in Baotou. *Chin Med J Metall Indus.* 2010; 27(4):373-5.
- [48] Jin GM, Piao LS, Li L. Relation of NADPH oxidase subunit p22phox gene C242T polymorphism in Korean patients with diabetic nephropathy in Yanbian. *Jinlin Medical Journal.* 2011; 32(7):1336-8.
- [49] Jin GM, Zheng SH. Relation of C-159T polymorphism in promoter region of CD14 and Korean patients with diabetic nephropathy in Yanbian area. *China Journal of Modern Medicine.* 2011; 21(31): 3876-80.
- [50] Wang Y, Wu WJ, Shi R. Meta-analysis of association of ApoE gene polymorphisms with diabetic nephropathy. *Journal of Shanghai Jiaotong University (Medical Science).* 2011; 31(5): 625-631.
- [51] Chen B, Song DP, Wu YB, et al. The study of the association between ApoC3 gene promoter region-482C→T polymorphism and diabetic nephropathy. *Chin J Diabetes.* 2012; 20(4): 263-6.
- [52] Jia HY, Wen SL. Relationship between diabetic nephropathy and angiotensin converting enzyme insertion/deletion polymorphism in Chinese Han population: A systematic review. *Chin J Diabetes.* 2012; 20(6): 414-7.
- [53] Lin CJ, Wen MJ, Hung YJ, et al. The impact of 5,10-methenyltetrahydrofolate synthetase polymorphism on diabetic nephropathy in the Taiwanese population. *Genet Test Mol Biomarkers.* 2012; 16(2): 142-5.
- [54] Liu XD, Fan QL, Yao L, et al. Association of ACE gene I/D polymorphism and diabetic nephropathy in Chinese Han population: a Meta-analysis. *Journal of Logistics University of CAPF (Medical Sciences).* 2012; 21(6): 412-416, 420.

- [55] Peng CL, Hong YZ, Fu LP. Associations of adiponectin gene polymorphism in type 2 diabetes with nephropathy. *J Med Res*. 2012; 41(4): 156-9.
- [56] Peng SH, Chen T, Li CR, et al. Insertion/Deletion polymorphism in the Angiotensin I-Converting Enzyme gene and risk of progression of diabetic nephropathy in noninsulin-dependant diabetes mellitus among the Han Chinese population: A Meta-analysis. *Chinese General Practice*. 2012; 15(12B): 4092-6.
- [57] Rao GF, Chen EF, Lin HY, et al. The correlation of eNOS gene polymorphism and susceptibility to diabetic nephropathy. *China Modern Doctor*. 2012; 50(7): 49-50.
- [58] Li YC, Chen WX, Xiao L. Single nucleotide polymorphism analysis Guangxi region elderly population screening diabetic nephropathy susceptibility genes Applications. *Chinese Journal of Gerontology*. 2013; 33(24): 6088-90.
- [59] Long AM, Zuo YF, Huang XQ, et al. Mets-analysis: Association of angiotensin converting enzyme gene insertion (I),deletion (D) polymorphism with diabetic nephropathy in Chinese Han population. *Chin J Diabetes*. 2013; 21(2):118-121.
- [60] Lv J, Cai CY, Wei FJ, et al. TOX and SMAD3 gene variants associated with microvascular complications of type 2 diabetes in Han Chinese. *Journal of Tianjin Medical University*. 2013; 19(1): 9-12.
- [61] Wu HY1, Wang Y, Chen M, et al. Association of ELMO1 gene polymorphisms with diabetic nephropathy in Chinese population. *J Endocrinol Invest*. 2013; 36(5): 298-302.
- [62] Yin X, Li H, Xuan J, Chen Y, et al. AGTR1 A1166C polymorphism is associated with risk of diabetic nephropathy. *Zhejiang Da Xue Xue Bao Yi Xue Ban*. 2013; 42(1): 45-51.
- [63] Zhang N, Zhuang M, Ma A, et al. Association of Levels of Mannose-Binding Lectin and the MBL2 Gene with Type 2 Diabetes and Diabetic Nephropathy. *Plos One*. 2013; 20(8): e83059.
- [64] Liu KH, Pang DZ, Zou CS, et al. Relationship of C6677T polymorphism of MTHFR gene with diabetic nephropathy in southeast Guangxi. *Int J Lab Med*. 2014; 35(13):1670-2.
- [65] Liao LN, Chen CC, Wu FY, et al. Identified single-nucleotide polymorphisms and haplotypes at 16q22.1 increase diabetic nephropathy risk in Han Chinese population. *BMC Gent*. 2014; 31(15): 113-24.
- [66] Peng D, Wang J, Pan J, et al. Lack of association between TLR4 genetic polymorphisms and diabetic nephropathy in a Chinese population. *Biomed Res Int*. 2014; 2014: 704167.
- [67] Yao QY, Lin JG, Ma Z, et al. Relationship between the polymorphism of RAGE 2245G/A gene and type 2 diabetic nephropathy. *Shandong Medical Journal*. 2014; 54(17): 53-5.
- [68] Zhao MM, Zhang R, Li C, et al. Relationship between variation of rs3760106 of PRKCB1 gene and incidence of end-stage renal disease of Shanghai Han people with type 2 diabetes mellitus. *Journal of Shanghai Jiao Tong University Medical Science*. 2014; 34(7):1027-33.
- [69] Zhou J, Peng R, Luo YJ, et al. Correlation of single nucleotide polymorphism of let-7a gene regulatory region with diabetic nephropathy in 108 cases of Chinese Han population. *J Third Mil Med Univ*. 2014; 36(12): 1312-6.
- [70] Tu X, Liu F, Jordan JB, et al. 'Huang Qi Elixir' for proteinuria in patients with diabetic nephropathy: a study protocol for a randomized controlled pilot trial. *Trials*. 2013; 18; 14: 223.

- [71] Wang H, Mu W, Zhai J, et al. The key role of Shenyan Kangfu tablets, a Chinese patent medicine for diabetic nephropathy: study protocol for a randomized, double-blind and placebo-controlled clinical trial. *Trials*. 2013; 5; 14: 165.
- [72] Wang B, Lin L, Ni Q, et al. Chinese medicine for treating diabetic nephropathy. *Chin J Integr Med*. 2011; 17(10): 794-800.
- [73] Xiao Y, Liu Y, Yu K, et al. The effect of Chinese herbal medicine on albuminuria levels in patients with diabetic nephropathy: a systematic review and meta-analysis. *Evid Based Complement Alternat Med*. 2013; 2013: 937549.
- [74] Wan Z, Chen X. Triptolide inhibits human immunodeficiency virus type 1 replication by promoting proteasomal degradation of Tat protein. *Retrovirology*. 2014; 11(1): 88.
- [75] Peng A, Huang X, Liu R, et al. Triptolide inhibits the inflammatory response of monocytes from rheumatoid arthritis patients by regulating miR-155. *Xi Bao Yu Fen Zi Mian Yi Xue Za Zhi*. 2014; 30(6): 635-8.
- [76] Ma R, Liu L, Liu X, et al. Triptolide markedly attenuates albuminuria and podocyte injury in an animal model of diabetic nephropathy. *Exp Ther Med*. 2013; 6(3): 649-656.
- [77] Ma RX, Xu Y, Zhang J, et al. Triptolide combined with irbesartan synergistically blocks podocyte injury in a type 2 diabetes rat model. *Zhonghua Nei Ke Za Zhi*. 2012; 51(2): 117-22.
- [78] Effect of tripterygium wilfordii in patients with diabetic nephropathy: a prospective randomized control clinical trial. *J Nephrol Dialy Transplant*. 2010; 19(6): 501-7, 533.
- [79] Ma ST, Liu DL, Niu Rui, et al. Tangjiang Shenkang granule in treatment of diabetic nephropathy: a double-blind, randomized, placebo-controlled multicenter clinical trial. *Chin J New Drugs Clin Rem*. 2011; 30(1): 16-9.
- [80] Ni Qing, Jiang Shan, Xiao YX, et al. Clinical observation of Qiyao Xiaoke capsule treating early diabetic nephropathy with multi-center, randomized, double-blind and placebo control groups. *China Journal of Traditional Chinese Medicine and Pharmacy*. 2013; 28(8): 2479-82.
- [81] Liu YH, Yang L, Liu J. Clinical observation on treatment of early diabetic nephropathy by milkvetch injection combined with captopril. *Zhongguo Zhong Xi Yi Jie He Za Zhi*. 2005; 25(11): 993-5.
- [82] Tu X, Ye XF, Xie CG, et al. Combination therapy with Chinese medicine and ACEI/ARB for the management of diabetic nephropathy: the promise in research fragments. *Curr Vasc Pharmacol*. 2014 Oct 14. [Epub ahead of print].
- [83] Shi JY, Huang Jing, Tian HM. Effect of breviscapine on diabetic nephropathy: a systematic review. *Chinese Journal of Evidence-based Medicine*. 2009; 9(10): 1099-109.
- [84] Wu WH, Wang Han, Zhang MP, et al. Effects of Tripterygium on diabetic nephropathy: a systematic review. *Chinese Journal of Evidence-based Medicine*. 2010; 10(6): 693-9.
- [85] Wang B, Ni Q, Wang X, et al. Meta-analysis of the clinical effect of ligustrazine on diabetic nephropathy. *Am J Chin Med*. 2012; 40(1): 25-37.
- [86] <http://www.ncbi.nlm.nih.gov/pubmedhealth/PMH0041299/>.
- [87] Zhang L, Mao W, Guo X, et al. Ginkgo biloba Extract for Patients with Early Diabetic Nephropathy: A Systematic Review. *Evid Based Complement Alternat Med*. 2013; 2013: 689142.
- [88] Wu Wei, Zhu ZZ, Li Hong, et al. Meta-analysis of puerarin treating early type 2 diabetic nephropathy. *Chinese Traditional Patent Medicine*. 2013; 35(7): 1399-406.

- [89] Li M, Wang W, Xue J, et al. Meta-analysis of the clinical value of Astragalus membranaceus in diabetic nephropathy. *J Ethnopharmacol*. 2011; 133(2): 412-9.
- [90] Wang Feng, Wu HM. Xuezhikang for diabetic nephropathy: a systematic review. *Chinese Journal of Evidence-based Medicine*. 2009; 9(1): 63-70.
- [91] Zhang MX, Cui Kai, Zhu YM, et al. Dan Hong injection for diabetic nephropathy: a systematic review. *Chinese Journal of Evidence-based Medicine*. 2009; 9(10): 1087-93.
- [92] Long Xuan, Wang Feng, Huang CQ. Tongxinluo capsule for diabetic nephropathy: a systematic review. *Chinese Journal of Evidence-based Medicine*. 2010; 10(1): 73-80.
- [93] Zhang Yan, Ye XF, Shen Wei. The efficacy of Chinese herbal medicine clearing heat and activating blood circulation in the management of urine protein in diabetic nephropathy: a meta-analysis. *Global Traditional Chinese Medicine*. 2009; 2(6): 435-8.
- [94] Cheng FF, Wang QQ, Zhong XG, et al. Systematic evaluation of traditional medicine plus losartan for diabetic nephropathy. *Chinese Journal of Integrated Traditional and Western Nephrology*. 2012; 13(9): 794-9.
- [95] Tao LH, Huang Ping, Qian Kang, et al. Effects of gypenosides on gene expression profile in diabetic nephropathy rats. *Chinese Journal of Integrated Traditional and Western Nephrology*. 2007; 8(12): 718-9.
- [96] Xiao Hua, Yang JJ, Liu Hua, et al. An experimental study on the effect of breviscapine on the expression of p47phox gene mRNA of diabetic rat's kidney. *Yunan Journal of Traditional Chinese Medicine and Materia Medica*. 2008; 29(7): 49-51.
- [97] Zhou FJ, Zhang LK, Su YJ, et al. Inhibitory effect of artemisinin on the up-regulation of c-fos and c-jun gene expression in kidney tissue of diabetic rats. *Modern Journal of Integrated Traditional Chinese and Western Medicine*. 2014; 23(21): 2294-8.
- [98] Hong XP, Zhang XZ, He XL, et al. Effects of astraglus on gene expression profile in diabetic nephropathy mice. *China Journal of Chinese Materia Medica*. 2008; 33(6): 676-80.
- [99] Qiao Yu, Xie Ying, Bai YB, et al. Effect of Shenxiaokang on the renal tissues expression of gene Gna12 in diabetic nephropathy rats. *Tianjin Journal of Traditional Chinese Medicine*. 2012; 29(5):469-74.
- [100] Wang Bin, Qiao Yu, Bai YB, et al. Effect of Shenxiaokang on diabetic nephropathy rats renal tissues gene – eIF3s8 expression. *Lishizhen Medicine and Materia Medica Research*. 2012; 23(1): 624.
- [101] Bai YB, Wang Bin, Liu CH, et al. Effect of Shenxiaokang on diabetic nephropathy rats renal tissues gene – myosin Va expression. *Journal of Liaoning University of Traditional Chinese Medicine*. 2011; 13(9): 57-9.
- [102] Liu CH, Wang Bin, Qiao Yu, et al. Effect of Shenxiaokang on gene S100A9 expression in diabetic rats. *Lishizhen Medicine and Materia Medica Research*. 2012; 23(10): 2518-20.
- [103] Jiang DY, Wang JX, Wang Bin, et al. Effect of Shenxiaokang on diabetic nephropathy rats renal tissues gene – Cxcl12 expression. *Chinese Archives of Traditional Chinese Medicine*. 2009; 27(11): 2254-6.
- [104] Jiang DY, Bai YB, Liu CG, et al. Effect of Shenxiaokang on gene – Ccl5 expression in diabetic rats. *World Science and Technology/Modernization of Traditional Chinese Medicine and Materia Medica*. 2006; 8(5): 37-42.

- [105] Xu Xun, Wang YM, Wu YY, et al. Effect of Tangshen Formula on susceptibility allele expression of hemodynamics in patients with diabetic nephropathy. *Chinese Traditional Patent Medicine*. 2011; 33(4):562-6.
- [106] Wang HW, Lin Lan, Ni Qing, et al. Affection of supplementing qi and nourishing yin drugs on prophase diabetic nephropathy Hcy and gene polymorphism of MTHFR. *China Journal of Traditional Chinese Medicine and Pharmacy*. 2013; 28(10): 2877-9.
- [107] Luo Ren, Cheng YB, Zhong XY. Study on correlation between therapeutic effect and ACE gene in diabetic nephropathy treated by “Minor Four and Five Decoction”. *Acta Universitatis Traditionis Medicalis Sinensis Pharmacologiaeque Shanghai*. 2001; 15(1): 24-6.

Chapter 46

BIOCHEMISTRY AND GENETICS OF ANSAMYCIN ANTIBIOTICS

Amit Kumar Jha and Jae Kyung Sohng*

Department of BT-Convergent Pharmaceutical Engineering,
Institute of Biomolecule Reconstruction, Sun Moon University,
Kalsan-ri, Tangjeong-myeon, Asan-si, Chungnam, Republic of Korea

ABSTRACT

Ansamycins are composed of secondary metabolites possessing a high degree of activity against numerous types of Gram-positive and Gram-negative bacteria. Structurally, ansamycins are characterized by the presence of core structures including an aromatic moiety (benzene or naphthalene derivative) and an aliphatic chain. Most ansamycins were isolated and characterized from Actinomycetes, while a few were mined in higher plants, for example, maytansine and colubrinol. Due to the development of microbiological techniques, genetic engineering and recombinant proteins, a wealth of different types of ansamycins have been mined along with their biosynthetic gene clusters. This resulted in developed strategies for enhancement of ansamycin production as well as synthesis of novel structure derivatives. In this chapter, we will describe the biochemistry and genetics of the most important members of ansamycin antibiotics and their applications as cytotoxic, anti-tumor, anti-parasitic and anti-bacterial agents. Thereafter, the future of ansamycins will be discussed to outline the most critically applied aspects in the last.

Keywords: biochemistry, ansamycin, genetics, antibiotics, *Streptomyces*

* Corresponding Author's Email: sohng@sunmoon.ac.kr (Fax: 82-41-544-2919; Tel: 82-41-530-2246).

INTRODUCTION

Since the 1950s, ansamycins, a prominent class of macrocyclic lactam antibiotics produced by various Actinomycetes have been investigated because of their wide range of biological activities. They are named ansamycins, from the Latin *ansa* (which stands for grip), because of their unique structure comprising of an aromatic moiety bridged by an aliphatic chain. Ansamycins contain an aromatic moiety (naphthalene or a benzene derivative) bridged by a polyketide chain (an aliphatic *ansa* chain) terminating in an amide linkage. The aromatic moiety involved in the biosynthesis of ansamycin antibiotics is derived from a 3-amino-5-hydroxybenzoic acid (AHBA) as a precursor of the mC₇N units, which are synthesized via a novel variant of the shikimate pathway. The branching was found to be modified by introduction of glutamine derived nitrogen at an early stage to give 3,4-dideoxy-4-amino-D-*arabino*-heptulosonic acid 7-phosphate (amino-DAHP) instead of the normal shikimate pathway intermediate, 3-deoxy-D-*arabino*-heptulosonic acid 7-phosphate (DAHP), followed by the formation of 5-amino-5-deoxy-3-dehydroshikimic acid (amino-DHS). Finally, DHS is aromatized by the enzyme AHBA synthase to AHBA (Figure 1 and 2).

AHBA appeared to be further activated by a nonribosomal peptide synthetase-like mechanism and enhanced via addition of methylmalonyl and malonyl extender units by a multimodular type 1 PKSs. On the basis of the folding and cyclization features of ansamycin polyketide, AHBA incorporation occurs in two different ways: either by a benzyl ring or a naphthyl ring with an aliphatic *ansa* chain. This reveals the formation of two main subclasses: the benzenoid ansamycins and the naphthalenoid ansamycins. Benzenoid and naphthalenoid ansamycins are further divided according to the differences in the length of their *ansa* chains. Thus, benzenoid ansamycins are divided into 2 groups, i.e., benzenoid ansamycins with C₁₅ *ansa* chains and benzenoid ansamycins with C₁₇ *ansa* chains. The naphthalenoid ansamycins are divided into 3 groups, i.e., naphthalenoid ansamycins with C₁₇ *ansa* chains, naphthalenoid ansamycins with C₂₃ *ansa* chains, and naphthalenoid ansamycins with C₉ *ansa* chains. Recently, new ansamycins were also reported from different species [1, 2, 3, 4].

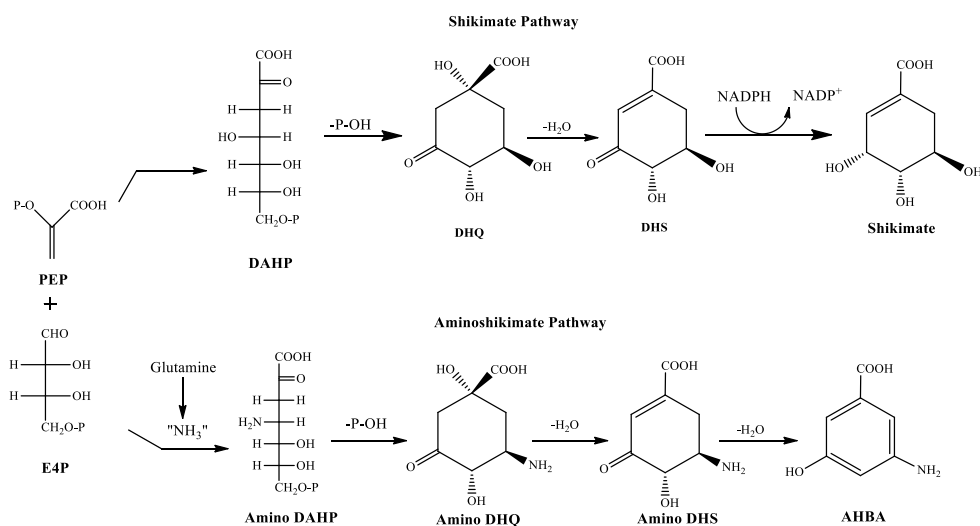


Figure 1. The shikimate pathway and the aminoshikimate pathway to AHBA.

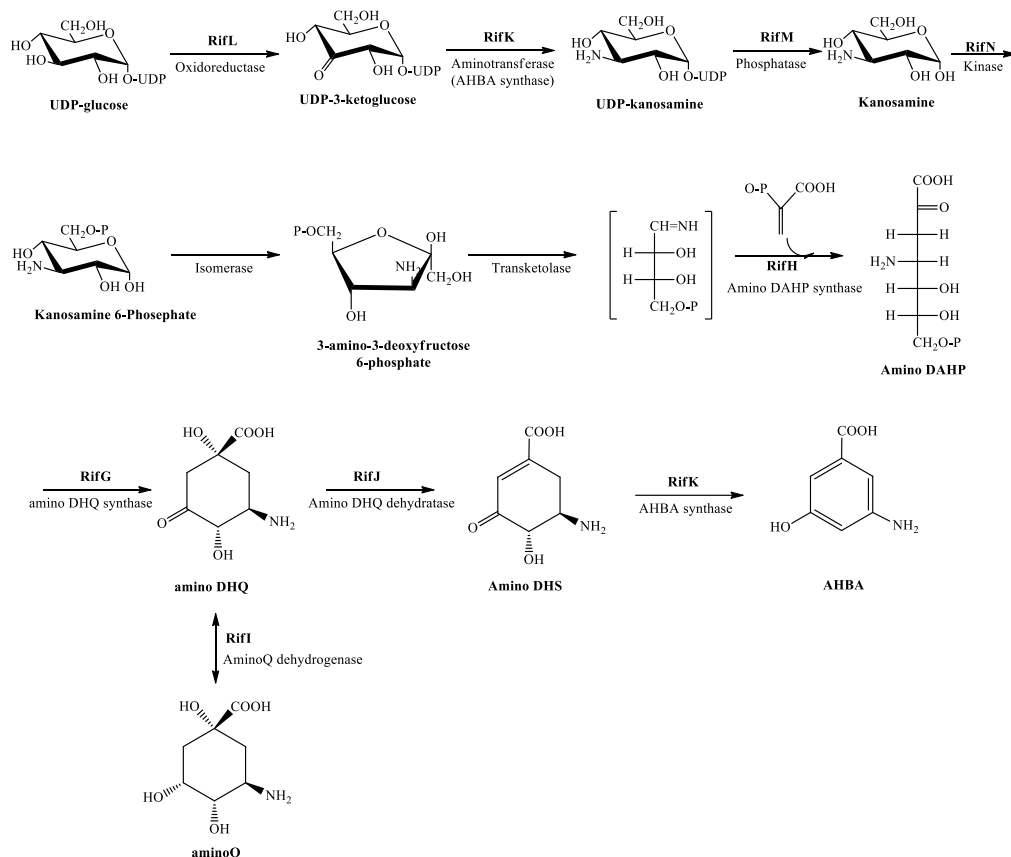


Figure 2. Biosynthesis of AHBA in *Amycolatopsis mediterranei* S699.

MAYTANSINOIDS

Maytansinoids are potent cytotoxic agents generally isolated from living organisms such as the higher plant families *Celastraceae*, *Rhamnaceae*, and *Euphorbiaceae*, as well as some mosses. They are 19-membered polyketide macrocyclic lactams belonging to the ansamycin family (Figure 3) [5, 6, 7]. Their molecular architecture is characterized by an aromatic chromophore with an aliphatic chain (ansa chain) connected back to a nonadjacent position through an amide linkage. Despite carrying either benzenic (ansamitocins, geldanamycin, etc) or naphthalenic (rifamycin, naphthomycin) chromophores, genetic and feeding experiments have revealed that all the ansamycins share the same polyketide starter unit, 3-amino-5-hydroxybenzoic acid (AHBA) [8]. Maytansinoids are under clinical trials with significant improvements achieved mainly by conjugation with tumor specific antibodies namely, trastuzumab emtansine (under Phase III), SAR3419 and BT062 (under Phase II), and several others such as BAY 94–9343, BIIB015, IMG529, lorvotuzumab mertansine, SAR566658, IMG529 [9]. The maytansinoid family comprises of several structural analogues such as maytansinol, maytansine, maytansinol 3-propionate, ansamitocin P3, and ansamitocin P4. They exhibit better cytotoxicity than several other known anticancer agents and had shown promising activity in B-16 melanocarcinoma murine solid tumors in addition to anti leukemic activity

against P388 murine lymphocytic leukemia [10, 11, 12]. Recent studies have shown that antibody-maytansinoid-conjugates block the cells at mitosis by suppressing microtubule dynamics similar to the results obtained involving the unconjugated maytansine [13, 14]. An actinomycete, *Actinosynnema pretiosum*, similarly displays the capacity to secrete large amounts of various maytansinoids known as ansamitocins [15].

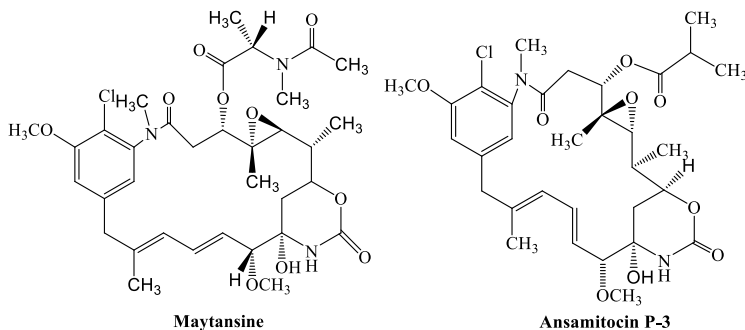


Figure 3. Structures of maytansinoids.

BENZENOID ANSAMYCINS

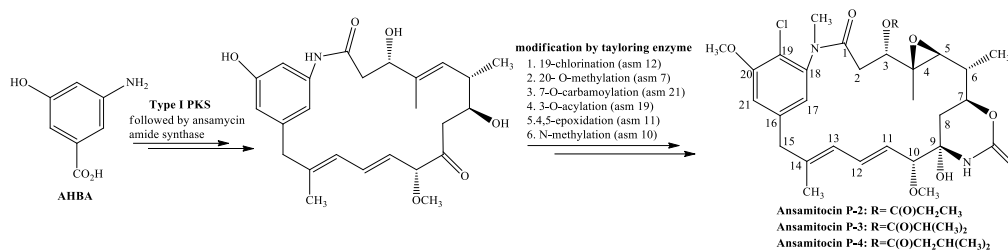
Benzenoid ansamycin antibiotics were first isolated in the late 1970s from the culture broths of several *actinomycetes* species and identified as inhibitors of eukaryotic Hsp90, an important cancer target. Considerable interest was generated by their unusual ansa bridge structure, and a number of compounds including ansamitocin, geldanamycin, herbimycin, macbecin, mycotrienin, cytotrienin, and trienomycin were identified.

ANSAMITOCIN

Ansamitocins, an ansamycin family of polyketide macrolactams with potent antitumor activities, were isolated from *Nocardia* sp. No. C-15003. Structurally, they were found to be similar to maytansine and related maytansinoids. Both the structures and antitumor activity of the ansamitocins are similar to those of the maytansinoids from plant sources [16]. Due to their high cytotoxicity, ansamitocins and homologous compounds have been used to conjugate with targeted molecules, such as monoclonal antibodies and folic acid for tumor-targeting therapy [17].

Ansamitocin P-3 is a secondary metabolite of *Actinosynnema pretiosum* ssp. Ansamitocins are a series of complex aromatic compounds including AP-0, AP-2, AP-3, AP-3', and AP-4, among which ansamitocin P-3 is found to be the most active ingredient [18, 19]. In recent years, Ansamitocin P-3 has received lot of attention because of its industrial importance. In the biosynthesis of ansamitocin P-3, 3-amino-5-hydroxybenzoic acid (AHBA) is used as a starter unit [20], and the incorporation of seven PKS extender units yields a 19-membered macrolactam, proansamitocin, which further undergoes a series of post-PKS modifications, including O- and N-methylation, chlorination, epoxidation, O-carbamoylation, and O-acylation for conversion into the bioactive compound, ansamitocin P-3 (Figure 4). Ansamitocin P-3 is

also known to depolymerize microtubules and bind to tubulin in a competitive manner with vinblastine and rhizoxin suggesting that it partially overlaps the vinblastine binding site [21]. Moreover, treatment of MCF-7 cells with ansamitocin P3 resulted in severe disruption of interphase and mitotic microtubules. The affected cells were found to be blocked from mitosis and accumulated p53 and its downstream partner p21 in the nucleus, which subsequently activated apoptotic cell death in the cells [22]. Recently, ansamitocin P-3 was identified as the most potent inducer of murine dendritic cell maturation, which resulted in augmentation of antitumor immunity in tumor-bearing hosts through the activation of *T* cells [23].



Beilstein J. Org. Chem. 2014, 10, 535-543.

Figure 4. Short representation of ansamitocin biosynthesis.

GELDANAMYCIN

Geldanamycin, a potent anticancer antibiotic, is a benzoquinone ansamacrolide related to herbimycin, reblastatin, and macbecin. It was first isolated in 1970 from *Streptomyces hygroscopicus* var. *geldanus* var. *nova*. Further, different geldanamycin analogs such as 4,5-dihydrothiazinogeldanamycin, reblastatin, 17-demethoxy-reblastatin, tetracyclic thiazinogeldanamycin and 19-hydroxy-4,5-dihydrogeldanamycin, were reported from the same strain (Figure 5) [24, 25]. Geldanamycin was found to be active against L-1210 and KB cells growing in culture and against the parasite *Syphacia ohlevata*, also shown to display antiviral activity [24, 26]. On the basis of its physical and chemical properties, geldanamycin is a complex molecule consisting of an unsaturated moiety attached to a quinone. The biosynthesis of geldanamycin involves the assembly of 3-amino-5-hydroxybenzoic acid as a starter unit, followed by the addition of extender units such as acetate, propionate, and glycolate to form a polyketide backbone, and further followed by the formation of geldanamycin via post-PKS modification, which includes C-17 hydroxylation, C-17 O-methylation, C-21 oxidation, C-7 carbamoylation, and C-4,5 oxidation. Geldanamycin binds to the N-terminal ATP binding pocket of heat shock protein 90 (Hsp90), which prevents stress-induced cellular damage, stabilizes various oncogenic kinases and influences gene expression e.g., by up-regulating NF-KB. As Hsp90 expression is particularly high in cancer cells and is associated with tumor cell progression, invasion and formation of metastases, as well as development of drug resistance, geldanamycin and its analogues have been considered for treatment of cancer in human cells. Their binding to Hsp90 disrupts the interaction of these proteins with the Hsp90 chaperone complex, preventing normal folding and leading to ubiquitylation and subsequent degradation. Second, geldanamycin treatment induces up-regulation of many proteins through the action of the transcription factor heat shock factor-1. These newly synthesized proteins then work to

restore cellular homeostasis after disruption of Hsp90 function [22]. Geldanamycin may also counteract neuronal injury, an effect attributed to destabilization of RIP1 protein. It was found that cisplatin (CDDP) blocks geldanamycin induced HSF-1-mediated transcription, resulting in decreased stress-responsive protein levels after treatment. The decreased transcription observed when geldanamycin is combined with CDDP is due to CDDP-mediated abrogation of HSF-1 chromatin binding, thereby preventing up-regulation of stress-responsive transcripts for genes such as Hsp70 and Hsp27 [27].

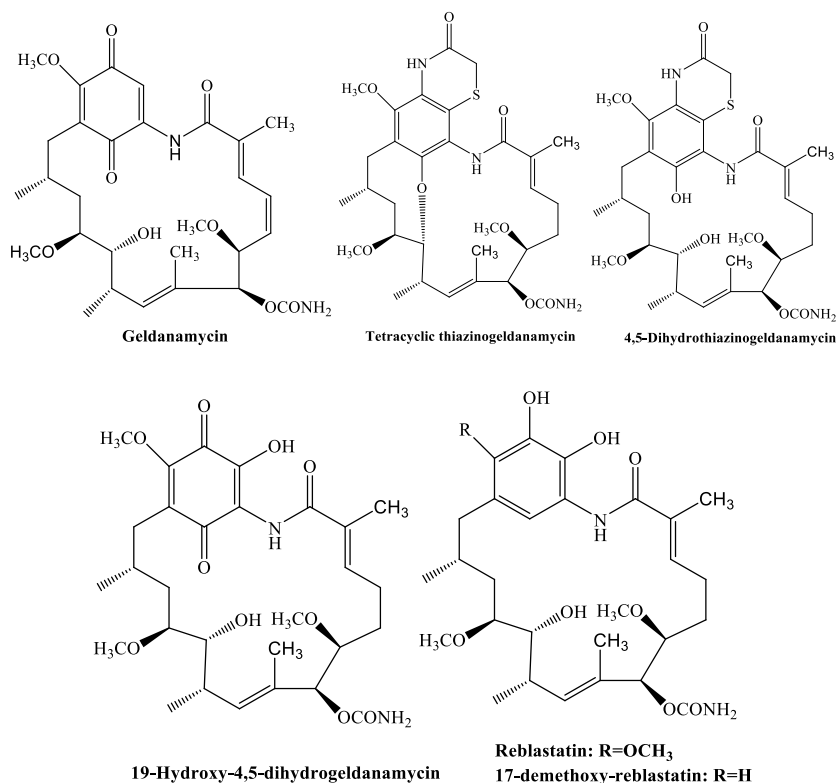


Figure 5. Structures of geldanamycin and its analogs from *Streptomyces hygroscopicus*.

So far, lots of geldanamycin derivatives have been reported and some of the less hepatotoxic derivatives entered into clinical trials, for instance, 17-allylamino-17-demethoxygeldanamycin (tanespimycin, 17-AAG) and 17-[2-(dimethylamino)ethyl]amino-17-demethoxygeldanamycin (alvespimycin, 17-DMAG) [28]. However, development of new generations of GA derivatives by structure modification has not yet been accomplished. 17-AAG showed some promise in phase II trials for the treatment of breast cancer. However, the clinical trial was terminated, indicating that geldanamycin derivatives with less hepatotoxicity are required. Recently, incorporation of phenylethylamine scaffold into the C-17 position of geldanamycin through application of structure-based bioisosterism was attempted to not only increase the binding of 17-phenylethylaminegeldanamycins to Hsp90, but also to decrease the hepatotoxicity [29]. SNX-25a, with a 2-(trans-4-hydroxy-cyclohexylamino) side chain, is a novel 2-aminobenzamide inhibitor of Hsp90 optimized through structure-activity relationship (SAR) explorations for high Hsp90 affinity and was tested against several human cancer cells

then compared with 17-AAG. Initially, the growth inhibitory effects of SNX-25a and 17-AAG on human cancer lines were investigated and demonstrated that SNX-25a displayed better activity than 17-AAG in human cancer cells. This may be mediated by G2 phase arrest and cell apoptosis; and provoked by the down-regulation of Hsp90 client proteins. In addition, SNX-25a, an Hsp90 inhibitor, displayed higher affinity for Hsp90 than the typical Hsp90 inhibitor 17-AAG in molecular docking. [30].

MACBECIN

In 1979, while screening for new antibiotics, the benzenoid ansamycins macbecins I and II (Figure 6) were detected after fermentation of the actinomycete strain No. C-14919 (N-2001). Macbecins I and II were antifungal and antiprotozoal, whereas macbecins I and II do not show any specific activity against eukaryotic microorganisms. Hence the biological activities of macbecins I and II are distinguishable from ansamitocins and other related compounds such as maytansine [31]. Macbecin II, a heat shock protein (Hsp90) inhibitor, is a DNA antimetabolite that induces breaks in double-stranded DNA, possibly by means of p53-mediated apoptosis. Although Macbecin II functions as a Hsp90 inhibitor, it is unclear whether this is its mechanism of action for increased activity in SMAD4-negative cells, as geldanamycin does not display similar activity. On the other hand, reduction in TSC2 (siRNAs) has been shown to activate mammalian target of rapamycin mediated cellular hyperplasia and hypertrophy. Cells with reduced levels of TSC2 are less susceptible to macbecin II. In some systems, loss of mammalian target of rapamycin activity is associated with p53-dependent apoptosis. Hence, enhancement of cellular growth through loss of TSC2 may help overcome the apoptotic effects of DNA damage. Furthermore, TSC2 knockdown affected the activity of macbecin II but not cisplatin, further implicating distinct p53-related cellular responses [32].

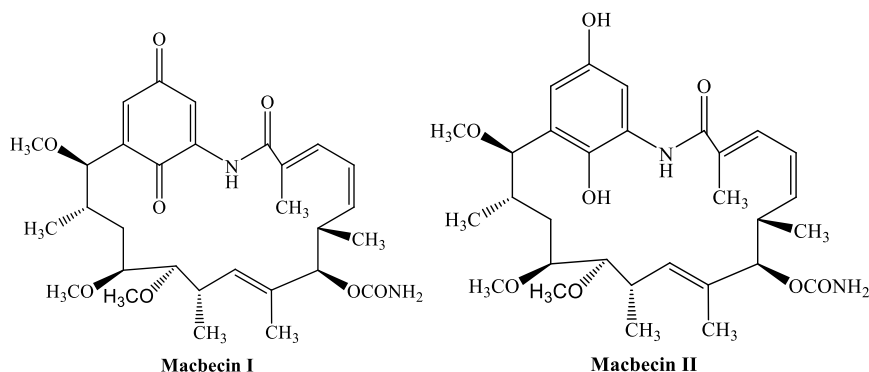


Figure 6. Structures of macbecin I and II.

HERBIMYCIN

In 1978, herbimycin was isolated from the fermentation broth of *Streptomyces hygroscopicus* strain No. AM-3672. In 1980, herbimycin B (Figure 7) was isolated from the

same strain, as was herbimycin C in 1986 (Figure 7). Herbimycin was found to have potent herbicidal activity against most mono- and di-cotyledonous plants, especially against *Cyperus microiria* STEUD. However, *Oryza sativa* showed strong resistance to herbimycin where as Herbimycin B was less effective. Moreover, herbimycin B showed potent anti-tobacco mosaic virus activity. Herbimycin A (Figure 7) may reduce hepatic ischemia-reperfusion injury through the protective effect of hepatocyte against radical injury, but not by preventing radical induced lipid peroxidation and oxidative stress. Herbimycin A did not inhibit lipid peroxidation or oxidative stress in reperfused liver tissue, but reduced liver damage and hepatocyte injury. It also appears to cause the heat shock transcription factor-1 (HSF-1) to bind the heat shock element, leading to transcriptional activation of heat shock protein (HSP) genes and providing protection against insult at nontoxic concentrations. It has been shown to induce transcription of HSP70, HSP90, HSP25, and HSP30 in a number of cell types.

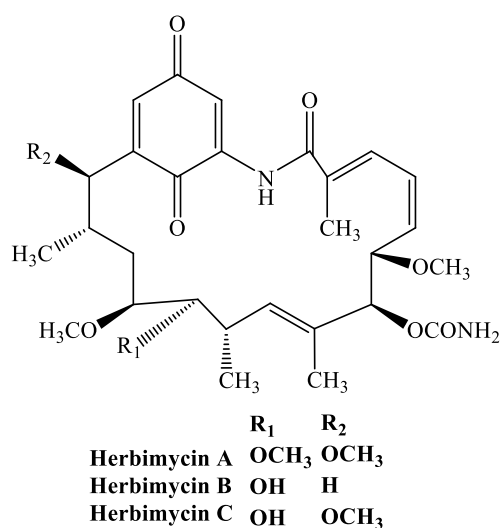


Figure 7. Structures of herbimycin A, B and C.

On the other hand, herbimycin A has been reported to inhibit the induction of apoptosis, which is mediated by intracellular signaling. If radical-induced cell injury is also mediated by intracellular signaling, it may be that herbimycin A, a potent tyrosine kinase inhibitor, inhibits the signaling via tyrosine phosphorylation. Because herbimycin A reduces hepatic reperfusion injury through a different mechanism from antioxidants, synergistic protective effects may be achieved by the addition of herbimycin A into preservation solutions. Herbimycin A also has protective effects on radical-induced injury in human vascular endothelial cells. Furthermore, herbimycin A promotes the degradation of transmembrane tyrosine kinase receptors by the 20S proteasome [33]. Based on the discovery of gonad-specific expression of the cellular Src tyrosine kinase SmTK3, inhibition experiments with the Src-kinase inhibitor herbimycin A led to reduced mitotic activity and egg production in paired *S. mansoni* females. The comparison of both inhibitor treatments pointed to a stronger reduction of both parameters following herbimycin A treatment. The strongest influence on mitotic activity and egg production was observed when both inhibitors were combined. Herbimycin A is an inhibitor of viral Src and c-Src activity, therefore, its effect on Hepatitis C virus replication is indeed mainly due to the

inhibition of c- Src function. This is further supported by the observation that down-regulation of c-Src expression by siRNA is accompanied by a reduction in the inhibitory concentration (IC₅₀) of herbimycin A, which is commensurate to the reduction of c-Src protein levels [34, 35].

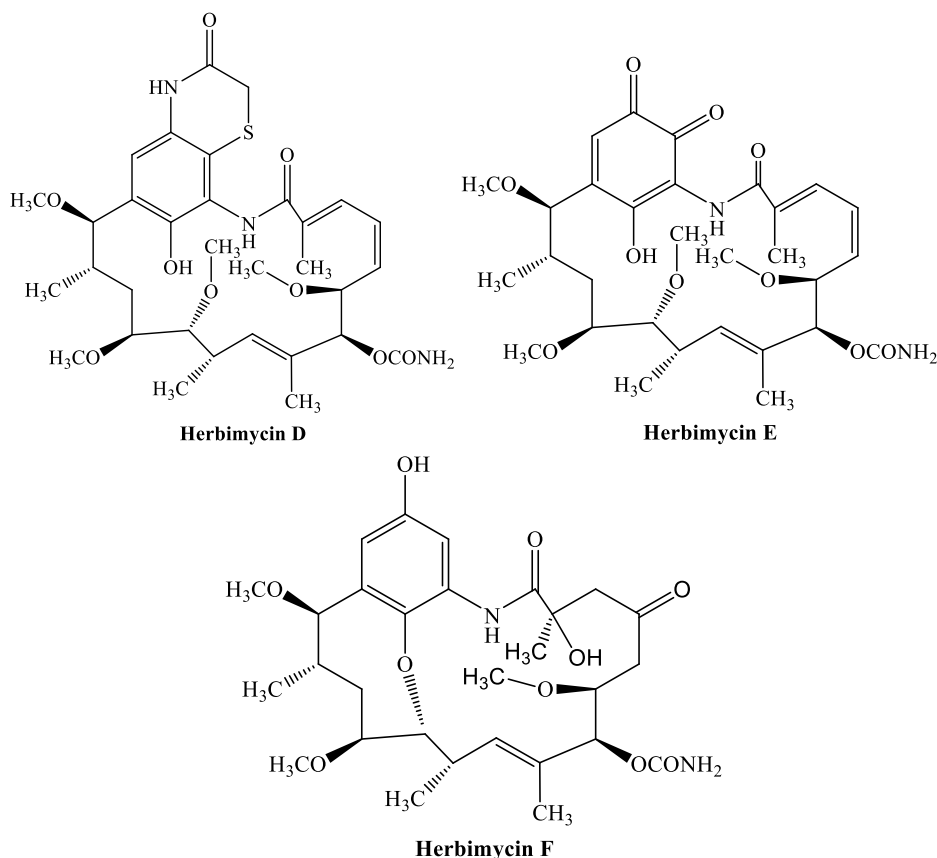


Figure 8. Structure of Herbimycin D, E and F.

Recently, three new ansamycin analogues, herbimycins D-F (Figure 8), were isolated from *Streptomyces* sp. RM-7-15. Herbimycin D was isolated as a pale yellow solid material composed of mercaptoacetamide-bearing ansamycins. Herbimycin E was isolated as a purple solid and its NMR chemical shifts were in agreement with those of the fully characterized *ortho*-quinone natural product 7-hydroxy-8-methoxymalbranicin-5,6-quinone. Similarly, herbimycin F was isolated as a white solid. Less than 25 bacterial *ortho*-quinones metabolites have been reported thus far. Within this context, the herbimycin E reported herein represents the first example of an *ortho*-quinone ansamycin. Similarly, unusual intramolecular etheric linkage was observed in herbimycin F. Unlike herbimycin A, which was equipotent in cancer cell line cytotoxicity assays, the new analogues (herbimycins D-F) did not exhibit any cytotoxicity or antibacterial and antifungal activities with similar affinity. Rather, these compounds were found to bind the Hsp90 α N-terminal domain with similar affinity to that of herbimycin A [36].

MYCOTRIENINS (ANSATRIENINS), TRIERIXIN AND QUINOTRIERIXIN

Mycotrienins, a unique class of benzoquinone ansamycins antibiotics, are structurally related to herbimycin A in that they also contain a quinone/hydroquinone system and are produced by *Streptomyces rishiriensis* T-23. The reduced form was identified, via comparison with the previously reported mycotrienin, as an antifungal antibiotic. A different form was characterized as its oxido- form, hence, the name mycotrienin I was given to the latter and mycotrienin II to the former. Ansatrienins A and B were also isolated from *Streptomyces collinus* and *Streptomyces rishiriensis* with different alanine stereochemistry in comparison to mycotrienin I and II (Figure 9). The isolation of two minor components, ansatrienins A2 and A3 were also reported, wherein the cyclohexyl group of ansatrienin A is replaced by a 2-methylbutyryl group and an isovaleryl group, respectively. Further screening for minor components of mycotrienins resulted in the isolation of two metabolites named 22-O-methylmycotrienin II and 19-deoxymycotrienin II. Mycotrienin I and mycotrienin II are active against fungi and yeasts, but inactive against bacteria [37, 38]. Mycotrienins are a possible target for pp60^{c-src} kinase in fetal rat long bones. The structural relations of different mycotrienin analogues that inhibit bone resorption closely reflect the ability of these compounds to inhibit pp60^{c-src}. Hence, pp60^{c-src} is essential for normal osteoclastic bone resorption and point to a potential for pharmacologic intervention in the bone resorption process at the level of pp60^{c-src}. Therefore, the mycotrienins may have some therapeutic potential as bone resorption inhibitors in diseases where bone resorption is increased [39]. Similarly, mycotrienin II, inhibited the cell-surface intercellular adhesion molecule-1 expression induced by TNF- α more strongly than that induced by IL-1 α in human lung carcinoma A549 cells. Mycotrienin II was found to inhibit protein synthesis in intact living cells, as well as in cell-free translation systems. Several compounds of the triene-ansamycin group (that is, mycotrienin I, trienomycin A, trierixin, quinotrierixin and quinotrierixin HQ) also inhibited intercellular adhesion molecule-1 expression, as well as cell-free translation in a manner similar to mycotrienin II, namely thru direct inhibition of translation. This limits inhibition of intercellular adhesion molecule-1 expression induced by pro-inflammatory cytokines [40].

Trierixin, a new member of the triene ansamycin group has also been isolated from *Streptomyces* sp. AC 654. Trierixin inhibits thapsigargin-induced XBP1-luciferase activation in HeLa/XBP1-luc cells and endogenous XBP1 splicing in HeLa cells. In the process of isolating trierixin, we isolated structurally related mycotrienin II and trienomycin A, both inhibitors of ER stress-induced XBP1 activation, from a culture broth of a trierixin-producing strain. Trierixin possessed a 21-membered macrocyclic lactam (Figure 10), which contains a methylthio-benzenediol structure and a cyclohexanecarbonylalanine moiety, and was thus labeled as 21-thiomethylmycotrienin II [41, 42].

Quinotrierixin, demethyltrienomycin A, demethyltrienomycin B and demethyltrienomycinol were isolated from a culture broth of *Streptomyces* sp. PAE37 and found to be inhibitors of ER stress-induced X-box binding protein 1 activation. All four possessed 21-membered macrocyclic lactams including triene moieties. Quinotrierixin inhibited thapsigargin-induced X-box binding protein 1 activation in HeLa cells, whereas demethyltrienomycin A and mycotrienin I inhibited ER stress-induced X-box binding protein 1 activation. A SAR study showed that the OH group at C-13 was crucial and that the CH₃ group at C-14 was important for the X-box binding protein 1 inhibitory activity [43, 44].

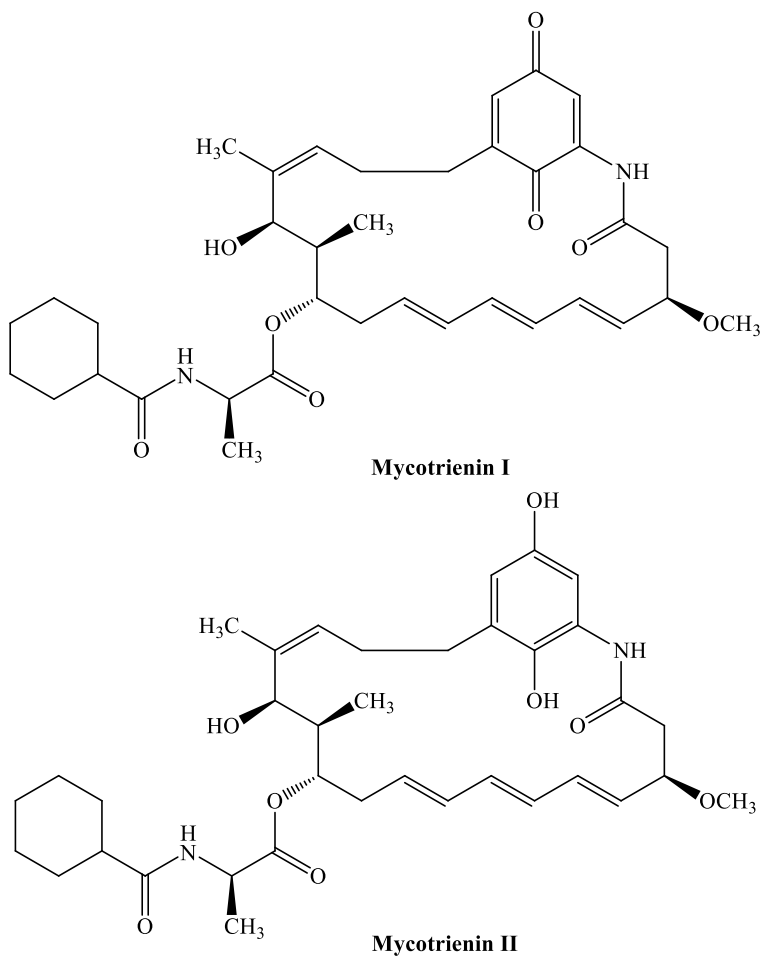


Figure 9. Structure of Mycotrienin I and II.

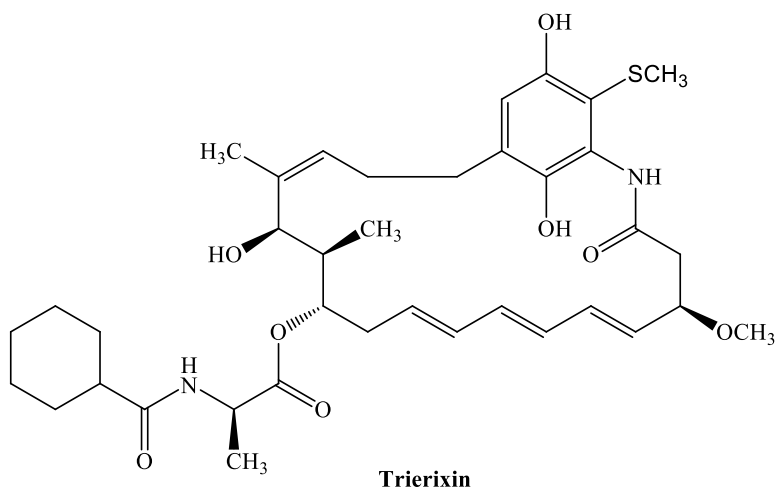


Figure 10. Structure of trierixin.

CYTOTRIENIN

Cytotrienin A (Figure 1A), 21-membered cyclic lactam of the triene-ansamycin family (Figure 11), was initially identified as an inducer of apoptosis in human leukemia HL-60 cells. Isolated from *Streptomyces* sp. RK95-74, cytotrienin A contains a trieneansamycin, 1-aminocyclopropane carboxylic acid and 1-cyclohexene-1-carboxylic acid moieties [45]. Cytotrienin A, as an apoptosis inducer, marks morphological changes of the cell including membrane blebbing, cell shrinkage, chromatin condensation, DNA cleavage, and fragmentation of the cell into apoptotic bodies. It also activates signaling pathways of JNK/SAPK and p38 MAPK. The p36 MBP kinase activated by cytotrienin A phosphorylates MBP to a greater extent than c-Jun and ATF-2. It was also reported that MST1/Krs2 is cleaved by caspase-3-like activity during Fas-induced apoptosis. It is possible that most apoptotic signals including the Fas-mediated signal and the cytotrienin-induced signal activate several pathways for apoptosis in target cells. The caspase-mediated proteolytic activation of MST/Krs proteins in cytotrienin A-sensitive and -resistant human tumor cell lines were further investigated. Cytotrienin A was also found to inhibit eukaryotic protein synthesis by targeting translation elongation and interfering with eukaryotic elongation factor 1A function. The molecule prevents HUVEC tube formation and reduces micro vessel formation in chorioallantoic membrane assays. Furthermore, it was found to inhibit the cell surface expression of intercellular adhesion molecule-1 (ICAM-1) induced by tumor necrosis factor (TNF)- α more strongly than the expression induced by interleukin (IL)-1 α . Hence, cytotrienin A is a translation inhibitor that induces the TACE-mediated ectodomain shedding of TNF receptor 1 via the activation of ERK and p38 MAP kinase [46].

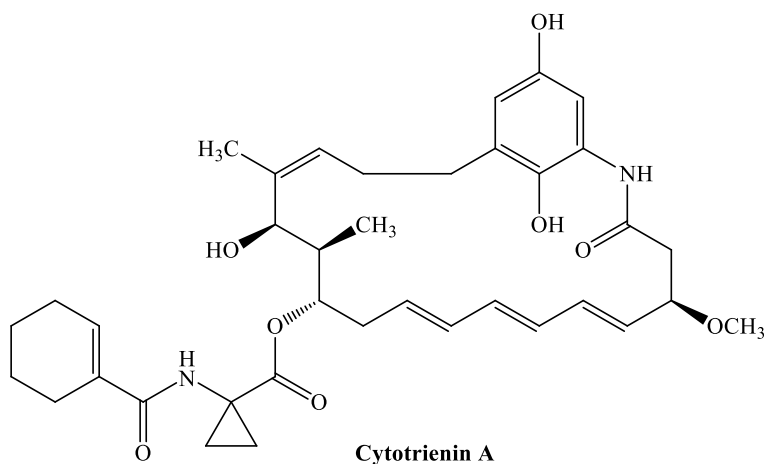


Figure 11. Structure of cytotrienin A.

TRIENOMYCIN

An antibiotic with triene-containing C17-benzene ansamycins known as trienomycins A (Figure 12) was isolated from the culture broth of *Streptomyces* sp. No. 83-16 in 1985 and from *Streptomyces* sp. G91614 in 2002. Isolation of trienomycins A was followed by the isolation

of a series of trienomycins which include trienomycins B-G. The trienomycin group is closely related to mycotrienins and ansatrienins in structure. However, trienomycins are unique among the benzenoid ansamycin antibiotics in that trienomycins A - E do not contain a p-quinone or p-hydroquinone moiety. The benzenoid moiety of the trienomycin group is rather similar to those of maytansinoids (higher plant origin) and ansamitocins (microbial origin). Similar to mycotrienins I and II and ansatrienins A and B, the hexahydrobenzoyl moiety was found in the structure of trienomycin A, whereas the 3-methylbutyryl (isovaleryl) and 2-methylbutyryl moieties were found in trienomycins B and C respectively. Instead of the hexahydrobenzoyl moiety of trienomycin A, a tetrahydrobenzoyl moiety is attached to the alanyl moiety in trienomycin D. In contrast, trienomycin E is quite similar to trienomycin B except for the acyl group attached to the alanyl group, which indicates that the 4-methylpentyl moiety is attached to the alanyl moiety of trienomycin E [47, 48].

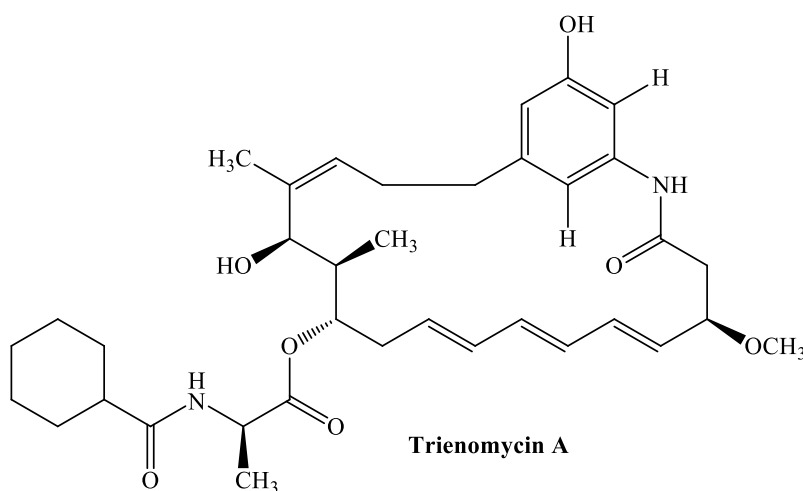


Figure 12. Structure of Trienomycin A.

In the year 2002, trienomycin G (Figure 13) along with trienomycin A, were reported to be isolated from *Streptomyces* sp. G91614. As in trienomycin A, the N-hexahydrobenzoyl alanine moiety of trienomycin G was found to be linked to C-13 instead of C-11. Thus, trienomycin G is a structural isomer of trienomycin A, only differing in the linkage position of the N-hexahydrobenzoyl alanine side chain to the ansa moiety. Trienomycins A-E exhibited considerable cytotoxic activity against HeLa S3 cells without showing activity against the microorganisms tested including bacteria, fungi and yeast. Trienomycins D and E exhibited 2 ~ 5-fold weaker activities against HeLa S3 cells than trienomycin A. On the other hand, trienomycin G and trienomycin A displayed potent inhibitory activity against Nitric oxide production in BV-2 microglia cells stimulated with LPS. Trienomycin G inhibited Nitric oxide production more strongly than that of trienomycin A. This result indicates that the position of the N-hexahydrobenzoyl alanine moiety affects the activity [49].

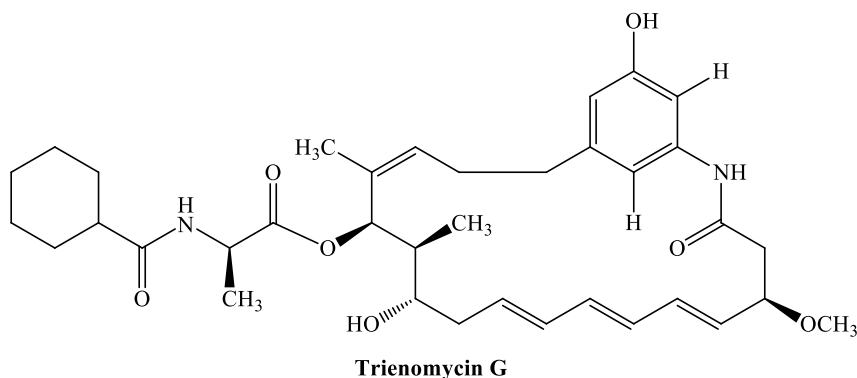


Figure 13. Structure of Trienomycin G.

NAPHTHALENOID ANSAMYCINS

Naphthalenoid ansamycins are best known for their antimicrobial activities, which are mediated by a specific inhibition of bacterial RNA polymerase. They are used for the treatment of leprosy, tuberculosis, and AIDS-related mycobacterial infections. Naphthalenoid ansamycins include salinisporamycin, hygrocins, rifamycin drugs as well as the rifamycin derivatives rifampicin (or rifampin), rifabutin, and rifapentine.

RIFAMYCIN

Rifamycin, a naphthalenic ansamycin antibiotic was isolated from two strains of *A. mediterranei*, S699 and LBGA 3136, and the major products of the complex were named rifamycin A, B, C, D and E (Figure 14). The rifamycin polyketide synthase (*rif* PKS) gene clusters were essentially identical in both of these strains. The partially characterized (~4 kb) *rif* PKS-like gene cluster from *A. rifamycinica* DSM 46095 showed 10% differences in nucleotide sequences compared to the *rif* PKS clusters from S699 and LBGA 3136.

The *rif* PKS gene cluster in strain 46095 was represented by a continuous fragment of ~85 kb that contained *rifA*, *rifB*, *rifC*, *rifD*, and *rifE* open reading frames (ORFs). The acyltransferase (AT) and ketosynthase (KS) domains present on these ORFs showed 90% homology to those on *A. mediterranei* S699 (4). Apart from a unique *rif* PKS cluster, eight other PKS, seven NRPS, and three hybrid NRPS/PKS clusters were present in the genome. The sequence information of a novel rifamycin biosynthetic gene cluster of *A. rifamycinica* 46095 will be helpful in the analysis of the structure of the polyketide produced by this strain [50]. Rifamycin B related compounds, such as rifamycin O and rifamycin S, were obtained by the chemical transformation of rifamycin B. Rifamycin L, rifamycin SV and rifamycin Y were isolated from the cultured broth of *Streptomyces mediterranei*. Rifamycin derivatives have also shown excellent activity against opportunistic pathogens in the AIDS complex of diseases, as well as against HIV and related oncogenic viruses due to the inhibition of the viral reverse transcriptase.

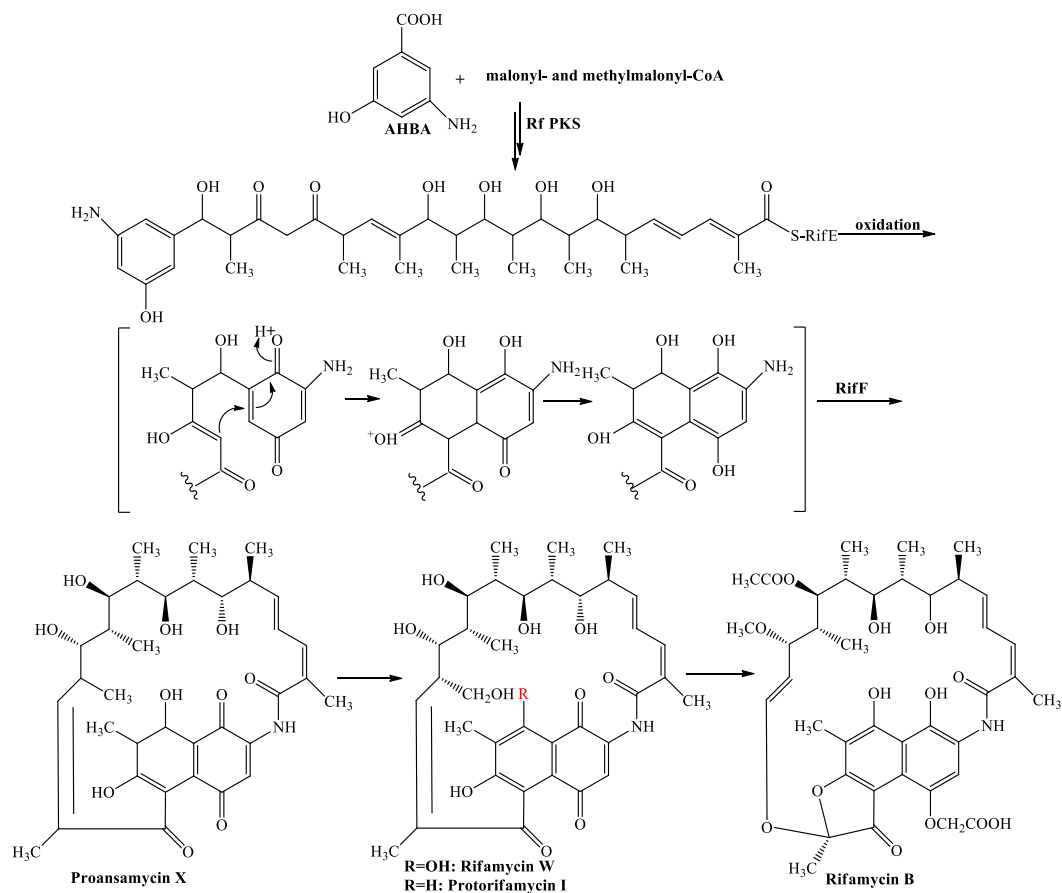


Figure 14. Pathway of Rifamycin B biosynthesis.

RIFAMPICIN

Since 1960, Rifampicin, a synthetically modified form of rifamycin (Figure 15), was used as the first-line therapy drug in the treatment of tuberculosis and other mycobacterial infections with low toxicity. Rifampicin prevents the assembly of DNA-dependent RNA polymerases, thereby inhibiting protein synthesis and cell growth in bacteria while preventing the assembly of DNA and protein into mature virus particles. The block, which occurs at the stage of the viral envelope formation, can be rapidly reversed by removal of the drug. It also inhibits the growth of poxviruses and adenovirus and can prevent the maturation of infectious viruses until quite late in the growth cycle. Rifampicin reaches maximal serum concentration in 1–4 h after application and its plasma half-time is 2–5 h. The lipophilic ansa chain is mainly responsible for the transport of the drug across the blood-brain barrier into the brain parenchyma [1, 51]. However, the emergence and spread of multidrug-resistant tuberculosis (MDR-TB) and extensively drug-resistant tuberculosis (XDR-TB) pose a major public health problem threat worldwide. According to the World Health Organization, 450,000 new cases of MDR-TB were documented worldwide in 2012. Thus, a sensitive and specific diagnostic tool is required to initiate the appropriate therapy and reduce the spread of multidrug-resistant *M. tuberculosis*.

strains. The emergence of resistance during therapy with rifamycin can generally be avoided with the use of adequate combination therapy. The clinically significant resistance mechanism is mutation in the 81-bp region of the *rpoB* gene, which encodes the target of rifamycin, the β subunit of bacterial DNA-dependent RNA polymerase. The portion of this sequence defined as cluster I is particularly important for high-level resistance. As a result of the high degree of conservation of DNA-dependent RNA polymerase, including in this β region, the mutations that determine resistance are also conserved across species. Mutations are most often found in three specific codons of the consensus sequence and Ser531Leu substitution in β predominates in mycobacteria and some other species. Cross-resistance appears to be complete among the rifamycins currently used to treat mycobacterial diseases. Mutations can have profound effects on transcription rate, initiation, pausing and termination. As a result of the central role played by DNA-dependent RNA polymerase in the bacterial cell, mutations often affect a number of physiological processes. On the other hand, better analogs of rifamycin can be generated to overcome the emergence of rifamycin resistance such as with rifampicin-resistant *M. tuberculosis* strains, OSDD 321 and OSDD 206, which contain the S531L mutation, and OSDD 55, which has the H526T mutation in the DNA-dependent RNA polymerase β subunit. These mutations appear to alter the affinity of DNA-dependent RNA polymerase to rifampicin. In addition, drug resistant mutations can disrupt hydrogen bonding in the polyketide ansa chain which was confirmed by the fact that 24-desmethylrifampicin showed improved activity against rifampicin resistant *M. tuberculosis*. The loss of the methyl group in this compound is postulated to lead to conformational changes in the ansa chain that allow for more flexibility of the compound to bind mutated DNA-dependent RNA polymerase[52, 53].

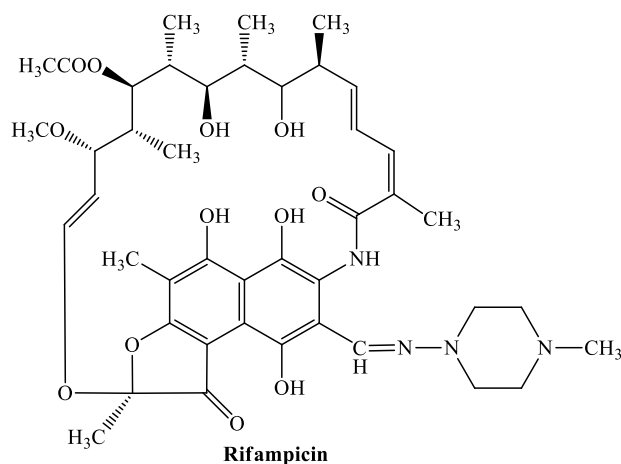


Figure 15. Structure of Rifampicin.

Recently, it was demonstrated that rifampicin inhibits the lipopolysaccharide -stimulated expression of Toll- like receptor-4. When cortical neurons were co-cultured with lipopolysaccharide -stimulated BV2 microglia, pre-treatment with rifampicin increased neuronal viability and reduced the number of apoptotic cells. Hence, rifampicin exerts strong brain protective effects in various in vivo and in vitro models of neurodegenerative diseases including Stroke, Parkinson, and Alzheimer. Pilot clinical studies indicate that patients with neurodegenerative diseases may benefit from rifampicin treatment. These promising findings

should be further explored in larger preclinical therapeutic studies and even further in randomized early phase clinical trials. Similarly, rifamycin SV is derived from rifamycin B by removal of the glycolic group bound to C-4 and has clinical uses in the treatment of bacterial infections due to binding and inhibition of bacterial DNA-dependent RNA polymerase, which is also able to bind the BCL6-POZ domain and inhibits BCL6 transcriptional repression. Hence, rifamycin SV can inhibit amyloid fibril formation through disruption of interactions between fibril aromatic rings required for elongation. Rifaximin does not have this effect [54, 55, 56].

RIFABUTIN

Since 1981, rifabutin, a spiro-piperidyl derivative of the parent compound rifamycin S (Figure 16), has been used for the treatment of disseminated *Mycobacterium avium intracellulare* infection in patients with AIDS. It was found to have significant activity in animal models against both *Mycobacterium tuberculosis* and *Mycobacterium leprae* compared to rifampin and studies in 1987 found that rifabutin is active against some rifampin-resistant strains of both species. Its significance for the treatment of *M. tuberculosis* was confirmed in 1996 and at lower doses than rifampicin. The drug has a long half-life (16 hr) in humans and a marked tissue tropism, with tissue levels five- to 10-fold higher than in serum. In animals, rifabutin was no more toxic than rifampin. In some cases, rifabutin may retain its activity against isolates resistant to rifampicin, possibly due to differences in affinity for mycobacterial RNA polymerase or additional inhibition of DNA biosynthesis [57, 58].

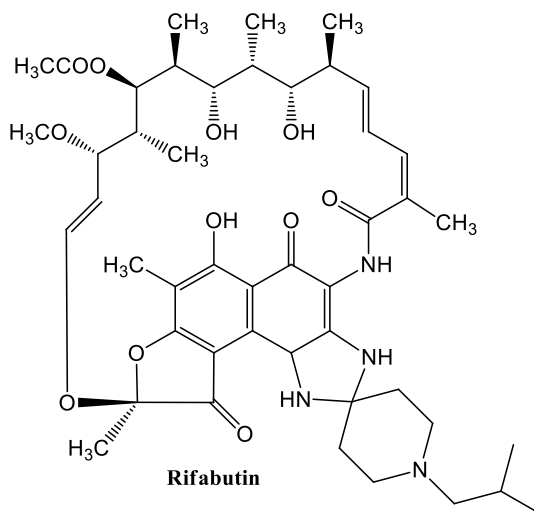


Figure 16. Structure of Rifabutin.

Rifabutin displays good in vitro activity against *Helicobacter pylori*, and the prevalence rate of rifabutin resistance is very low at only about 1%. Rifabutin should be limited to patients where previous (multiple) eradication regimens with key antibiotics such as amoxicillin, clarithromycin, metronidazole, tetracycline and levofloxacin have failed. Thus, it is suggested that the position of rifabutin in the algorithm of *Helicobacter pylori* treatment is as a fourth-line rescue regimen; that is, administered only after failure of three eradication regimens [59].

Rifabutin has long post antibiotic effect against *M. tuberculosis* and *M. avium* Complex, shows extensive distribution in various tissues, and readily penetrates cell membranes of leucocytes. These characteristics and their variations among patients can considerably influence the outcome of rifabutin-containing anti-mycobacterial therapy and therefore might be one of the explanations of favorable efficacy despite lower plasma concentrations of rifabutin. Low isoniazid plasma concentration is independently related to treatment failure of HIV/TB co-infection [60].

The American Thoracic Society guidelines classify rifabutin as a first-line anti-TB drug. In addition the WHO guidelines recently added rifabutin to the list of group 1 drugs, the most potent and best-tolerated anti-TB agents. Rifampicin is an essential component of TB treatment and has been referred to as the most important anti-TB agent because of its excellent sterilizing capacity. Considering that rifabutin is more active than rifampicin against slow-growing mycobacteria, including *Mycobacterium tuberculosis*, and that rifabutin has a similar potency to rifampicin in the treatment of drug susceptible TB, the promising outcome in the rifabutin group is most likely related to the outstanding potency of rifabutin against *M. tuberculosis*. Minimum inhibitory concentration distributions for rifabutin were above the epidemiological cut-off, but below the standard critical concentration. The WHO Guidelines for the treatment of HIV-associated tuberculosis recommend that treatment is given daily throughout the intensive and continuation phases of antituberculosis treatment. A daily dose of rifabutin is in line with these recommendations. This would facilitate the important programmatic issue of combining rifabutin with other antituberculosis medications as a fixed-dose combination pill to be taken on a daily basis [61, 62, 63].

RIFAPENTINE

In 1998, rifapentine, a potent antimycobacterial rifamycin (Figure 17) antibiotic was approved by the Food and Drug Administration as an alternative to rifampin in treatment regimens for patients with tuberculosis. Rifapentine is under active investigation as a potent drug that may help shorten the duration of tuberculosis treatment. There have been many reports on the antimicrobial activity of rifapentine in vitro, in macrophages, and in mice. It was found to be 2–4 times as active as rifampicin against *Mycobacterium tuberculosis*. The serum half-life of rifapentine is several times longer and it has greater bactericidal dosing than rifampicin. It has been approved as a first-line drug for once or twice-weekly dosing in the treatment of tuberculosis.

Rifapentine is being evaluated as an agent to shorten the duration of tuberculosis treatment. Applying a nonlinear mixed-effect modeling approach to rich population pharmacokinetic data collected from individuals receiving daily doses of rifapentine ranging from 450 mg to 1,800 mg revealed that the bioavailability of rifapentine decreases with increasing doses beginning at the lowest dose tested. Rifapentine pharmacokinetics is time dependent, but the magnitude of auto induction does not vary with concentration. Clinical trial simulations demonstrate that splitting the dose or taking the dose with food could substantially increase daily rifapentine exposures, which is a significant finding; given that rifapentine activity appears to correlate best with the area under the concentration time ratio and that current dosing achieves concentrations that are still on the steep part of the dose-response curve. With tuberculosis still

globally out of control, it is necessary to undertake studies that allow us to maximize the benefits from the drugs and drug combinations that are available for treatment [64, 65].

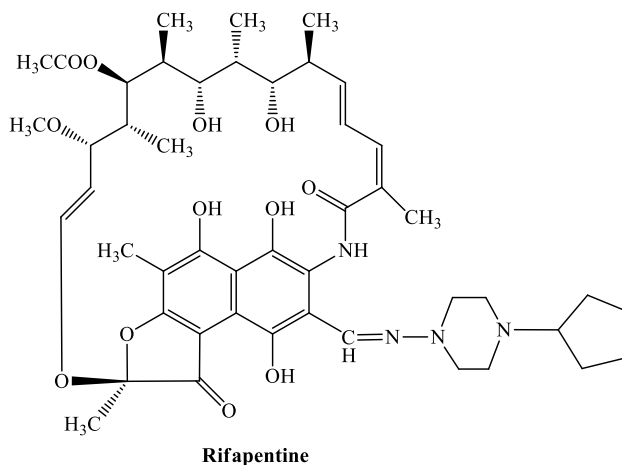


Figure 17. Structure of Rifapentine.

SALINISPORAMYCIN

In 2009, salinisporamycin, a new rifamycin antibiotic (Figure 18) was isolated from a culture broth of marine actinomycete identified as *Salinispora arenicola* on the basis of the 16S rRNA sequence. Subsequent biological studies revealed that salinisporamycin not only showed antimicrobial activity, but also inhibited the growth of A549 cells (the human lung adenocarcinoma cell line) [66].

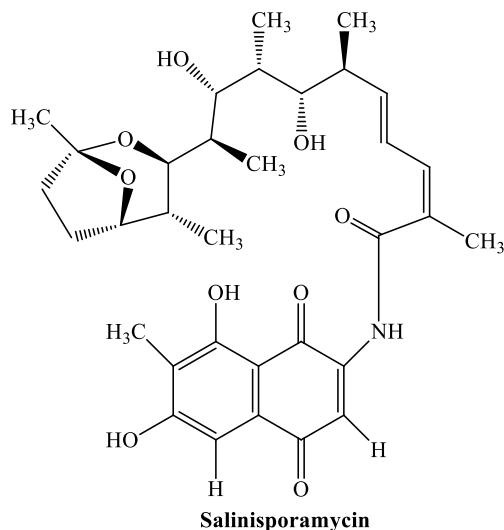


Figure 18. Structure of Salinisporamycin.

HYGROCIN

Streptomyces sp. LZ35 is a 3-amino-5-hydroxy-benzoic acid (AHBA) containing strain with two AHBA synthases. It was isolated from intertidal soil collected in Jimei, Xiamen, P. R. China. This strain can produce two different types of ansamycins: geldanamycin (benzenic ansamycin) and a hygrocins (naphthalenic ansamycin) (Figure 19). Hence, AHBA is common in the biosynthesis of naphthalenic and benzenic ansamycins.

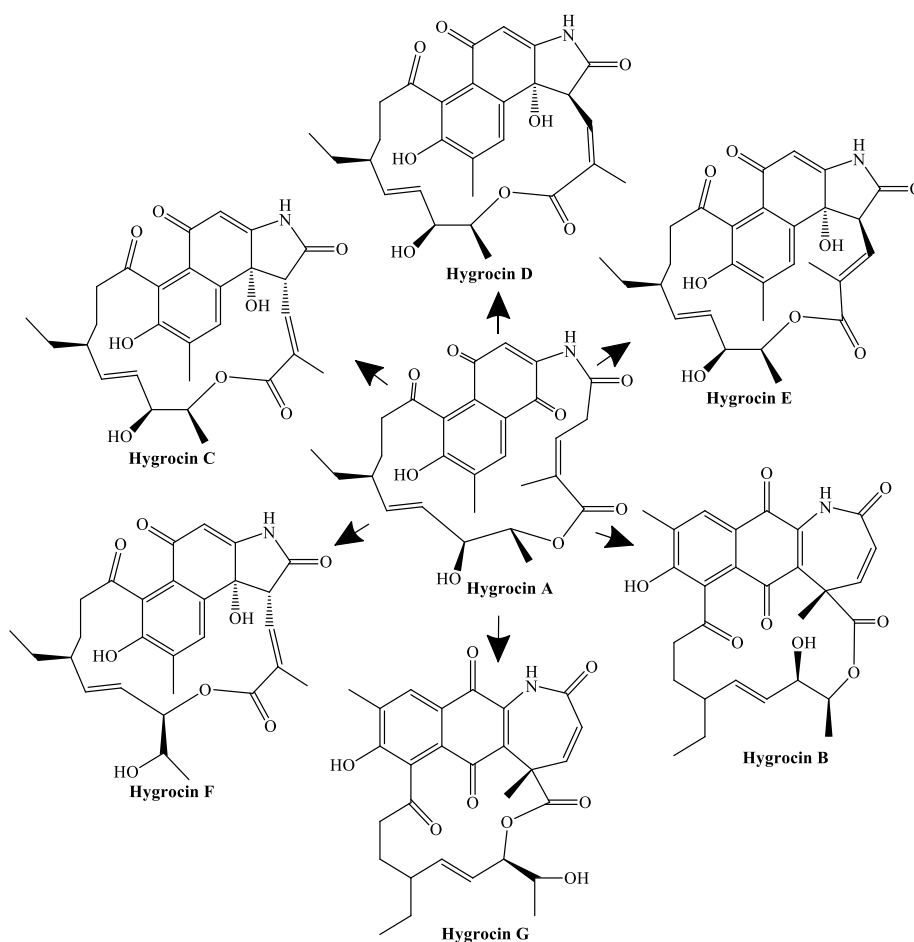


Figure 19. Structure of hygrocins C–E, derivatives of hygrocin A but differ in the configuration at C-2 and the orientation of the C-3,4 double bond. Hygrocins F and G were isomers of hygrocins C and B, respectively, due to the different alkyl oxygen participating in the macrolide ester linkage. (J. Nat. Prod. 2013, 76, 2175–2179).

Hygrocin A and B, along with a derivative of hygrocin A, were previously isolated from *Streptomyces hygroscopicus* ATCC25293. The chemical structures of hygrocins indicated that an AHBA starter unit and eight extender units, including four malonyl-CoA molecules, three methylmalonyl-CoA molecules, and one (2S)-ethylmalonyl-CoA molecule, are employed in the polyketide chain extension. Similarly, hygrocins C–E were shown to have the same planar structure as the “degradation” product. However, since the relative configuration of the

“degradation” product was not elucidated, it cannot be identified as one of the hygroscins C–E. In contrast, hygroscins B–G may be derived from the reactive precursor hygroscin A. The γ lactam of hygroscins C–F is formed in an intramolecular aldol reaction, and formation of the seven-membered lactam ring in hygroscin B and G is possibly due to the attack of the carbonyl-activated methylene on the quinone in a vinylogous fashion. The hygroscins (1–6) were evaluated for cytotoxicity against various human cancer cell lines (breast cancer MDAMB-431 cells, prostate cancer PC3 cells, alveolar basal epithelial cells A549, colorectal cancer SW620 cells, and hepatocellular liver carcinoma cells HepG2). Though the hygroscins tested were structurally quite similar, they showed dramatically different biological activities. Hygroscins C, D, and F were observed to be cytotoxic to human breast cancer MDA-MB-431 cells and prostate cancer PC3 cells, whereas hygroscins B, E, and G showed no toxicity relative to the control. As hygroscins C, D, and E have the same planar structure, the lack of activity of E might be ascribed to the configuration of E with a 3, 4 double bond [67, 68].

NEW BIOACTIVE ANSAMYCIN

Chaxamycins

In 2011 chaxamycins A-D, new bioactive ansamycin type polyketides were isolated from the culture broth of *Streptomyces* sp. strain C34. A unique structural feature of the new ansamycins (A-D) (Figure 20) that distinguishes them from the previously known members of this family is the lack of a methyl group at the olefinic C-16 next to the amide link when compared with previously isolated naphthalenic ansamycins such as the rifamycins, streptovaricins, tolypomycins, the halomicins, naphthomycin, actamycin, and damavaricins, and the bransarols as well as the benzenic ansamycins such as geldanamycin, and the herbimycins and macbecins. Thus, chaxamycin A is the 8-hydroxy-16-demethyl analogue of protostreptovaricin-I while chaxamycin B is the 16-demethyl analogue of protostreptovaricin-I.

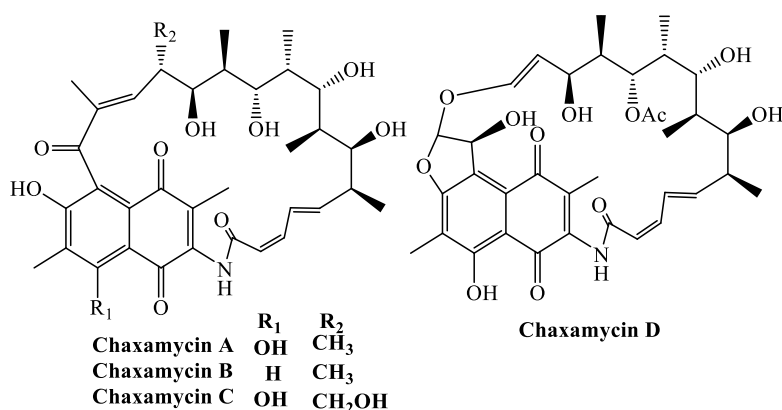
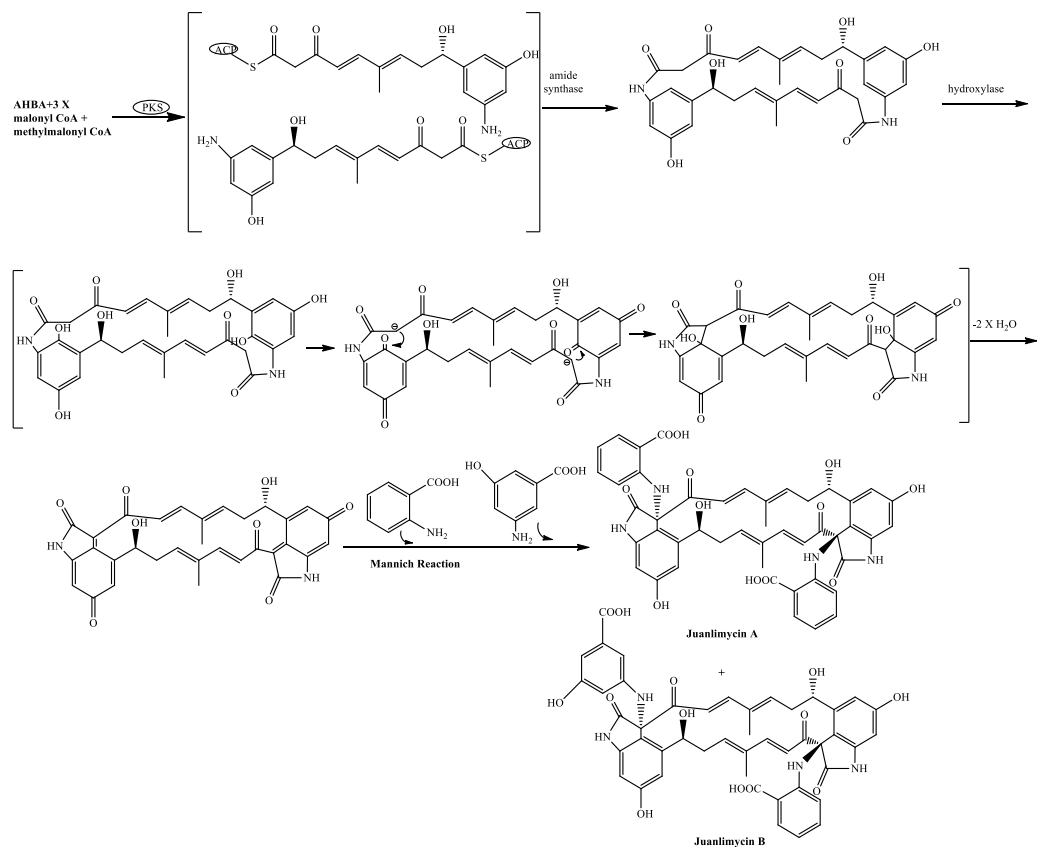


Figure 20. Structure of Chaxamycin A-D.



dx.doi.org/10.1021/ol501072x | Org. Lett. 2014, 16, 2752-2755

Figure 21. Proposed Biosynthetic Pathway and Post-PKS Modifications of Juanlimycins A and B.

Chaxamycins A-D possess antibacterial activities against the Gram-positive *S. aureus* ATCC 25923 and the Gram-negative *E. coli* ATCC 25922. Chaxamycin D showed a highly selective antibacterial activity against *S. aureus* ATCC 25923, displaying activity against almost all strains [69, 70].

Juanlimycins

In 2014, juanlimycins A and B, two novel ansamycin macrodilactams with unprecedented feature (Figure 21), were reported in the culture broth of *Streptomyces* sp. LC6. This was determined by analysis of high-resolution ESIMS, 1D and 2D NMR spectroscopic data and X-ray single crystal diffraction study. Juanlimycins A and B were found to be active against *Staphylococcus aureus* ATCC 25923, *Mycobacterium smegmatis* mc2 155 and *Candida albicans* 5314. Juanlimycins A and B are also the first macrodilactams among the reported ansamycins [71].

REFERENCES

- [1] Funayama S, Cordell GA. "Ansamycin antibiotics^A discovery, classification, biosynthesis and biological activities" *Stud. Nat. Prod. Chem.* 2000, 23, 51-106.
- [2] Arakawa K, Müller R, Mahmud T, Yu TW, Floss HG. "Characterization of the early stage aminoshikimate pathway in the formation of 3-amino-5-hydroxybenzoic acid: the RifN protein specifically converts kanosamine into kanosamine 6-phosphate" *J. Am. Chem. Soc.* 2002, 124, 10644-5.
- [3] Watanabe K, Rude MA, Walsh CT, Khosla C. "Engineered biosynthesis of an ansamycin polyketide precursor in *Escherichia coli*" *Proc. Natl. Acad. Sci. USA.* 2003,100, 9774-8.
- [4] Jeso V, Iqbal S, Hernandez P, Cameron MD, Park H, LoGrasso PV, Micalizio GC. "Synthesis of benzoquinone ansamycin-inspired macrocyclic lactams from shikimic acid" *Angew Chem. Int. Ed. Engl.* 2013, 52, 4800-4.
- [5] Cassady JM, Chan KK, Floss HG, Leistner E. "Recent developments in the maytansinoid antitumor agents" *Chem. Pharm. Bull (Tokyo).* 2004, 52, 1–26.
- [6] Yu TW, Bai L, Clade D, Hoffmann D, Toelzer S, Trinh KQ, Xu J, Moss SJ, Leistner E, Floss HG. "The biosynthetic gene cluster of the maytansinoid antitumor agent ansamitocin from *Actinosynnema pretiosum*" *Proc. Natl. Acad. Sci. USA.* 2002, 99, 7968-73.
- [7] Rinehart KL Jr, Shield LS. "Chemistry of the ansamycin antibiotics" *Fortschr Chem. Org. Naturst.* 1976, 33, 231–307.
- [8] Zhao P, Bai L, Ma J, Zeng Y, Li L, Zhang Y, Lu C, Dai H, Wu Z, Li Y, Wu X, Chen G, Hao X, Shen Y, Deng Z, Floss HG "Amide N-Glycosylation by Asm25, an N-Glycosyltransferase of Ansamitocins" *Chem. Biol.* 2008,15,863-74.
- [9] Sievers EL, Senter PD. "Antibody-drug conjugates in cancer therapy" *Annu. Rev. Med.* 2013, 64, 15–29.
- [10] Cassady JM, Chan KK, Floss HG, Leistner E. "Recent developments in the maytansinoid antitumor agents" *Chem. Pharm Bull (Tokyo).* 2004, 52, 1–26.
- [11] Issell BF, Crooke ST. "Maytansine" *Cancer Treat Rev.* 1978, 5, 199–207.
- [12] Ikeyama S, Takeuchi M. "Antitubulin activities of ansamitocins and maytansinoids" *Biochem. Pharmacol.* 1981, 30, 2421–2425.
- [13] Oroudjev E, Lopus M, Wilson L, Audette C, Provenzano C, Erickson H, Kovtun Y, Chari R, Jordan MA. "Maytansinoid-antibody conjugates induce mitotic arrest by suppressing microtubule dynamic instability" *Mol. Cancer Ther.* 2010, 9, 2700–2713.
- [14] Venghateri JB, Gupta TK, Verma PJ, Kunwar A, Panda D. "Ansamitocin P3 Depolymerizes Microtubules and Induces Apoptosis by Binding to Tubulin at the Vinblastine Site Jubina B" *PLoS One.* 2013, 8, e75182.
- [15] Ma N, Wei L, Fan Y, Hua Q. "Heterologous expression and characterization of soluble recombinant 3-deoxy-d-arabino-heptulosonate-7-phosphate synthase from *Actinosynnema pretiosum* ssp. *auranticum* ATCC31565 through co-expression with Chaperones in *Escherichia coli*" *Protein Expr Purif.* 2012, 82, 263–269.
- [16] Higashide E, Asai M, Ootsu K, Tanida S, Kozai Y, Hasegawa T, Kishi T, Sugino Y, Yoneda M. "Ansamitocin a group of novel maytansinoid antibiotics with antitumor properties from *Nocardia*" *Nature.* 1977, 270,721-2.

- [17] Yao Y, Cheng Z, Ye H, Xie Y, He J, Tang M, Shen T, Wang J, Zhou Y, Lu Z, Luo F, Chen L, Yu L, Yang JL, Peng A, Wei Y. "Preparative isolation and purification of anti-tumor agent ansamitocin P-3 from fermentation broth of *Actinosynnema pretiosum* using high-performance countercurrent chromatography" *J. Sep. Sci.* 2010, 33,1331-7.
- [18] Ma N, Wei L, Fan Y, Hua Q. "Heterologous expression and characterization of soluble recombinant 3-deoxy-d-arabino-heptulosonate-7-phosphate synthase from *Actinosynnema pretiosum* ssp. *auranticum* ATCC31565 through co-expression with Chaperones in *Escherichia coli*" *Protein Expr. Purif.* 2012, 82, 263–269.
- [19] Gao Y, Fan Y, Nambou K, Wei L, Liu Z, Imanaka T, Hua Q. "Enhancement of ansamitocin P-3 production in *Actinosynnema pretiosum* by a synergistic effect of glycerol and glucose" *J. Ind. Microbiol. Biotechnol.* 2014, 41,143-52.
- [20] Floss HG, Yu TW, Arakawa K. "The biosynthesis of 3-amino-5- hydroxybenzoic acid (AHBA), the precursor of mC7N units in ansamycin and mitomycin antibiotics: a review" *J. Antibiot.* 2011, 64, 35–44.
- [21] Hamel E. "Natural products which interact with tubulin in the vinca domain: maytansine, rhizoxin, phomopsin A, dolastatins 10 and 15 halichondrin B" *Pharmacol. Ther.* 1992, 55, 31–51.
- [22] Venghateri JB, Gupta TK, Verma PJ, Kunwar A, Panda D. "Ansamitocin P3 Depolymerizes Microtubules and Induces Apoptosis by Binding to Tubulin at the Vinblastine Site" *PLoS One.* 2013,8,e75182.
- [23] Martin K, Müller P, Schreiner J, Prince SS, Lardinois D, Heinzelmann-Schwarz VA, Thommen DS, Zippelius A. "The microtubule-depolymerizing agent ansamitocin P3 programs dendritic cells toward enhanced anti-tumor immunity" *Cancer Immunol Immunother.* 2014, 63, 925-38.
- [24] DeBoer C, Meulman PA, Wnuk RJ, Peterson DH. "Geldanamycin, a new antibiotic" *J. Antibiot. (Tokyo).* 1970, 23, 442-7.
- [25] Wu CZ, Jang JH, Ahn JS, Hong YS. "New geldanamycin analogs from *Streptomyces hygroscopicus*" *J. Microbiol. Biotechnol.* 2012, 22, 1478-81.
- [26] Li YH, Li M, He WQ, Wang YG, Shao RG. "Inactivation of putative PKS genes can double geldanamycin yield in *Streptomyces hygroscopicus* 17997" *Genet Mol. Res.* 2013, 12, 2076-85.
- [27] McCollum AK, Lukasiewicz KB, Teneyck CJ, Lingle WL, Toft DO, Erlichman C. "Cisplatin abrogates the geldanamycin-induced heat shock response" *Mol. Cancer Ther.* 2008, 7, 3256-64.
- [28] Li Z, Jia L, Wang J, Wu X, Hao H, Xu H, Wu Y, Shi G, Lu C, Shen Y. "Design, synthesis and biological evaluation of 17-arylmethylamine-17- demethoxygeldanamycin derivatives as potent Hsp90 inhibitors" *Eur. J. Med. Chem.* 2014, 85, 359-70.
- [29] Li Z, Jia L, Wang J, Wu X, Shi G, Lu C, Shen Y. "Discovery of Novel 17 Phenylethylaminegeldanamycin Derivatives as Potent Hsp90 Inhibitors" *Chem. Biol. Drug Des.* 2014, doi: 10.1111/cbdd.12371.
- [30] Wang S, Wang X, Du Z, Liu Y, Huang D, Zheng K, Liu K, Zhang Y, Zhong X, Wang Y. "SNX-25a, a novel Hsp90 inhibitor, inhibited human cancer growth more potently than 17-AAG" *Biochem. Biophys. Res. Commun.* 2014, 450, 73-80.

- [31] Tanida S, Hasegawa T, Higashide E. "Macbecins I and II, new antitumor antibiotics. I. Producing organism, fermentation and antimicrobial activities" *J. Antibiot. (Tokyo)*. 1980, 33, 199-204.
- [32] Bailey SN, Sabatini DM, Stockwell BR. "Microarrays of small molecules embedded in biodegradable polymers for use in mammalian cell-based screens" *Proc. Natl. Acad. Sci. USA*. 2004, 101, 16144-9.
- [33] Hatakeyama T, Sakai H, Yamaguchi M, Shimura H, Kuzume M, Matsumoto T, Matsumiya A, Yoshizawa Y, Midorikawa T, Kumada K, Sanada Y. "Protective effects of herbimycin A on hepatic reperfusion injury" *Transplant Proc*. 2000, 32, 2303-5.
- [34] Pfannkuche A, Büther K, Karthe J, Poenisch M, Bartenschlager R, Trilling M, Hengel H, Willbold D, Häussinger D, Bode JG. "c-Src is required for complex formation between the hepatitis C virus-encoded proteins NS5A and NS5B: a prerequisite for replication" *Hepatology*. 2011, 53, 1127-36.
- [35] Hamel L, Kenney M, Jayyosi Z, Ardati A, Clark K, Spada A, Zilberstein A, Perrone M, Kaplow J, Merkel L, Rojas C. "Induction of heat shock protein 70 by herbimycin A and cyclopentenone prostaglandins in smooth muscle cells" *Cell Stress Chaperones*. 2000, 5, 121-31.
- [36] Shaaban KA, Wang X, Elshahawi SI, Ponomareva LV, Sunkara M, Copley GC, Hower JC, Morris AJ, Kharel MK, Thorson JS. "Herbimycins D-F, ansamycin analogues from *Streptomyces* sp.RM-7-15" *J. Nat. Prod*. 2013, 76, 1619-26.
- [37] Hiramoto S, Sugita M, Ando C, Sasaki T, Furihata K, Seto H, Otake N. "Studies on mycotrienin antibiotics, a novel class of ansamycins. V. Isolation and structure determination of novel mycotrienin congeners" *J. Antibiot. (Tokyo)*. 1985, 38, 1103-6.
- [38] Reynolds KA, Wang P, Fox KM, Floss HG. "Biosynthesis of ansatrienin by *Streptomyces collinus*: cell-free transformations of cyclohexene- and cyclohexadiene-carboxylic acids" *J. Antibiot. (Tokyo)*. 1992, 45, 411-9.
- [39] Feuerbach D, Waelchli R, Fehr T, Feyen JH. "Mycotrienins. A new class of potent inhibitors of osteoclastic bone resorption" *J. Biol. Chem*. 1995, 270, 25949-55.
- [40] Yamada Y, Tashiro E, Taketani S, Imoto M, Kataoka T. "Mycotrienin II, a translation inhibitor that prevents ICAM-1 expression induced by pro-inflammatory cytokines" *J. Antibiot. (Tokyo)*. 2011, 64, 361-6.
- [41] Futamura Y, Tashiro E, Hironiwa N, Kohno J, Nishio M, Shindo K, Imoto M. "Trierixin, a novel Inhibitor of ER Stress-induced XBP1 activation from *Streptomyces* sp. II. structure elucidation" *J. Antibiot. (Tokyo)*. 2007, 60, 582-5.
- [42] Tashiro E, Hironiwa N, Kitagawa M, Futamura Y, Suzuki S, Nishio M, Imoto M. "Trierixin, a novel Inhibitor of ER stress-induced XBP1 activation from *Streptomyces* sp. I. Taxonomy, fermentation, isolation, and biological activities" *J. Antibiot. (Tokyo)*. 2007, 60, 547-53.
- [43] Kawamura T, Tashiro E, Shindo K, Imoto M. "SAR Study of a novel triene-ansamycin group compound, quinotrierixin, and related compounds, as inhibitors of ER Stress-induced XBP1 Activation" *J. Antibiot. (Tokyo)*. 2008, 61, 312-7.
- [44] Kawamura T, Tashiro E, Yamamoto K, Shindo K, Imoto M. "SAR Study of a novel triene-ansamycin group compound, quinotrierixin, and related compounds, as inhibitors of ER Stress-induced XBP1 Activation" *J. Antibiot. (Tokyo)*. 2008, 61, 303-11.

- [45] Kakeya H, Zhang HP, Kobinata K, Onose R, Onozawa C, Kudo T, Osada H. "Cytotrienin A, a Novel Apoptosis Inducer in Human Leukemia HL-60 Cells" *J. Antibiot. (Tokyo)*. 1997, 50, 370-2.
- [46] Yamada Y, Taketani S, Osada H, Kataoka T. "Cytotrienin A, a translation inhibitor that induces ectodomain shedding of TNF receptor 1 via activation of ERK and p38 MAP kinase" *Eur. J. Pharmacol.* 2011, 667, 113-9.
- [47] Funayama S, Okada K, Iwasaki K, Komiyama K, Umezawa I. "Structures of trienomycins A, B and C, novel cytotoxic ansamycin antibiotics" *J. Antibiot. (Tokyo)*. 1985, 38, 1677-83.
- [48] Nomoto H, Katsumata S, Takahashi K, Funayama S, Komiyama K, Umezawa I, Omura S. "Structural studies on minor components of trienomycin group antibiotics trienomycins D and E" *J. Antibiot. (Tokyo)*. 1989, 42, 479-81.
- [49] Kim WG, Song NK, Yoo ID. "Trienomycin G, a New Inhibitor of Nitric Oxide Production in Microglia Cells, from *Streptomyces* sp. 91614" *J. Antibiot. (Tokyo)*. 2002, 55, 204-7.
- [50] Saxena A, Kumari R, Mukherjee U, Singh P, Lal R. "Draft Genome Sequence of the Rifamycin Producer *Amycolatopsis rifamycinica* DSM 46095" *Genome Announc.* 2014, 2, e00662-14.
- [51] Tang L, Yoon YJ, Choi CY, Hutchinson CR. "Characterization of the enzymatic domains in the modular polyketide synthase involved in rifamycin B biosynthesis by *Amycolatopsis mediterranea*" *Gene*. 1998, 216, 255-65.
- [52] Goldstein BP. "Resistance to rifampicin: a review" *J. Antibiot. (Tokyo)*. 2014, 67, 625-630.
- [53] Nigam A, Almabruk KH, Saxena A, Yang J, Mukherjee U, Kaur H, Kohli P, Kumari R, Singh P, Zakharov LN, Singh Y, Mahmud T, Lal R. "Modification of Rifamycin Polyketide Backbone Leads to Improved Drug Activity against Rifampicin-resistant *Mycobacterium tuberculosis*" *J. Biol. Chem.* 2014, 289, 21142-21152.
- [54] Evans SE, Goult BT, Fairall L, Jamieson AG, Ko Ferrigno P, Ford R, Schwabe JW, Wagner SD. "The ansamycin antibiotic, rifamycin SV, inhibits BCL6 transcriptional repression and forms a complex with the BCL6-BTB/POZ domain" *PLoS One*. 2014, 9, e90889.
- [55] Bi W, Zhu L, Jing X, Zeng Z, Liang Y, Xu A, Liu J, Xiao S, Yang L, Shi Q, Guo L, Tao E. "Rifampicin improves neuronal apoptosis in LPS-stimulated co-cultured BV2 cells through inhibition of the TLR-4 pathway" *Mol. Med. Rep.* 2014, 10, 1793-9.
- [56] Yulug B, Hanoglu L, Kilic E, Schabitz WR. "RIFAMPICIN: An antibiotic with brain protective function" *Brain Res. Bull.* 2014, 107C, 37-42.
- [57] O'Brien RJ, Lyle MA, Snider DE Jr. "Rifabutin (ansamycin LM 427): a new rifamycin-S derivative for the treatment of mycobacterial diseases" *Rev. Infect Dis.* 1987, 9, 519-30.
- [58] Davies G, Cerri S, Richeldi L. "Rifabutin for treating pulmonary tuberculosis" *Cochrane Database Syst. Rev.* 2007, 4, CD005159.
- [59] Gisbert JP, Calvet X. "Review article: rifabutin in the treatment of refractory *Helicobacter pylori* infection" *Aliment Pharmacol. Ther.* 2012, 35, 209-21.
- [60] Tanuma J, Sano K, Teruya K, Watanabe K, Aoki T, Honda H, Yazaki H, Tsukada K, Gatanaga H, Kikuchi Y, Oka S. "Pharmacokinetics of rifabutin in Japanese HIV-infected patients with or without antiretroviral Therapy" *PLoS One*. 2013, 8, e70611.

- [61] Jo KW, Ji W, Hong Y, Lee SD, Kim WS, Kim DS, Shim TS. "The efficacy of rifabutin for rifabutin-susceptible, multidrug-resistant tuberculosis" *Respir Med.* 2013, 107, 292-7.
- [62] Sirgel FA, Warren RM, Böttger EC, Klopper M, Victor TC, van Helden PD. "The Rationale for Using Rifabutin in the Treatment of MDR and XDR Tuberculosis Outbreaks" *PLoS One.* 2013, 8, e59414.
- [63] Lan NT, Thu NT, Barrail-Tran A, Duc NH, Lan NN, Laureillard D, Lien TT, Borand L, Quillet C, Connolly C, Lagarde D, Pym A, Lienhardt C, Dung NH, Taburet AM, Harries AD. "Randomised pharmacokinetic trial of rifabutin with lopinavir/ritonavir-antiretroviral therapy in patients with HIV-associated tuberculosis in Vietnam" *PLoS One.* 2014, 9, e84866.
- [64] Weiner M, Egelund EF, Engle M, Kiser M, Prihoda TJ, Gelfond JA, Mac Kenzie W, Peloquin CA. "Pharmacokinetic interaction of rifapentine and raltegravir in healthy volunteers" *J. Antimicrob. Chemother.* 2014, 69, 1079-85.
- [65] Savic RM, Lu Y, Bliven-Sizemore E, Weiner M, Nuermberger E, Burman W, Dorman SE, Dooley KE. "Population pharmacokinetics of rifapentine and desacetyl rifapentine in healthy volunteers: nonlinearities in clearance and bioavailability" *Antimicrob. Agents Chemother.* 2014, 58, 3035-42.
- [66] Matsuda S, Adachi K, Matsuo Y, Nukina M, Shizuri Y. "Salinisporamycin, a novel metabolite from *Salinispora arenicola*" *J. Antibiot. (Tokyo).* 2009, 62, 519-26.
- [67] Li S, Wang H, Li Y, Deng J, Lu C, Shen Y, Shen Y. "Biosynthesis of hygrocins, antitumor naphthoquinone ansamycins produced by *Streptomyces* sp. LZ35" *Chembiochem.* 2014, 15, 94-102.
- [68] Lu C, Li Y, Deng J, Li S, Shen Y, Wang H, Shen Y. "Hygrocins C–G, cytotoxic naphthoquinone ansamycins from *gdmAI*-disrupted *Streptomyces* sp. LZ35" *J. Nat. Prod.* 2013, 76, 2175-9.
- [69] Rateb ME¹, Houssen WE, Arnold M, Abdelrahman MH, Deng H, Harrison WT, Okoro CK, Asenjo JA, Andrews BA, Ferguson G, Bull AT, Goodfellow M, Ebel R, Jaspars M. "Chaxamycins A–D, bioactive ansamycins from a hyper-arid desert *Streptomyces* sp" *J. Nat. Prod.* 2011, 74, 1491-9.
- [70] Chen M, Roush WR. "Crotylboron-based synthesis of the polypropionate units of chaxamycins A/D, salinisporamycin, and rifamycin S" *J. Org. Chem.* 2013, 78, 3-8.
- [71] Zhang J, Qian Z, Wu X, Ding Y, Li J, Lu C, Shen Y. "Juanlimycins A and B, ansamycin macrodilactams from *Streptomyces* sp" *Org. Lett.* 2014, 16, 2752-5.

Chapter 47

BRCA GENE MUTATIONS MEDIATE PARTICULARLY HIGH TNBC RISK BY DEFECTIVE ESTROGEN SIGNALING

Zsuzsanna Suba*

National Institute of Oncology,
Surgical and Molecular Tumor Pathology Centre
Budapest, Hungary

ABSTRACT

Ubiquitously expressed protein products of *BRCA1* and *BRCA2* genes are implicated in processes fundamental to all cells, including DNA repair and recombination, checkpoint control of cell cycle, and transcription. *BRCA* gene mutations lead to disruption of *BRCA* proteins in mutation carrier cases and induce susceptibility to specific types of cancer. Among women with germline *BRCA* mutations near 50% of mammary malignancies are triple negative breast cancer (TNBC) presenting with a high grade histologically. Among women with breast cancer, TNBC was established in 57.1% of *BRCA1*-mutation positive and in 23.3% of *BRCA2*-mutation positive cases, whereas in only 13.8% of *BRCA*-proficient women. Although *BRCA* gene mutation carrier women usually exhibit clinical symptoms of defective estrogen receptor (ER) signaling; such as anovulatory infertility and early menopause, the serum estrogen levels of these patients are consequently elevated. In these cases, a compensatory feedback mechanism aims to break through the inherited or acquired ER resistance by increased estrogen synthesis so as to maintain the cellular estrogen surveillance. The higher the estrogen overproduction of *BRCA*-mutation positive cases, the higher the possibility of tumor-free survival. In conclusion, *BRCA1* and *BRCA2* gene mutations seem to increase the breast cancer risk, particularly that of TNBC, in case of insufficient compensation of defective ER signaling. Upregulation of these genes by means of elevated estrogen levels of high parity, artificial hormonal cycle created by oral contraceptives or a pregnancy mimicking high estrogen dose may decrease the excessive cancer risk of *BRCA* mutation positive women.

* Corresponding Author's Email: subazdr@gmail.com (Tel: 00 36 1 224 86 00; Fax: 0036 1 224 86 20).

INTRODUCTION

A small proportion of breast cancer cases, in particular those arising at a young age, are attributable to a highly penetrant, autosomal dominant predisposition to the disease. Twenty years ago, the identification of *BRCA1* gene was announced, the first gene known to strongly predispose to breast cancer when becoming mutated [1]. Identification of *BRCA2* gene was revealed the next year, in 1995 [2]. *BRCA1* and *BRCA2* are classic deoxyribonucleic acid (DNA) stabilizing, tumor-suppressor genes that repair breaks in double-stranded DNA. Now, twenty years later, we can summarize the scientific values of this discovery and its consequences for breast cancer prevention and treatment today.

Inherited mutations in *BRCA1* or *BRCA2* genes predispose to breast, ovarian, and other cancers. Their ubiquitously expressed protein products are implicated in processes fundamental to all cells, including DNA repair and recombination, checkpoint control of cell cycle, and transcription [3]. *BRCA* gene mutations lead to a defect of DNA double-strand break repair through homologous recombination. Disruption of *BRCA* proteins in mutation carriers can induce susceptibility to specific types of cancer [4]. However, the so-called safeguard activities of *BRCA1* gene do not appear to be cell-type specific. Thus, they do not, by themselves, fully elucidate the strong association of *BRCA1* gene mutations with specific cancer types, such as breast cancer.

The incidence of hereditary breast and ovarian cancers reveals close correlation with *BRCA* mutations [5]. *BRCA1/BRCA2* mutations are responsible for 3-8% of all breast cancer cases, whereas for 30-40% of familial cases. Ten percent of patients with ovarian cancer have a genetic predisposition. About 80% of families with a history of ovarian cancer have mutations in the *BRCA1*, while 15% in the *BRCA2* gene. These correlations suggest strong parallelism between the initiations of breast and ovarian cancers, which female organs are mistakenly regarded as being highly endangered by increased endogenous estrogen levels [6,7]. Poorly differentiated triple negative breast cancers (TNBCs) were first characterized in the literature in 2005 by the absence of steroid receptors for estrogen (ER) and progesterone (PR) as well as the lack of tyrosine kinase human epidermal growth factor receptor 2 (HER-2) [8]. Clinically, TNBCs exhibit fairly aggressive local growth, rapid progression and account for a high rate of early metastases, most commonly to visceral organs and central nervous system [9]. Among women with germline *BRCA1* mutation near 50% of breast cancers is triple negative presenting with a high grade histologically [10,11]. Among women with breast cancer, TNBC was established in 57.1% of *BRCA1*-mutation positive and in 23.3% of *BRCA2*-mutation positive cases, whereas in only 13.8% of *BRCA* negative women [12].

Strong correlation between *BRCA* gene mutations and the high risk of TNBC proposes certain mediators between germline mutations and the risk for poorly differentiated breast cancers.

INTERACTIONS BETWEEN ENDOGENOUS ESTROGENS AND WELL FUNCTIONING, WILD *BRCA* GENES

Wild type *BRCA1* gene expression is strongly enhanced during puberty and pregnancy, when estrogen levels exhibit a dramatic increase and an excessive development of mammary

gland takes place [13,14,15]. These observations suggest that high estrogen levels might strongly stimulate the expression of DNA stabilizing *BRCA* genes [16]. Since during pregnancy, the explosion-like cell proliferation in the female organs and fetal structures is associated with extreme increase in estrogen levels, the concomitantly elevated expression of the *BRCA1* gene may serve as a strict safeguard of abrupt DNA synthesis and mitotic activity.

In the mammary gland of ovariectomized mice, estradiol administration increased the level of *BRCA1* expression [13]. Further studies on ER-positive MCF-7 and BT20T human breast cancer cell lines indicated that depletion of estrogens significantly reduced *BRCA1* mRNA expression, and the expression was increased again by treatment with estradiol [17,18]. In *BRCA* proficient human breast cancer cell lines, increased *BRCA1* mRNA expression by estradiol administration suggests that proper estrogen exposure induces apoptosis or differentiation of tumor cells by quick DNA repair and recombination.

Evidences support that estrogen-induced activation of the normal *BRCA1* gene may be important in protecting the breast against cancer initiation. The phytoestrogen genistein, a major active component in soy products, may reduce the risk of premenopausal breast cancer development [19], and increases the expression of *BRCA1* gene in human breast cancer cells in culture [20]. Besides weakly estrogenic genistein, estrogens may also up-regulate the *BRCA1* gene and reduce breast cancer risk. In animal models before the onset of puberty, administration of 17 β estradiol reduced the later risk of breast cancer by the increased expression of *BRCA1* gene [21].

Clinical and experimental data support the cancer preventive and curative impacts of estrogens by means of the augmented DNA stabilizing effect of the increased *BRCA1* gene expression. By contrast, estrogen depletion or defective estrogen signaling may induce cancer initiation and even tumor dedifferentiation attributed to low expression of DNA stabilizer *BRCA1* gene.

DEFECTIVE ESTROGEN SIGNALING ASSOCIATED WITH *BRCA* GENE MUTATIONS

While many specific roles of *BRCA1* gene remain to be clarified, it is clear that functional *BRCA1* protein is required to prevent the malignant transformation of breast cells [15]. Selective inactivation of *BRCA1* gene in mammary epithelial cells results in blunted ductal development, breast hyperplasia, as well as tumor formation [22]. In the breast of *BRCA1* mutation carrier women, the persistent presence of least differentiated type 1 lobules is a characteristic finding [23], suggesting that functional *BRCA1* is necessary for proper estrogen signaling and normal mammary lobular differentiation.

BRCA gene mutation carrier women frequently exhibit the clinical symptoms of estrogen deficiency, such as anovulatory infertility and ovarian failure [24,25,26], in spite of their elevated estrogen levels. By contrast, in certain premenopausal cases with *BRCA* gene mutation, the defects of ER signaling are clinically disguised by reactively increased estrogen synthesis and/or other compensatory mechanisms. With ageing, however, the relatively higher but decreasing estradiol levels are not enough for the breakthrough of ER signal transduction defects and these women have a lifelong increased risk for breast cancer [27].

Since both germline mutations in *BRCA* genes and anovulatory infertility are associated with high susceptibility for breast and ovarian cancers, correlations between *BRCA1* mutation and low response to fertility treatments were examined [24]. In *BRCA1* mutation positive women, the low response rate of ovaries to fertility treatment was significantly increased (33.3%) as compared with *BRCA1* mutation negative patients (3.3%). These results support that *BRCA1* mutations are associated with defective estrogen signaling reflected by an increased rate of ovulatory failure.

In women with *BRCA1/2* mutations, ovarian failure leads to earlier age at natural menopause (under the age of 40) and it was a significantly more frequent observation than among mutation-free cases ($p < 0.001$) [25,26]. The high risk of premature ovarian failure among *BRCA1/2* mutation carriers reflects the disturbances in estrogen synthesis or estrogen receptor signaling pathways. Disorders of estrogen signaling may confer the risk of tumors associated with *BRCA1/2* mutation [27,28].

DECREASED ER-ALPHA EXPRESSION IN CELLS WITH *BRCA* GENE MUTATION

Estrogens, estrogen receptors and *BRCA* genes are in strict crosstalk and interplay in order to maintain cellular health. In *BRCA* gene mutation carrier women, difficulties in estrogen signaling may be associated with the fairly decreased cellular ER-alpha expression.

Molecular basis of the failure of ER-alpha expression was studied in *BRCA* mutant cells. In the breast of women with *BRCA1* gene mutation, low ER-alpha expression and the associated decrease in estrogen signaling may explain the persistent presence of the least differentiated type 1 lobules [23]. *ESR1* gene expression was found to be 5.4-fold lower in *BRCA1*-mutant tumors than in *BRCA* gene proficient ones [29]. These findings suggest that a deficient *BRCA1* gene may cause decreased ER-alpha expression in both developing mammary glands and breast cancers. In patients with *BRCA1* gene defect, the loss of ER expression leads to the predominance of poorly differentiated breast cancers [30], which may be attributed to the decreased estrogen surveillance [28].

Breast tumors in *BRCA1* mutant cases are typically ER α -negative, whereas the majority of sporadic, ER-positive breast cancers express wild-type *BRCA1* [10]. These observations suggest causal associations between *BRCA1* mutation-linked breast cancer and low *ESR1* gene expression.

Considering that in *BRCA* gene deficient cases the crucial compensatory mechanisms are the restoration and nursing of sufficient ER signal transduction, it becomes obvious that selective ER α blockers seriously aggravate the emergency situation of low *ESR1* and ER expressions.

DECREASED ESTRADIOL-LIGANDED TRANSCRIPTIONAL ACTIVITY OF ERS IN *BRCA* GENE DEFICIENT CELLS

The *BRCA1* gene is regarded as an inhibitor of the transcriptional activity of ER-alpha in human breast cancer cells lines [31]. Even recent publications mistakenly established that

BRCA1 defect deliberates estrogen-induced DNA damage resulting in genomic instability and strong risk for tumor formation in highly estrogen regulated tissues [32].

Nevertheless, direct interaction between *BRCA1* protein and the transcriptional activity of ERs is a complex regulatory process. Inhibition of ER-alpha activity by *BRCA1* was demonstrated either in cell lines with endogenous wild type *BRCA1* or in a breast cancer cell line that has defective *BRCA1* [33]. The *BRCA1* protein was found to bind with ER-alpha in vivo and in vitro, by an estrogen independent interaction that mapped to the amino-terminal region of *BRCA1*. *BRCA1* proteins containing the amino-terminal binding with ER-alpha region lost their inhibitory effect on ER-alpha activity. These findings suggest that the amino-terminus of *BRCA1* protein positively interacts with ER-alpha, while the carboxyl-terminus of *BRCA1* may function as a simple transcriptional repression domain [33]. In conclusion, *BRCA1* gene may ambiguously regulate ER-alpha activity, depending on the momentary necessities.

Further studies established that *BRCA1* gene is upregulated in response to estradiol in mammary epithelial cells, and *BRCA1* in turn positively regulates the transcriptional activity of ERs. This implies the existence of a positive feedback mechanism that regulates functional interaction between *BRCA1* gene and ER-alpha in breast cancer cells [18].

In conclusion, *BRCA1* regulates ER expression positively; hence loss of *BRCA1* function may lead to ER-negative phenotype providing a molecular basis for the loss of ER expression in the majority of *BRCA1* mutant cancers [29].

There is a physiologic link between *BRCA1* gene and mammary gland development. *BRCA1* up-regulation is required for controlled proliferation and differentiation of breast cells in puberty and pregnancy [34], while loss of *BRCA1* results in impaired differentiation but increased poorly controlled proliferation in human mammary epithelial cell line [35].

In conclusion, the experienced controversial interactions between *BRCA1* gene and ER transcriptional activity may be attributed to the existence of a balanced feedback mechanism, which serves as a safeguard of healthy mammary cells in every phase of proliferative and resting periods. In breast cancer cells, active *BRCA1* gene and *BRCA* protein increase both the expression and transcriptional activity of ERs resulting in a progressive differentiation and even apoptotic death of tumor cells. *BRCA1* gene defect destroys the defensive transcriptional activity of ER signaling against mutant breast cancer cells, deliberating the initiation and progression of malignancies.

DEFENSIVE COUNTERACTIONS AGAINST BRCA GENE DEFECTS

Although *BRCA* gene mutation carrier women have an increased lifelong risk for cancer development, certain defensive counteractions may help to improve the safeguard of cell proliferation resulting in a tumor-free life.

Increased Aromatase Activity Associated with BRCA Gene Mutation

Transcriptional regulation of target genes is an important function of the *BRCA1* gene [36]. One of its target genes is aromatase CYP19A1, regulating the expression of aromatase enzyme that catalyses the conversion of C₁₉ steroids into bioactive estrogens [37]. In vitro studies

demonstrated the direct binding of *BRCA1* gene to the proximal promoter region of CYP19A1 gene (promoter II) in breast adipose fibroblasts and as a consequence the repression of transcriptional activity [38]. This finding was erroneously evaluated as a defensive mechanism against the danger of elevated estrogen synthesis, while both repression and activation are important players in balanced regulation.

Conversely, gene silencing of *BRCA1* seemed to be associated with deliberation of aromatase gene expression in human stromal adipose cells resulting in increased enzyme activity [39].

These observations suggest that defective *BRCA1* gene reactively induces excessive estrogen synthesis; however, the increased hormone level was erroneously regarded as causal factor of the cancers of breast and ovary.

Aromatase expression was significantly higher in *BRCA1* mutation carriers, either in patients who had experienced breast cancer or in women who had a high risk for breast cancer and had prophylactic removal of their breast tissue [36]. Increases in aromatase and its proximal promoter I.3/II transcripts were also observed in these *BRCA1* mutation carriers, supporting the hypothesis that with decreased *BRCA1* function there is an overregulation of aromatase transcription leading to excessive estrogen synthesis. Lack of functional *BRCA1* protein correlated to higher aromatase levels in 85% of *BRCA1* mutation carriers. The authors erroneously concluded that the link between defective *BRCA1* protein and increased aromatase activity is significant in terms of understanding why carcinogenesis is concentrated to estrogen-producing tissues in *BRCA1* mutation carriers.

Measurements of aromatase activity in the different quadrants of mastectomy specimens from patients with breast cancer indicated that activity was always higher in quadrants associated with tumor as compared with non-involved quadrants. In sequential biopsies of large primary tumors, the measurement of in vitro aromatase content before and during treatment with aminoglutethimide-hydrocortisone showed a marked but apparently paradoxical rise in enzyme activity following therapy [40]. These results may be estimated as the importance of local estrogen synthesis within the breast in terms of both the natural history and behavior of breast cancers.

BRCA1 gene mutation is associated with a combination of excessive aromatase expression and activity, predominantly in estrogen receptor negative phenotypes of tumors [30]. This link between excessive tissular aromatase synthesis and the development of poorly differentiated breast cancer is usually regarded as a justification of the carcinogenic impact of increased local estrogen level.

Nevertheless, in young breast cancer cases (<40 yrs.), locoregional control after breast conserving surgery highlighted that the absence of CYP19-aromatase activity in removed tumor samples carried a strongly significant risk for locoregional tumor recurrence and prophesied poor prognosis [41]. Lack of intratumoral estrogen synthesis seems to be strongly associated with low differentiation of breast cancers, and at the same time tumor growth and rapid tumor spread [28,42].

Prophylactic or therapeutic use of aromatase inhibitors in women with high breast and ovarian cancer risk fairly jeopardizes the health of female organs exhibiting intense defensive estrogen synthesis.

Hyperestrogenism as Compensatory Mechanism in Women with BRCA Gene Mutation

Characteristic high aromatase activity in women with *BRCA1* mutation harmonizes with the consequent finding of reactively increased estrogen concentration. Elevated estrogen synthesis counteracts decreased ER α expression so as to preserve cellular estrogen surveillance.

Excessive estrogen synthesis seems to be an effective counteraction against decreased ER expression in women with *BRCA* gene mutation. Higher serum titers of estradiol were measured in *BRCA1/BRCA2* mutation carriers as compared with women who were free of mutations [43]. The authors erroneously concluded that the high penetrance for breast and ovarian cancers or both can be explained by the increased estrogen synthesis in carriers of *BRCA1/BRCA2* mutations, as elevated estradiol level is regarded as a risk factor for these female tumors. Nevertheless, estradiol is beneficial even in sky-high doses and its long lasting reactive overproduction may save the mutation carrier patients from tumor development [27,28].

Stronger hyperestrogenism (71.7 pg/ml) was observed in *BRCA2* mutation carrier patients as compared with the modestly increased estrogen levels of women with *BRCA1* mutations (45.5 pg/ml) and the normal levels in cases without *BRCA* mutation (38.5 pg/ml) [44]. This result erroneously strengthens the practice of antiestrogen treatment in cases with breast cancer. Higher estrogen production in *BRCA2* mutation carriers seems to be an effective counteraction against defective estrogen signal transduction as a mediator of their markedly lower cancer risk, while the moderate estrogen overproduction in *BRCA1* mutation carriers is associated with higher tumor risk.

In *BRCA* gene mutation carriers, elevated estrogen levels associated with high parity, artificial hormonal cycle created by oral contraceptives or estrogen administration may decrease the high cancer risk. Parity in *BRCA1* mutation carriers significantly reduced the risk for ovarian cancer [45], moreover, the risk was reduced with each additional full-term pregnancy in women with germline mutation [46]. Furthermore, parity with its highly elevated estrogen level also seemed to be protective against TNBC, similarly like against the predominant ER-positive tumors [47].

Use of oral contraceptives (OCs) was found to highly reduce the risk of ovarian cancers in women with both *BRCA1* (OR: 0.56) and *BRCA2* mutations (OR: 0.39) [45]. Ovarian cancer risk decreased with each year of long term contraceptive use in women carrying *BRCA1* or *BRCA2* mutations [48]. Protective effect of OC was established as a chemoprevention against ovarian cancers in young women with *BRCA* mutations; whereas the OC associated risk of breast cancer in *BRCA* mutation carriers seemed to be heterogeneous with inconsistent results [49]. Nevertheless, the use of OCs for at least 12 months was associated with strongly decreased breast cancer risk for *BRCA1* mutation carriers (OR: 0.22) [50].

Consumption of phytoestrogen-rich foods such as soy emerged as preventive measure against breast cancer. Soy consumption may be beneficial in early life before puberty or during adolescence, according to results of immigrant and epidemiological studies [19]. In animal experiments, prepubertal administration of 17 β -estradiol reduced the later risk of breast cancer by inducing a persistent up-regulation of *BRCA1* gene [21].

Increased Ligand-Independent Transcriptional Activity in Cells with *BRCA1* Mutation

A relative decrease in estradiol-liganded transcriptional activity of ERs was observed in *BRCA1* gene deficient human ovarian cancer cells [31]. Nevertheless, ER α showed an unexpectedly increased ligand-independent transcriptional activity in cells with *BRCA1* mutation that was not observed in *BRCA1*-proficient cells [51]. The authors mistakenly concluded that normal *BRCA1* gene mediates the repression of the ligand-independent transcription of the estrogen receptors, while *BRCA1* mutation deliberates the overwhelming, harmful estrogen signaling by ligand independent pathways.

Nevertheless, increased estrogen independent stimulation of ERs in *BRCA1*-deficient tumor cells, may be regarded as a counteraction against defective estrogen-dependent ER transcription aiming the restoration of omnipotent estrogen signaling in emergency situations. High risk of cancer may be manifested if the defect of ligand activated ER signal is not supplemented by compensatory non-liganded ER signaling.

In conclusion, *BRCA1* gene mutation seems to increase the risk of breast cancer initiation and progression particularly that of TNBC by the defective ligand activated transcriptional activity of estrogen. In cells with *BRCA1* mutation, the increased non-liganded transcriptional activity of ERs may be a compensatory mechanism for the improvement of estrogen signaling.

HORMONAL RISK FACTORS FOR OVERALL BREAST CANCER AND TNBC

All justified risk factors for poorly differentiated TNBC development seem to be in close correlation with estrogen loss or defective estrogen signaling and further associated metabolic disorders [28]. The question arises whether *BRCA* mutations may lead to breast cancer and preferentially TNBC development by the defect of estrogen signaling or by quite different pathways.

The majority of breast cancer risk factors seem to be common to both ER-positive and ER-negative tumors, including TNBCs [28]. Well known risk factors for overall breast cancer, such as metabolic syndrome, type 2 diabetes, obesity, African-American race and *BRCA* gene mutation all proved to be particularly dangerous for the development of poorly differentiated, ER-negative tumors and TNBCs as well [12,52-58]. Moreover, the stronger the risk factors, the higher the danger of ER-negative breast cancers, included TNBCs, as compared with ER-positive tumors.

Estrogen-related cancer risk seems to be apparently different, even quite inverse concerning ER-positive breast cancers and TNBCs [59]. These controversies suggest that the biologic mechanisms behind the initiation and progression of both TNBCs and non-TNBCs are completely obscure until now.

Correlations between Menopausal Status and TNBC Risk

Literary data strongly support that young age and premenopausal status in women means an equivocally higher incidence rate of TNBCs as compared with older, postmenopausal cases [60-64]. Nevertheless, it is hardly conceivable that TNBC, which is regarded as an apparently hormone independent tumor, would exhibit increased incidence rate at higher, premenopausal estrogen concentrations, while rarely occurs in hormonally challenged postmenopausal women [28].

In a case control study, breast cancers in women less than 40 years of age, exhibited a significantly higher rate of ER-negative tumors (33.8%) as compared with older cases (21.9%), suggesting a disproportional risk for poorly differentiated tumors in the young age group [61]. Nevertheless, the raw numbers of age dependent incidence of breast cancers clearly show a close to twofold increase in ER-negative tumors and an almost fourfold increase in ER-positive tumors with ageing (Table. 1.). All the statistical risk indices, such as odds ratios (ORs), hazard ratios (HRs) and incidence risk ratios (IRRs) are based on the percentage of cancer cases. Consequently, the majority of studies erroneously report on higher TNBC risk among young women, disregarding the low number of overall breast cancer cases among these cases [28].

Table 1. Age related increases in ER-positive and ER-negative breast cancer incidences

Menopausal status	Average age of patients	Number of patients		Percentage of patients	
		ER+ tumors	ER- tumors	ER+ tumors	ER- tumors
Premenopausal cases	35 years	52	26	66.2%	33.8%
Postmenopausal cases	59 years	178	50	78.1%	21.9%

Notes: With ageing, raw number of tumors shows a close to twofold increase in ER-negative subtypes and a much higher – almost fourfold – increase in ER-positive subtypes, while the percentage of ER-negative cancers exhibits a misleading decreasing trend. Data derived from Hartley et al. (ref. 61). Abbreviation: ER, estrogen receptor.

What can the advantageous factor be in young women suppressing strongly the ER-positive and moderately the ER-negative breast cancers? In premenopausal cases, healthy or slightly defective estrogen synthesis supplies the ligands for the available ERs of tumor cells. In case of sufficient estrogen exposure both the initiation and progression of ER-positive breast cancers may be more effectively suppressed as compared with ER-negative tumors [28]. ER negativity of breast cancer is the crucial histochemical marker defining the poorest prognosis of the disease [61,62]. The antitumor capacity of preserved estrogen level in young women may be in correlation with the fact that estrogen is the ligand of ERs in highly differentiated tumors with better prognosis.

Correlations between Reproductive History and the Risk of Different Breast Cancer Subtypes

An additional puzzling phenomenon is the apparently controversial correlation between the risk of TNBC and reproductive factors in women. High TNBC risk in multiparous women with excessive estrogen signaling and low TNBC risk in nulliparous cases with defective

estrogen exposure are strongly supported by literary data [59,65,66]. These correlations are highly embarrassing if one considers the apparent independence of ER-negative tumors from the environmental concentration of estrogens.

High parity shows strong tumor protective effect even against the cancers of highly hormone dependent female organs including overall breast cancer, endometrial and ovarian tumors [67,68]. Recently, a significantly decreased overall cancer risk was reported after ovulation induction and in vitro fertilization assisted childbirth, mainly due to a lower than expected incidence of breast cancer [69].

Parity and particularly multiparity are associated with a decreased risk of the predominant ER-positive breast cancer type [65,70,71]. Among parous women, even the number of births was inversely associated with the risk of ER-positive breast cancer [59]. Among women who had at least four terminated pregnancies, a strongly decreased risk of ER-positive breast cancer was observed as compared with nulliparous cases (OR=0.55) [48].

Conversely, TNBC incidence exhibited an apparently unchanged ratio in parous women [48,64,71], whereas in certain studies even an increased risk of TNBC was reported in multiparous cases [72-74]. In a recent study the number of births was found to be directly associated with the risk of TNBC [75].

Nulliparity is generally in correlation with anovulatory disorders, thus these hormone deficient cases may be regarded as opposite extremes as compared with multiparous women [6]. Delayed first childbirth may also be associated with prolonged defective estrogen synthesis and ovulatory failures. Postpubertal sexual hormone imbalances are frequently associated with definite or prolonged fertility disorders resulting in nulliparity and delayed first childbearing [76] and inducing increased overall breast cancer risk among premenopausal women [77-78]. High overall breast cancer risk in correlation with defective estrogen synthesis and anovulatory disorders justifies the role of physiologic estrogen signaling in preservation of mammary health [6]. Administration of pregnancy mimicking estrogen and progesterone doses to nulliparous women seems to be a useful strategy for protection against breast cancer [79].

Certain studies suggested that nulliparity plays quite inverse role in the risks of ER-positive breast cancer and TNBC as compared with multiparity [66]. Nulliparous status of women was associated with a 35% higher risk of ER-positive breast cancer (HR=1.35), whereas with a 39% lower risk for TNBC (HR=0.61) [75]. Delayed first childbirth was also directly associated with risk for ER-positive cancers, but showed no remarkable effect on the risk of TNBCs [59].

Considering the apparently contradictory results, if women undertake more childbirth they may be exposed to a stronger risk of developing TNBC, conversely, if they remain nulliparous they are exposed to higher risk for ER-positive cancers. So what should they do?

In multiparous women, good fertility associated estrogen supply and excessive estrogen levels during pregnancies strongly and equivocally reduce the development of overall breast cancers and the predominant ER-positive tumors in particular. A plausible explanation is that estrogen, being the specific ligand for ERs, may preferentially block the development and progression of ER-positive cancers. However, its killing capacity against ER-negative cancers is slower and weaker as the specific tumor receptors are missing. In its complexity, in estrogen rich milieu a fairly decreased number and percentage of ER-positive tumors is associated with a moderately decreased number and an unchanged or deceptively increased percentage of ER-negative breast cancers [28].

By contrast, in nulliparous hormonally challenged women, the weakness of estrogen surveillance results in enhanced overall breast cancer risk. The insufficient estrogen supply has

a defective killing capacity even against the predominant, hormone sensitive ER-positive breast cancer cells resulting in an increased number and percentage of surviving ER-positive tumors. The survival possibility for hormonally weakly controlled ER-negative cancers like TNBCs may disproportionately improve and their raw number may be somewhat decreased, while the relative number (percentage) increases [80].

For women who are afraid of developing breast cancer, it is plausible to choose parity either by natural way or by in vitro fertilization to prevent the development of both TNBC and non-TNBC type tumors.

CONCLUSION

Considering the disproportionally lower tumor suppressor impact of estrogens on ER-negative breast cancers, BRCA gene mutation associated defective estrogen signaling may have close correlation with an easier deliberation of mitotic failures and development of poorly differentiated ER-negative tumors. In cases with BRCA gene mutation, increased aromatase activity, elevated estrogen levels and higher ligand-independent transcriptional activity of ERs are defensive endogenous mechanisms so as to improve the defective estrogen surveillance. In high risk women, the key for breast cancer prevention may be the restoration of triangle partnership among estrogens, estrogen receptors and DNA stabilizer BRCA genes.

REFERENCES

- [1] Miki Y, Swensen J, Shattuck-Eidens D, Futreal PA, Harshman K, Tavtigian S et al. A strong candidate for the breast and ovarian cancer susceptibility gene BRCA1. *Science* 1994; 266(5182): 66-71.
- [2] Wooster R, Bignell G, Lancaster J, Swift S, Seal S, Mangion J et al. Identification of the breast cancer susceptibility gene BRCA2. *Nature*. 1995; 378(6559):789-92.
- [3] Venkitaraman AR. Cancer susceptibility and the functions of BRCA1 and BRCA2. *Cell* 2002; 108: 171-182.
- [4] Farmer H, McCabe N, Lord CJ, Tutt AN, Johnson DA, Richardson TB, Santarosa M, Dillon KJ, Hickson I, Knights C, Martin NM, Jackson SP, Smith GC, Ashworth A. Targeting the DNA repair defect in BRCA mutant cells as a therapeutic strategy. *Nature*. 2005;434: 917-21.
- [5] Lux MP, Fasching PA, Beckmann MW. Hereditary breast and ovarian cancer: review and future perspectives. *J. Mol. Med. (Berl)* 2006; 84:16-28.
- [6] Suba Z. Circulatory estrogen level protects against breast cancer in obese women. *Recent Pat Anticancer Drug Discov*. 2013; 8(2):154-67.
- [7] Suba Z. Common soil of smoking-associated and hormone-related cancers: estrogen deficiency. *Oncol. Rev*. 2010; 4:73-87.
- [8] Brenton JD, Carey LA, Ahmed AA, Caldas C. Molecular classification and molecular forecasting of breast cancer: ready for clinical application? *J. Clin. Oncol*. 2005; 23(29): 7350-60.

- [9] Boyle P. Triple-negative breast cancer: epidemiological considerations and recommendations. *Ann. Oncol.* 2012; 23 Suppl 6:vi7-12.
- [10] Lakhani SR, Van De Vijver MJ, Jacquemier J, Anderson TJ, Osin PP, McGuffog L, Easton DF. The pathology of familial breast cancer: predictive value of immunohistochemical markers estrogen receptor, progesterone receptor, HER-2, and p53 in patients with mutations in BRCA1 and BRCA2. *J. Clin. Oncol.* 2002; 20(9): 2310-8.
- [11] Lee E, McKean-Cowdin R, Ma H, Spicer DV, Van Den Berg D, Bernstein L, Ursin G. Characteristics of triple-negative breast cancer in patients with a BRCA1 mutation: results from a population-based study of young women. *J. Clin. Oncol.* 2011; 29(33): 4373-80.
- [12] Atchley DP, Albarracin CT, Lopez A, Valero V, Amos CI, Gonzalez-Angulo AM, Hortobagyi GN, Arun BK. Clinical and pathologic characteristics of patients with BRCA-positive and BRCA-negative breast cancer. *J. Clin. Oncol.* 2008; 26(26): 4282-8.
- [13] Marquis S. T., Rajan J. V., Wynshaw-Boris A., Xu J., Yin G. Y., Abel K. J., Weber B. L., Chodosh L. A. The developmental pattern of BRCA1 expression implies a role in differentiation of the breast and other tissues. *Nat. Genet.* 11: 17-26, 1995.
- [14] Hennighausen L, Robinson GW: Think globally, act locally: the making of a mouse mammary gland. *Genes Dev* 1998, 12: 449-455.
- [15] Mueller CR, Roskelley CD. Regulation of BRCA1 expression and its relationship to sporadic breast cancer. *Breast Cancer Res.* 2003, 5:45-52.
- [16] Hilakivi-Clarke L. Estrogens, BRCA1, and Breast Cancer. *Cancer Res.* 2000; 60: 4993-5001.
- [17] Gudas J. M., Nguyen H., Li T., Cowan K. H. Hormone-dependent regulation of BRCA1 in human breast cancer cells. *Cancer Res.* 55: 4561-4565, 1995.
- [18] Spillman M., Bowcock A. BRCA1 and BRCA2 mRNA levels are coordinately elevated in human breast cancer cells in response to estrogen. *Oncogene*, 13: 1639-1645, 1996.
- [19] Adlercreutz HJ. Phytoestrogens and breast cancer. *Steroid Biochem. Mol. Biol.* 2002; 83(1-5): 113-8.
- [20] Bernard-Gallon D. J., Maurizis J. C., Rio P. C., Bignon Y. J. Influence of genistein and daidzein on BRCA1 protein levels in human breast cancer cell lines. *J. Natl. Cancer Inst.* 90: 862-863, 1998.
- [21] Cabanes A, Wang M, Olivo S, DeAssis S, Gustafsson JA, Khan G, Hilakivi-Clarke L. Prepubertal estradiol and genistein exposures up-regulate BRCA1 mRNA and reduce mammary tumorigenesis. *Carcinogenesis.* 2004; 25(5):741-8.
- [22] Xu X, Wagner KU, Larson D, Weaver Z, Li C, Ried T, Hennighausen L, Wynshaw-Boris A, Deng CX: Conditional mutation of Brca1 in mammary epithelial cells results in blunted ductal morphogenesis and tumour formation. *Nat. Genet* 1999, 22:37-43.
- [23] Russo J., Russo I. H. Toward a unified concept of mammary carcinogenesis Aldaz M. C. Gould M. N. McLachlan J. Slaga T. J. eds. *Progress in Clinical and Biological Research*,: 1-16, Wiley-Liss New York 1997.
- [24] Oktay K, Kim JY, Barad D, Babayev SN. Association of BRCA1 mutations with occult primary ovarian insufficiency: a possible explanation for the link between infertility and breast/ovarian cancer risks. *J. Clin. Oncol.* 2010; 28(2): 240-4.
- [25] Lin WT, Beattie M, Chen LM, Oktay K, Crawford SL, Gold EB, Cedars M, Rosen M. Comparison of age at natural menopause in BRCA1/2 mutation carriers with a non-clinic-based sample of women in northern California. *Cancer* 2013; 119(9): 1652-9.

- [26] Finch A, Valentini A, Greenblatt E, Lynch HT, Ghadirian P, Armel S, Neuhausen SL, Kim-Sing C, Tung N, Karlan B, Foulkes WD, Sun P, Narod S; Hereditary Breast Cancer Study Group. Frequency of premature menopause in women who carry a BRCA1 or BRCA2 mutation. *Fertil Steril* 2013; 99(6): 1724-8.
- [27] Suba Z. Diverse pathomechanisms leading to the breakdown of cellular estrogen surveillance and breast cancer development: new therapeutic strategies. *Drug Design Devel Ther.* 2014; 8:1381—1390.
- [28] Suba Z. Triple-negative breast cancer risk in women is defined by the defect of estrogen signaling: preventive and therapeutic implications. *Onco Targets Ther.* 2014; 7:147-64.
- [29] Hosey AM, Gorski JJ, Murray MM, Quinn JE, Chung WY, Stewart GE et al. Molecular basis for estrogen receptor alpha deficiency in BRCA1-linked breast cancer. *J. Natl. Cancer Inst.* 2007; 99(22):1683-94.
- [30] Bernstein LM. Endocrinology of the wild and mutant *BRCA1* gene and types of hormonal carcinogenesis. *Future Oncology.* 2008; 4(1): 23-39.
- [31] Fan S, Wang J, Yuan R, Ma Y, Meng Q, Erdos MR, Pestell RG, Yuan F, Auborn KJ, Goldberg ID, Rosen EM. BRCA1 inhibition of estrogen receptor signaling in transfected cells. *Science.* 1999; 284(5418): 1354-6.
- [32] Savage KI, Matchett KB, Barros EM, Cooper KM, Irwin GW, Gorski JJ et al. BRCA1 deficiency exacerbates estrogen-induced DNA damage and genomic instability. *Cancer Res.* 2014; 74(10):2773-84.
- [33] Fan S, Ma YX, Wang C, Yuan RQ, Meng Q, Wang JA, Erdos M, Goldberg ID, Webb P, Kushner PJ, Pestell RG, Rosen EM. Role of direct interaction in BRCA1 inhibition of estrogen receptor activity. *Oncogene.* 2001; 20(1): 77-87.
- [34] Buckley N. E., Conlon S. J., Jirstrom K., Kay E. W., Crawford N. T., O'grady A., et al. The DeltaNp63 proteins are key allies of BRCA1 in the prevention of basal-like breast cancer. *Cancer Res.* 2011; 71: 1933–1944.
- [35] Gorski J. J., Kennedy R. D., Hosey A. M., Harkin D. P. The complex relationship between BRCA1 and ER alpha in hereditary breast cancer. *Clin. Cancer Res.* 2009; 15: 1514–1518.
- [36] Chand AL; kConFab, Simpson ER, Clyne CD. Aromatase expression is increased in BRCA1 mutation carriers. *BMC Cancer.* 2009; 9:148. doi: 10.1186/1471-2407-9-148.
- [37] Simpson ER, Dowsett M. Aromatase and its inhibitors: significance for breast cancer therapy. *Recent Prog. Horm Res.* 2002; 57:317–338.
- [38] Lu M, Chen D, Lin Z, Reierstad S, Trauernicht AM, Boyer TG, Bulun SE. BRCA1 negatively regulates the cancer-associated aromatase promoters I.3 and II in breast adipose fibroblasts and malignant epithelial cells. *J. Clin. Endocrinol. Metab.* 2006; 91:4514–4519.
- [39] Ghosh S, Lu Y, Katz A, Hu Y, Li R. Tumor suppressor BRCA1 inhibits a breast cancer-associated promoter of the aromatase gene (CYP19) in human adipose stromal cells. *Am. J. Physiol. Endocrinol. Metab.* 2007; 292:246–252.
- [40] Miller WR, O'Neill J. The importance of local synthesis of estrogen within the breast. *Steroids.* 1987; 50(4-6):537-48.
- [41] Bollet MA, Savignoni A, De Koning L, Tran-Perennou C, Barbaroux C, Degeorges A, Sigal-Zafrani B, Almouzni G, Cottu P, Salmon R, Servant N, Fourquet A, de Cremoux P. Tumor aromatase expression as a prognostic factor for local control in young breast

- cancer patients after breast-conserving treatment. *Breast Cancer Res.* 2009; 11(4):R54. doi: 10.1186/bcr2343.
- [42] Suba Z. Failures and controversies of the antiestrogen treatment of breast cancer. In: Suba Z. Ed. *Estrogen prevention for breast cancer*. Chapter 6. Pp. 105-125. Nova Science Publishers Inc. New York, 2013.
- [43] Widschwendter M, Rosenthal AN, Philpott S, Rizzuto I, Fraser L, Hayward J, Intermaggio MP, Edlund CK, Ramus SJ, Gayther SA, Dubeau L, Fourkala EO, Zaikin A, Menon U, Jacobs IJ. The sex hormone system in carriers of BRCA1/2 mutations: a case-control study. *Lancet Oncol.* 2013; 14(12):1226-32.
- [44] Kim J, Oktay K. Baseline E(2) levels are higher in BRCA2 mutation carriers: a potential target for prevention? *Cancer Causes Control.* 2013; 24(3):421-6.
- [45] McLaughlin JR, Risch HA, Lubinski J, Moller P, Ghadirian P, Lynch H, Karlan B, Fishman D, Rosen B, Neuhausen SL, Offit K, Kauff N, Domchek S, Tung N, Friedman E, Foulkes W, Sun P, Narod SA; Hereditary Ovarian Cancer Clinical Study Group. Reproductive risk factors for ovarian cancer in carriers of BRCA1 or BRCA2 mutations: a case-control study. *Lancet Oncol.* 2007; 8(1): 26-34.
- [46] Antoniou AC, Rookus M, Andrieu N, Brohet R, Chang-Claude J, Peock S et al. Reproductive and hormonal factors, and ovarian cancer risk for BRCA1 and BRCA2 mutation carriers: results from the International BRCA1/2 Carrier Cohort Study. *Cancer Epidemiol. Biomarkers Prev.* 2009; 18(2):601-10.
- [47] Whittemore AS, Balise RR, Pharoah PD, Dicioccio RA, Oakley-Girvan I, Ramus SJ et al. Oral contraceptive use and ovarian cancer risk among carriers of BRCA1 or BRCA2 mutations. *Br. J. Cancer.* 2004; 91(11): 1911-5.
- [48] Ma H, Wang Y, Sullivan-Halley J, Weiss L, Marchbanks PA, Spirtas R, Ursin G, Burkman RT, Simon MS, Malone KE, Strom BL, McDonald JA, Press MF, Bernstein L. Use of four biomarkers to evaluate the risk of breast cancer subtypes in the women's contraceptive and reproductive experiences study. *Cancer Res.* 2010; 70(2): 575-87.
- [49] Cibula D, Zikan M, Dusek L, Majek O. Oral contraceptives and risk of ovarian and breast cancers in BRCA mutation carriers: a meta-analysis. *Expert Rev. Anticancer Ther.* 2011; 11(8):1197-207.
- [50] Milne RL, Knight JA, John EM, Dite GS, Balbuena R, Ziogas A et al. Oral contraceptive use and risk of early-onset breast cancer in carriers and noncarriers of BRCA1 and BRCA2 mutations. *Cancer Epidemiol. Biomarkers Prev.* 2005; 14(2):350-6.
- [51] Zheng L, Annab LA, Afshari CA, Lee WH, Boyer TG. BRCA1 mediates ligand-independent transcriptional repression of the estrogen receptor. *Proc. Natl. Acad. Sci. USA* 2001; 98(17): 9587-92.
- [52] Davis AA, Kaklamani VG. Metabolic syndrome and triple-negative breast cancer: a new paradigm. *Int. J. Breast Cancer.* 2012; 2012:809291. doi: 10.1155/2012/809291.
- [53] Davison Z, de Blacqui re GE, Westley BR, May FE. Insulin-like growth factor-dependent proliferation and survival of triple-negative breast cancer cells: implications for therapy. *Neoplasia.* 2011; 13(6): 504-15.
- [54] Maiti B, Kundranda MN, Spiro TP, Daw HA. The association of metabolic syndrome with triple-negative breast cancer. *Breast Cancer Res. Treat.* 2010; 121(2): 479-83.
- [55] Colonna SV, Douglas Case L, Lawrence JA. A retrospective review of the metabolic syndrome in women diagnosed with breast cancer and correlation with estrogen receptor. *Breast Cancer Res. Treat.* 2012; 131(1): 325-31.

- [56] Amadou A, Hainaut P, Romieu I. Role of obesity in the risk of breast cancer: lessons from anthropometry. *J. Oncol.* 2013; 2013:906495. doi: 10.1155/2013/906495.
- [57] Harris HR, Willett WC, Terry KL, Michels KB. Body fat distribution and risk of premenopausal breast cancer in the Nurses' Health Study II. *J. Natl. Cancer Inst.* 2011; 103(3): 273-8.
- [58] Amirikia KC, Mills P, Bush J, Newman LA. Higher population-based incidence rates of triple-negative breast cancer among young African-American women : Implications for breast cancer screening recommendations. *Cancer* 2011; 117(12):2747-53.
- [59] Phipps AI, Malone KE, Porter PL, et al. Reproductive and hormonal risk factors for postmenopausal luminal, HER-2-overexpressing, and triple-negative breast cancer. *Cancer.* 2008; 113(7):1521–1526.
- [60] Bauer KR, Brown M, Cress RD, Parise CA, Caggiano V. Descriptive analysis of estrogen receptor (ER)-negative, progesterone receptor (PR)-negative, and HER2-negative invasive breast cancer, the so-called triple-negative phenotype: a population-based study from the California cancer Registry. *Cancer.* 2007; 109(9): 1721-8.
- [61] Hartley MC, McKinley BP, Rogers EA, Kalbaugh CA, Messich HS, Blackhurst DW, Lokey JS, Trocha SD. Differential expression of prognostic factors and effect on survival in young (< or =40) breast cancer patients: a case-control study. *Am Surg.* 2006; 72(12): 1189-94.
- [62] Talley LI, Grizzle WE, Waterbor JW, Brown D, Weiss H, Frost AR. Hormone receptors and proliferation in breast carcinomas of equivalent histologic grades in pre- and postmenopausal women. *Int J Cancer.* 2002; 98(1):118-27.
- [63] Gaudet MM, Press MF, Haile RW, Lynch CF, Glaser SL, Schildkraut J, Gammon MD, Douglas Thompson W, Bernstein JL. Risk factors by molecular subtypes of breast cancer across a population-based study of women 56 years or younger. *Breast Cancer Res Treat.* 2011; 130(2): 587-97.
- [64] Dolle JM, Daling JR, White E, Brinton LA, Doody DR, Porter PL, Malone KE. Risk factors for triple-negative breast cancer in women under the age of 45 years. *Cancer Epidemiol. Biomarkers Prev.* 2009; 18(4): 1157-66.
- [65] Palmer JR, Boggs DA, Wise LA, Ambrosone CB, Adams-Campbell LL, Rosenberg L. Parity and lactation in relation to estrogen receptor negative breast cancer in African American women. *Cancer Epidemiol. Biomarkers Prev.* 201; 20(9): 1883-91.
- [66] Yang XR, Chang-Claude J, Goode EL, Couch FJ, Nevanlinna H, Milne RL, Gaudet J. Associations of breast cancer risk factors with tumor subtypes: a pooled analysis from the Breast Cancer Association Consortium studies. *J. Natl. Cancer Inst.* 2011; 103(3): 250-63.
- [67] Key TJ. Hormones and cancer in humans. *Mutat Res.* 1995; 333(1-2): 59-67.
- [68] Britt K, Ashworth A, Smalley M. Pregnancy and the risk of breast cancer. *Endocrine-Related Cancer.* 2007; 14(4): 907-33.
- [69] Källén B, Finnström O, Lindam A, Nilsson E, Nygren KG, Olausson PO. Malignancies among women who gave birth after in vitro fertilization. *Hum. Reprod.* 2011; 26(1):253-8.
- [70] Kwan ML, Kushi LH, Weltzien E, Maring B, Kutner SE, Fulton RS, Lee MM, Ambrosone CB, Caan BJ. Epidemiology of breast cancer subtypes in two prospective cohort studies of breast cancer survivors. *Breast Cancer Res* 2009; 11(3): R31.

- [71] Trivers KF, Lund MJ, Porter PL, Liff JM, Flagg EW, Coates RJ, Eley JW. The epidemiology of triple-negative breast cancer, including race . *Cancer Causes Control*. 2009; 20(7):1071-82.
- [72] Yang XR, Sherman ME, Rimm DL, Lissowska J, Brinton LA, Peplonska B, Hewitt SM, Anderson WF, Szeszenia-Dabrowska N, Bardin-Mikolajczak A, Zatonski W, Cartun R, Mandich D, Rymkiewicz G, Ligaj M, Lukaszek S, Kordek R, García-Closas M. Differences in risk factors for breast cancer molecular subtypes in a population-based study. *Cancer Epidemiol. Biomarkers Prev*. 2007; 16(3): 439-43.
- [73] Millikan RC, Newman B, Tse CK, et al. Epidemiology of basal-like breast cancer. *Breast Cancer Res Treat*. 2008; 109(1): 123–139.
- [74] Xing P, Li J, Jin F. A case-control study of reproductive factors associated with subtypes of breast cancer in Northeast China. *Med. Oncol*. 2010; 27(3): 926–931.
- [75] Phipps AI, Chlebowski RT, Prentice R, McTiernan A, Wactawski-Wende J, Kuller LH, Adams-Campbell LL, Lane D, Stefanick ML, Vitolins M, Kabat GC, Rohan TE, Li CI. Reproductive history and oral contraceptive use in relation to risk of triple-negative breast cancer. *J. Natl. Cancer Inst*. 2011; 103(6): 470-7.
- [76] Velentgas P, Daling JR. Risk factors for breast cancer in younger women. *J. Natl. Cancer Inst. Monogr* 1994; (16): 15-24.
- [77] Opdahl S, Alsaker MD, Janszky I, Romundstad PR, Vatten LJ. Joint effect of nulliparity and other breast cancer risk factors. *Br. J. Cancer* 2011; 105(5): 731-6.
- [78] Newcomb PA, Trentham-Dietz A, Hampton JM, Egan KM, Titus-Ernstoff L, Warren Andersen S, et al. Late age at full term birth is strongly associated with lobular breast cancer. *Cancer* 2011; 117(9): 1946-56.
- [79] Tsubura A, Uehara N, Matsuoka Y, Yoshizawa K, Yuri T. Estrogen and progesterone treatment mimicking pregnancy for protection from breast cancer. *In Vivo*. 2008; 22(2):191-201.
- [80] Suba Z. Triple negative breast cancer risk in women is increasing by the loss of estrogen supply or by defective estrogen signaling. In: Suba Z. Ed. *Recent Avenue to Cancer Prevention*. Chapter 6. Pp. 119-147. Nova Science Publishers Inc. New York, 2014.

Chapter 48

GENOTYPE-PHENOTYPE RELATIONSHIPS IN LANGUAGE PROCESSES IN RETT SYNDROME

***Alessandra Falzone, Antonio Gangemi
and Rosa Angela Fabio****

Department of Cognitive Science, University of Messina, Messina, Italy

ABSTRACT

Rett syndrome (RTT) is a neurodevelopmental disorder mainly caused by mutations in the MECP2 gene affecting around 1 in 10,000 female births. Mutations in the MECP2 gene have been associated with the onset of RTT. Clinical manifestations include severe linguistic and motor impairments that are the core of phenotype symptoms. Some patients show a moderate level of conservation of linguistic functions while others lose the use of functional verbal communication. The objectives of the present chapter are to study in depth the latest theoretical approaches to the link between linguistic processes and the specific RTT genotype.

This chapter begins with a theoretical overview on cognitive alterations and then focuses on linguistic specific impairments characterized by the loss of articulation or the production of few functional sounds. A restricted sample shows the presence of verbal speech (Preserved Speech Variant). Renieri et al. (2009) proposed the term “Zappella variant” rather than “preserved speech variant” to describe milder forms of RTT, because other aspects, besides speech, are involved.

The second part proposes a preliminary research which analyses the correlation between linguistic phenotype and specific genotype.

1. INTRODUCTION

Complex diseases are a set of pathologies for which environmental and genetic factors act together to form a phenotypic manifestation and, unlike monogenic diseases, don't present a standard model of Mendelian inheritance. The phenotype of monogenetic diseases can be

* Corresponding Author's Email: rafabio@unime.it.

complex, but in any case, it depends on the action of a single gene, while in a complex disease, different phenotypic components depend on interactions of more genes that interact with each other in an epistatic way. The difficulty in studying these complex pathologies consists in correlating different phenotypic components to a single specific genetic / environmental factor (Elston, 1995).

1.1. Rett Syndrome (RTT)

Rett syndrome is a neurodevelopmental disorder, affecting around 1/10.000 female births.

In 1998, in a study on genetic RTT cases, Xq28, a critically important region was identified. In 1999, the mutational screening of candidate genes in the region allowed the identification of MECP2 as a cause of the standard form. Ten years later, different research studies regarding the causes of RTT variations have also been carried out.

In 2000, it was also shown that the “Zappella Variant” is caused by mutations in the MECP2 gene.

Another study in 2005 demonstrated that a second gene called CDKL5 localised in the X chromosome, is involved in the variant with early onset convulsions.

Recently, the FOXP1 gene, localized in chromosome 14 has been identified as the first autosomal gene associated to RTT, in particular to the congenital variant.

These results demonstrated that RTT presents genetic and clinical heterogeneity and they provide data for molecular bases to understand pathogenic mechanisms of the disease and to establish targeted therapeutic strategies.

1.2. Standard Form of Clinical Features

RTT presents a clinical characteristic course that we can divide into four levels.

Prenatal and perinatal history is normal although, retrospectively, mild factors of the disease as stereotyped occasional movements or dystonic postures can be observed.

After a period of about 6-18 months, females present a developmental arrest (level I, stagnation), followed by a phase of regression (level II). At this level (1-4 years), they lose previously acquired skills as the hand use and verbal communication.

A rapid decline of social interactions, associated with the appearance of autistic traits is also evident. Females manifest involuntary stereotyped movements of the hand such as torsion, “hand washing” as well as bruxism, breathing irregularities such as apneas and hyperventilation. At this level, head growth slows down, which often results in microcephaly (improperly defined acquired microcephaly). At the next level (level III), called “pseudo stagnation” (4-7 years), there is a decrease in autistic symptomatology and an improvement in social interactions, even though the inability to speak, apraxia and manual stereotypies persist.

Scoliosis and somatic underdevelopment becomes more evident and seizures often occur.

The fourth and final level (5-15 years) is characterized by a global progressive deterioration that can extend to the condition of spastic quadriplegia (IV stage of late motor degeneration).

1.3. Clinical Features of Variant Form

In addition to the standard form, five RTT variants that differ in age of onset and severity of symptoms have also been described.

The “Zappella Variant” (Z-RTT) or Preserved speech Variant, described for the first time by Zappella in 1992, is the most common. This variant presents a more favorable clinical course. Differently from standard RTT patients, they typically present a normal head size; kyphoscoliosis is milder and somatic hypoevolution is reduced, sometimes with a trend of being overweight.

During the third stage, patients regain some previously lost skills, such as verbal communication. Some patients are able to say only a few words, while others are also able to communicate with complex sentences. We can observe an improvement in the use of hands, although considerable dyspraxia persists and the classic stereotyped movements are present. Their motor skills improve to the point that some girls are also able to go up and down stairs independently.

In the variant with early onset convulsions, described for the first time by Hanefeld in 1985, the initial period is masked by the onset of convulsions, usually in the form of flexor spasms. The onset of convulsions is accompanied by a delayed psychomotor development. Only later, girls develop the typical features of RTT as manual stereotypies, above all “Hand-Mouth”. Furthermore, the head size, weight and height are normal in most cases.

In the congenital variant, the psychomotor delay is evident from the first months of life, often with hypotonia and early alterations of EEG. In the following months the different levels previously described in the standard syndrome appear and some generalized convulsions can also appear.

The “Forme Frusta” are RTT variants that don’t have the typical features of the disease. Generally, the first level appears later (1-3 years). Females show initial milder symptoms and present a protracted clinical deterioration.

Usually, they retain some form of communication and developmental abnormalities are much less obvious. The classic stereotypes of the hands may be atypical or absent.

The clinical picture become more similar to RTT when these females reach adolescence and adulthood.

There are very few females with a late variant regression. In these cases, the first level is protracted and regression may arise during primary school.

2. GENOTYPE CHARACTERIZATION

Currently, gene MECP2 mutations [Xq28] are found in the majority of cases (90%) of standard Rett and in 30% of cases of atypical Rett. Since there are many clinically documented RETT cases who don’t present the MECP2 mutation, many scientists in this field believe that the pathology is caused by genetic heterogeneity. Between 2004 and 2005 mutations were identified in another gene; CDKLS, situated in the X chromosome near the Xp22 cytogenetic band in subjects affected by an atypical Rett form characterized by the early onset of resistance to the epilepsy drug (Hanfield’s variant). (Weaving et al. (2004); Tao et al. (2004); Scala et al. (2005)).

MECP2 gene was described for the first time in 1992 (Meehan et al. (1992)). MECP2 is situated on the long arm of chromosome X in the vicinity of the cytogenetic band q28 and it is subject to X inactivation (Adler et al. (1995); D'Esposito et al. (1996); Vilain et al. (1996)).

MECP2 gene occupies a part of genomic DNA of about 76 Kb, it is transcribed in telomere – centromero. Though it is expressed ubiquitously, its levels of expression appear to be regulated in manner, tissue and specific development particularly at the neuronal level. (Jung et al. (2003); Balmer et al. (2003); Castelli et al. (2013); Cohen et al. (2003); Kishi et al. (2004); Mullaney et al. (2004); Shahbazian et al. (2002b)).

MECP2 counts four exons that code 2 different isoforms of the protein, indicated as MeCp2, due to alternative splicing of exon 2. The two isoforms MeCp2 therefore differ only at the N terminal. The most abundant isoform, MeCP2B, contain exons 1-3 e 4, isoform Mecp2 A contain exons 2, 3 e 4. (Kriaucionis and Bird (2004); Mnatzakanian et al. (2004)). The portion 3°UTR of exon 4 of this gene is unusually long (approximately 8.5 Kb) and phylogenetically well preserved. (Coy et al. (1999)).

2.1. Mutations in MECP2 Gene

To date, at least 300 different mutations of this gene have been identified. The spectrum of abnormalities identified include missense, nonsense, frameshift mutations, as well as large deletions and duplications of the entire gene.

The MECP2 mutations are distributed along the entire coding sequence of the gene and 70% of alteration attributable to at least 8 variants, all transitions C > T (Lee et al. (2001)), presumably resulting from spontaneous deamination of cytosine residues in the nucleotide level:(c.316C>T, c.397C>T, c. 473C>T, c.502C>T, c.763C>T, c. 808C>T, c. 880C>T, c.916C>T).

2.2. MECP2 Proteine's Structure

MeCp2 protein (Swiss-Prot P51608) is situated in the nucleus and it consists of three functional domains:

- A Methyl-binding domain (MBD), 85 amino acids
- A Transcriptional Repression Domain (TRD) that counts 100 amino acids (amino acids 207 – 310) important in recruiting other components of the repressor complex. (Nan et al. (1997));
- A C – terminal domain, whose function has not yet been characterized.

Protein MeCp2 is part of a complex transcriptional repressor that contains the Sin3A corepressor and histone deacetylase HDAC1 and HDAC2. The functioning model of this complex predicts that MeCP2 repress the transcription through a mechanism that involves the binding to residues and recruitment of co repressor Sin3A and HDACs to change the structure of the chromatin.

MeCP2 recognizes and binds CpG methylated sites present in the promoters of the gene target, HDACs deacetyl histone tails and this consents DNA to wrap itself around nucleosomes causing the compaction of chromatin. In these conditions, the transcriptional machine loses access to the promoter and the genome is no longer transcribed (Jones et al. (1998), Nan et al. (1998)).

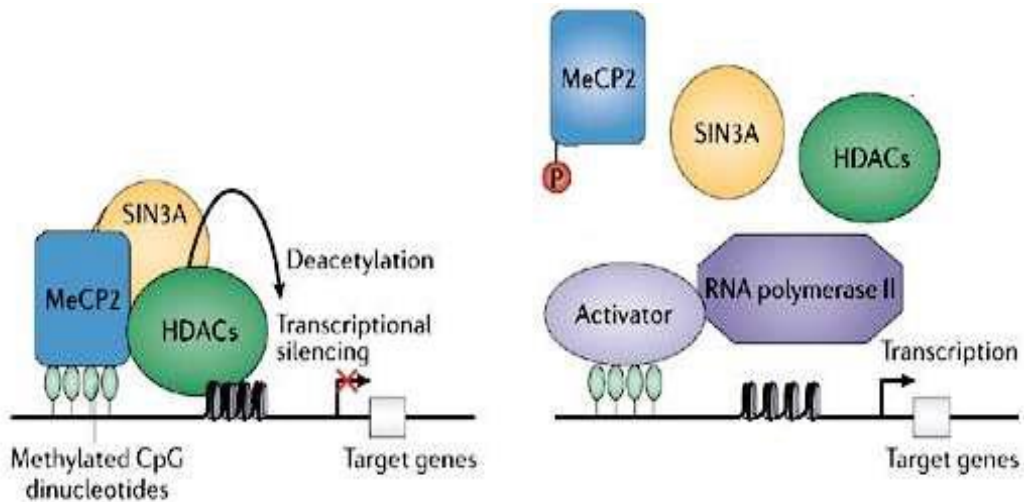


Figure 1. Operation model of the protein MecP2. MeCP2 recognizes and binds CpG and, through the recruitment of cofactors SIN3A and HDACs causes deacetylation of histones and thus the compaction of DNA around nucleosomes. In this way, target genes are not transcribed.

3. PHENOTYPES WITH GENE MUTATION MECP2

Protein MeCP2 has a fundamental role in the repression of the transcription of numerous target genes during neuronal differentiation therefore this mechanism would explain why MeCP2 alterations were detected in different phenotypes in subjects of both sexes characterized by neurodevelopmental disorders.

Alterations have been identified in the autistic spectrum, and also in subjects with Angelman syndrome.

In fact, the full range of phenotypes in which alterations in MECP2 may have a pathogenic role remains to be investigated; it can be hypothesized that alterations of this gene are the basis of both diseases with clinical features that fall within the phenotypic Rett spectrum as well as other forms of pervasive developmental disorders.

Knowledge regarding the importance of the involvement of alterations of this gene is essential to understand the role of the MECP2 protein in a complex functional network that underlies the correct neuronal maturation and to understand better the correlation of alteration with specific phenotypes in the Rett syndrome.

More than 200 different mutations of the MECP2 gene have been reported in the Rett Base (IRSA MECP2 Variation Database; but eight mutations (Arg106Trp, Arg133Cys, Thr158Met, Arg168X, Arg255X, Arg270X,

Arg294X, Arg306Cys) affect around 67% of RTT females. A remaining 10% of RTT cases show a large group of C-terminal frameshift mutations.

Several studies have reported genotype-phenotype correlations, but with conflicting results. Most authors reporting data from different cohorts of RTT patients demonstrated that no correlation exists between missense vs. truncating mutations, whereas others reported that the truncating defects are more severe than the missense ones. Studies aimed at comparing mutations affecting the different functional domains share the opinion that defects affecting the C-terminal domain give a milder clinical score. Differences in clustering the mutations, the heterogeneity in the size of the analyzed cohorts, the selected clinical parameters, and variation in the age of the subjects are likely to explain the conflicting results.

Another important factor modulating phenotype expression is the randomization of X-inactivation; when present, it is expected to influence phenotypical severity. Nevertheless, most studies reported that RTT patients' lymphocytes fail to show a skewed X chromosome inactivation (XCI)

Yet, Knudsen found a significant increase of skewed XCI in RTT patients and their mothers. However, the inclusion of this data in genotype-phenotype correlation studies remains controversial because of the bias of tissue mosaicism. In fact, the brain cannot respect the same ratios as those found in the blood or fibroblasts, which are the investigated tissues. Moreover, a role of SNPs and CNVs as modifier elements of phenotype in patients carrying the same MECP2 mutations should be considered.

Given that conflicting findings about genotype-phenotype relationships in RTT are still under discussion, it seems worthwhile to further investigate this relationship in order to overcome some methodological flaws

In a previous study, Fabio et al. (2014) examined the effect of MECP2 mutations on the phenotypic variability within a group of 114 RTT patients, focusing on specific methodological issues. More precisely, the study was performed taking into account what was recommended by Ham et al. (2005) concerning the weak points of the previous studies. Firstly, biases originating from missing data, multiple testing, and age differences were minimized. Secondly, a main new feature was added to this analytic examination of the RTT phenotype, by using the Rett Assessment Rating Scale (R.A.R.S.), a specific instrument devised to measure the intensity of RTT symptoms and to provide a specific RTT behavioral profile. The specific aim of this study was to correlate disease-causing genetic mutations to the variety and intensity of the detailed symptomatic parameters assessed by R.A.R.S. and to provide insights into the effect of MECP2 mutations on RTT phenotype. The results showed that a specific kind of genotypes can be associated with the severity of symptoms showed by RTT patients. On the basis of these results, our study aims at providing a phenotype/genotype correlation in relation to a specific cognitive process, language production and comprehension.

4. LANGUAGE IN RETT SYNDROME

After its nosographical description, RTT was classified among Pervasive Developmental Disorders (APA, 2000), but it has been moved to the genetic disorder category because of its primary ethiology (DSM V, APA 2013).

An evident feature of Rett Syndrome is that the disorder is characterized by severe alteration in motor and speech capabilities that are considered as inclusive diagnostic criteria (Neul et al., 2010).

In particular, the residual linguistic abilities are very different and each individual could manifest various degrees of severity in language production: whereas some RTT patients completely lose their activity in verbal sound production, which is functional to communication, others preserve functional vocal sound and/or words (Budden, 1997; De Bona et al. 2000; Fabio et al., 2009; 2011; 2013; Zappella et al. 2001).

Linguistic deficits typically arise after the regression phase: indeed, RTT females are characterized by a normal language development comparable with a healthy one before the regression phase. They often exhibit babbling and phonological coupling except for the early onset variant.

It has been demonstrated that linguistic competence levels are correlated to the language acquisition stage in which individuals were at the onset of regression (Marschik et al., 2012).

It has been found that residual linguistic capabilities can be considered as an outcome predictor for the syndrome: the greater the verbal ability before regression, the more likely it is that patients can regain speech functionality to communication and interaction during the pseudo-stationary stage.

There are different linguistic phenotype variants in RTT. A restricted sample shows the presence of verbal speech (Preserved Speech Variant). Renieri et al. (2009) proposed the term “Zappella variant” rather than “preserved speech variant” to describe milder forms of RTT, because other aspects, besides speech, are involved. This form presents a less severe clinical condition, i.e., regular skull dimensions and a relevant reduction of epileptic seizures and breathing alterations.

Few studies have evaluated a genotype/phenotype correlation between linguistic residuals and specific genotype. Uchino et al. (2001) tried to correlate the grade of disability in locomotion and that of microcephalus with a language disability in RTT in a preliminary study and after they correlated language RTT to the loci of MECP2 mutation. This correlation was found on the basis of a qualitative evaluation of spoken language.

This study proposes a genotype/linguistic phenotype correlation based on an articulation capability test by a phonetic evaluation test (Fanzago, 1983). Our hypothesis moves off the assumption that linguistic alterations in RTT derive from alterations in breathing and facial-laryngeal muscle's coordination rather than from general motor disease. As a consequence, genotypes producing severe breathing and facial-laryngeal muscles alteration are expected to show a severe linguistic phenotype.

Indeed, girls with RTT show a complex breathing phenotype that could include hypoventilation, hyperventilation, apneic episodes with clusters of arrhythmic breathing, and breath hold terminated by Valsalva maneuvers (Katz et al. 2009). It seems that modifications of subcortical nuclei in the brainstem, which regulate breathing rhythm, are connected to speech motor control.

It is supposed that this relationship is due to the same neurological altered basis: both respiratory capacity and speech are controlled by subcortical nuclei in the brainstem and they are affected in Rett patients (Ogier et al. 2008, Ramirez et al. 2013). Many studies show that the functional integrity of the brainstem respiratory network (Trevvarthen & Daniel, 2005; Katz et al. 2009; Ogier & Katz, 2008) is affected, but to date this aspect has not yet been analyzed for speech. On the other hand, many studies in the normal and neuropsychological population

show that brainstem nuclei (i.e., basal ganglia, pontine reticular nuclei, substantia nigra, central gray of the midbrain, frontal cortex, caudate, putamen, globus pallidus and thalamus (Deguchi et al., 2000; Saito et al., 2001) as well as part of PNS (i.e., spinal trigeminal tract, solitary tract and nucleus) are involved in both breathing control and in language production (Falzone 2014) (Figure 2)

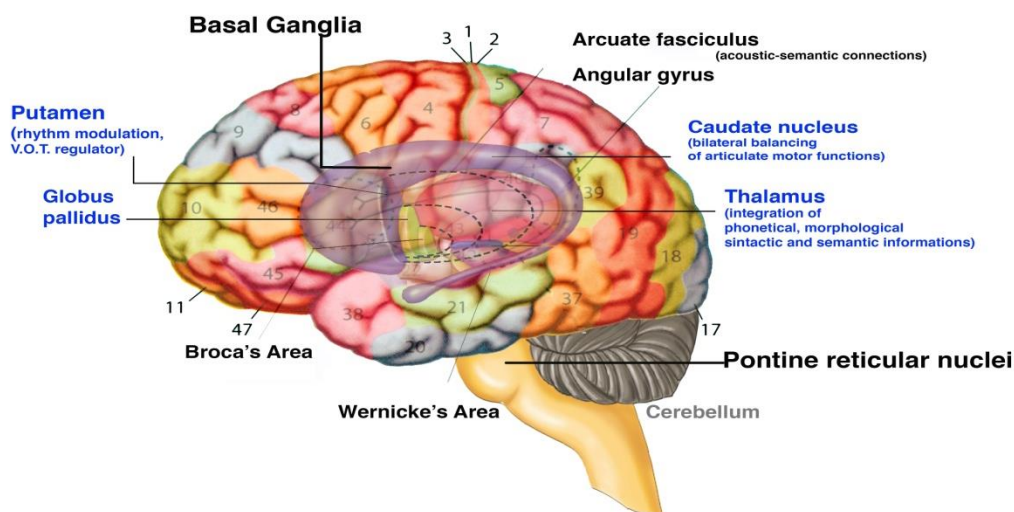


Figure 2. Cortical and subcortical network for language. The subcortical network for language involves the same brainstem nuclei for breathing control and the respiratory rhythm (i.e., basal ganglia, pontine reticular nuclei, substantia nigra, central gray of the midbrain, caudate, putamen, globus pallidus and thalamus (cf. Falzone, Anastasi, Pennisi, 2014).

Furthermore, there's no doubt that the linguistic phenotype of Rs breathing relies on arousal, i.e., breath and speech alterations can be exacerbated if cognitive arousal increases (Katz et al., 2009).

Some studies have shown that girls with RTT have a modified breathing brainstem network (Trevvarthen & Daniel, 2005; Katz et al. 2009; Ogier & Katz, 2008) but no one has correlated this aspect to articulated language alteration in RS.

On the basis of previous research on residual linguistic capacity in RS patients and on the basis of the evaluation of residual communication ability (both comprehension and articulation) (cf. Fabio et al., 2009) real linguistic phenotype to genotype has been correlated.

In order to do this, speech ability has to be evaluated by using the Fanzago test, a phonetic evaluation instrument which follows the typical degree of language acquisition phase in phonological difficulties aspects. After this, the spontaneous production of language sounds and the presence of breathing alteration at baseline condition and during a cognitive task were evaluated. The more arousal rises, the more that breathing alteration is expected to increase. As a result, even speech ability worsens.

5. METHODS

5.1. Participants

Twenty-one girls with a diagnosis of RTT, ranging from age 4 to 31 (mean= 16,34 years, SD=5,98), took part in the experiment. Their families had been contacted by the Italian Rett Association, which asked them to participate in the study. Twenty girls were met in Tuscany during a summer campus organized by the Association and one girl's parents were interviewed at the Hospital of the Messina University Hospital. According to the paperwork provided by the parents, all the patients examined for MECP2 mutations were positive. All of the participants were diagnosed with Rett Syndrome following both the guidelines established by the Criteria Work Group and the mutation analysis of the methyl-CpG binding protein 2 gene.

A general assessment was carried out by a psychologist through the Vineland Adaptive Behavior Scale (VABS) (Sparrow, Balla, & Cicchetti, 1984), the standardized test for the Rett Assessment Rating Scale (RARS, Fabio et al., 2005) and Modified Colored Progressive Matrices.

The Fanzago phonetic articulation test was administered to evaluate the status of vocal sound articulation and objective production articulated voice functional to communication.

During the cognitive assessment, behavioural breathing parameters were evaluated in order to verify the presence of hypoventilation, hyperventilation, apnea, and breath hold. All the girls showed breathing alteration and none required any Valsalva maneuvers.

5.2. Material

Functional scales. The Vineland Adaptive Behavior Scales are used to diagnose intellectual and developmental disabilities. The Scales are organized into four domains: Communication (Receptive, Expressive, Written); Daily Living (Personal, Domestic, Community); Socialization (Interpersonal Relationships, Play and Leisure Time, Coping Skills); and Motor Skills (Gross, Fine).

The Rett Assessment Rating Scale (RARS) is a standardized scale used to evaluate subjects with Rett's syndrome (Fabio et al., 2005). It is constructed by following the diagnostic criteria for RTT proposed by DSM-IV-TR (APA, 2010) and recent research and clinical experience. It follows a structure similar to that used for the diagnosis of the pervasive developmental disorders included in the same nosographical category as RTT (i.e., Childhood Autism Rating Scale, CARS).

A total of 31 items was generated as representatives of the profile of RTT. Each item concerns a specific phenotypic characteristic and describes four increasing levels of severity. Each item is provided with a brief glossary explaining its meaning in a few words. Each item is rated on a 4-point scale, where 1 = within normal limits, 2 = infrequent or low abnormality, 3 = frequent or medium- high abnormality, and 4 = strong abnormality. Intermediate ratings are possible; for example, an answer between 2 to 3 points is rated as 2.5. For each item, the evaluator circles the number corresponding to the best description of the patient. After a patient has been rated on all 31 items, a total score is computed by summing the individual ratings.

This total score allows the evaluator to identify the level of severity of RTT, conceptualized as a continuum ranging from mild symptoms to heavy deficits.

Cognitive Measures

Modified Raven's Coloured Progressive Matrices were used (Antonietti, Castelli, Fabio, Marchetti, 2003). Differently from the standard Raven's Colored Progressive Matrices (1940), in this adapted scale the A series was administered to girls and each table was larger (42 cm x 29,7 cm). Girls have to choose between two items (one target and one distractor) placed separately in front of them. Both items (target and distractor) were shown 3 times and the spatial position of the target and distractor was randomized. When the girl replied with two consecutive and correct answers the examiner presented the following table; when the girl replied wrongly three times, the test was interrupted.

Linguistic Measures

The Fanzago phonetic articulation test (1983) is used to evaluate the articulation capabilities in children and it is a good measure of their phonetic development. This instrument is based on spontaneous/repetition elicited denomination of 114 figures grouped in 22 tables. Each table represents one image whose name starts with a vocal sound and other objects in which the same sound is placed in a second or third position or is coupled with a vowel or consonant sounds. This study uses items which represent perceptively salient and commonly used objects for RTT.

This test proposes an ontogenetic approach to speech sound production and it groups the phonemes on manner articulation (occlusive, fricative, affricate, nasal, lateral, vibrant).

Target phonemes are reported in a specific template and it's possible to indicate if it has been produced correctly, omitted or distorted. If girls utter the entire word painted in the image during spontaneous or elicited production, it can be marked apart.

Respiratory Evaluation

Breathing parameters were evaluated in order to verify the presence of hypoventilation, hyperventilation, apnea, and breath hold. RTT girls typically present an altered respiratory rhythm. Some studies report alteration during both sleep and when awake. For sleep evaluation, previous respiratory evaluation, the clinical data, and parent reports were used. The respiratory rhythm was measured through behavioural analysis.

All the girls were video taped, two observers independently transcribed the first five minutes of each tape (one during sleep and one in the waking state). The final transcription was then coded independently by both observers for the behavioural breathing parameters. The inter-rater agreement concordance was high (Kappa index = .98).

5.3. Procedure

All the activities were performed in a setting suitable for language activity with patients: all distracting stimuli were removed so the girls focused only on the task.

After the initial assessment, each girl was evaluated at breathing baseline. After the Fanzago test was administered starting with spontaneous vocal production. In the second step girls had to produce phonemes at the request of the linguistic therapist. In this study, girls are requested to produce first vowel sounds, that are easier to articulate and that appear early on during normal development. Then the figures with objects whose name starts with consonants were shown (starting with easier sounds like /p/ /n/ /b/ /m/ and following by more complex ones /r/ /z/). In this study only images with objects starting with phoneme targets were presented, whereas words with phoneme targets in a second or third position were not shown. All correct phoneme (spontaneous/elicited) production was marked on the specific template, in which therapists can write if girls utter the entire word painted in the image during spontaneous or elicited production.

In order to check the effective results of the requested sound production task, the alteration of breathing rhythms during a cognitive, not linguistic task was evaluated in this study.

5.4. Results

Results are presented in relation to both the breathing problems and the language ability.

With reference to the first analysis, the respiratory rhythm was normal during sleep and abnormal in the waking state. To calculate the abnormality of the breathing in the waking state, the sum of each type of respiratory dysfunctions for five minutes was analyzed. With reference to the second analysis, four parameters of the Fanzago test were used: the number of vowels spontaneously produced, the number of consonants spontaneously produced, the number of vowels with elicited denomination and the number of consonants with elicited denomination.

Before proceeding with the analysis of the genotype-phenotype correlation, Pearson's correlation coefficient between the sum of each type of respiratory dysfunctions and the four parameters of Fanzago test were calculated. Results show that breathing problems display an inverse correlation with each of the Fanzago parameters, namely with the number of vowels spontaneously produced ($r(21) = -.378$, $p < .137$), the number of consonants spontaneously produced ($r(21) = -.61$, $p < .001$), the number of vowels with elicited denomination ($r(21) = -.498$, $p < .03$), and the number of consonants with elicited denomination ($r(21) = -.29$, $p < .23$).

To proceed with the genotype-phenotype correlation, since the number of participants was low, dichotomized scores to classify participants into a particular breathing type were used. Based on high ($>$ median) or low ($\leq$ median) scores on respiratory rhythm ($Mdn = 1.2$), mild breathing problems ($Mdn = 0.6$), and severe breathing problems ($Mdn = 2.8$), participants were placed within one of the two type categories. Since some of the girls with RTT shows only clinical features and not mutation in MECP2, only 14 patients were included in the analysis.

As shown in figure 3, patients with a truncating mutation after NLS manifested a lower degree of impairment than patients with a truncating mutation within NLS in the breathing dysfunctions ($\chi^2(2, N = 14) = 1.74$, $p < .05$).

With reference to the second analysis, the genotype-phenotype correlation was carried out with four parameters of the Fanzago test: the number of vowels spontaneously produced, the number of consonants spontaneously produced, the number of vowels with elicited denomination and the number of consonants with elicited denomination. No girl produced single consonants, but consonants together with vowels, i.e., syllables. To calculate these correlations, the sum of each type of syllable and the sum of each type of vowels was

considered. Again, we used dichotomized scores to classify participants into a particular category. Based on high ($>$ median) or low ($\leq$ median) scores on the number of vowels spontaneously produced (Mdn = 2), low level (Mdn = 0.6), and high level (Mdn = 2,8), participants were placed within one of the two type categories. Because some of the girls with RTT show only clinical features and not mutation in MECP2, again, only 14 patients were included in the analysis. As shown in figure 4, patients with a truncating mutation after NLS show a similar pattern to those observed in patients with a truncating mutation within NLS ($p<.48$). Since the Italian language has only 7 vowels (in many regional variations, only 5 with phonological differentiations), it may be that a floor effect does not permit the performances of the girls within each domain to be differentiated.

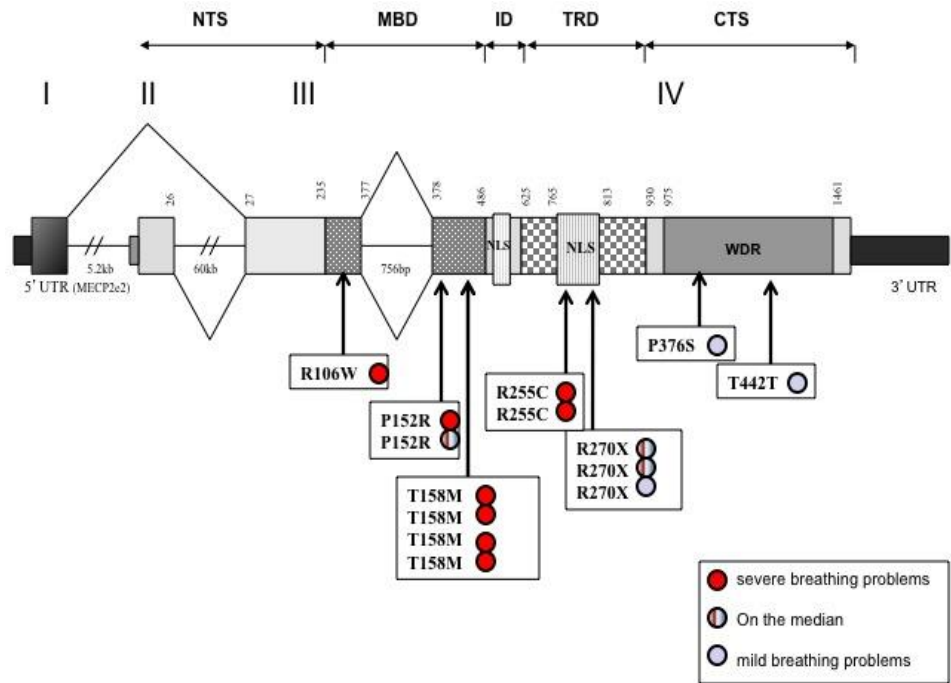


Figure 3. Genomic structure of the MECP2 gene and localization of breathing problems in the coding regions.

Based on high ($>$ median) or low ($\leq$ median) scores on the number of consonants spontaneously produced (Mdn = 3), low level (Mdn = 1.2), and high level (Mdn = 5,8), participants were placed within one of the two type categories. As can be seen in figure 5, patients with a truncating mutation after NLS show similar pattern to patients with a truncating mutation within NLS ($p<.34$).

With reference to the number of vowels with elicited denomination, based on high ($>$ median) or low ($\leq$ median) scores on respiratory rhythm (Mdn = 1), low level of vowels with elicited Denomination (Mdn = 0.4), and high level of vowels with elicited denomination (Mdn = 2,7), participants were placed within one of the two type categories. As shown in figure 6, patients with a truncating mutation after NLS manifested a higher level of vowels denominations than patients with a truncating mutation within NLS (χ^2 (2, N = 14) = 2.15, $p<.05$). With reference to the number of consonants with elicited denomination (namely the

syllables), only two girls (P376S and T442T) were able to repeat a high number of syllables (respectively 12 and 10). For this reason the relative figure was not produced.

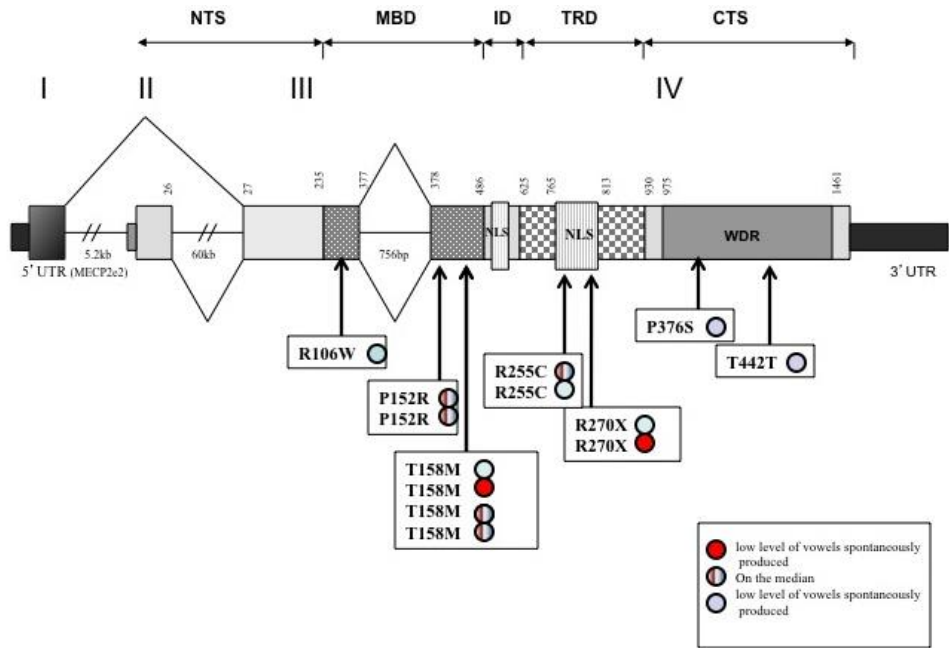


Figure 4. Genomic structure of the MECP2 gene and localization of vowels spontaneously produced in the coding regions.

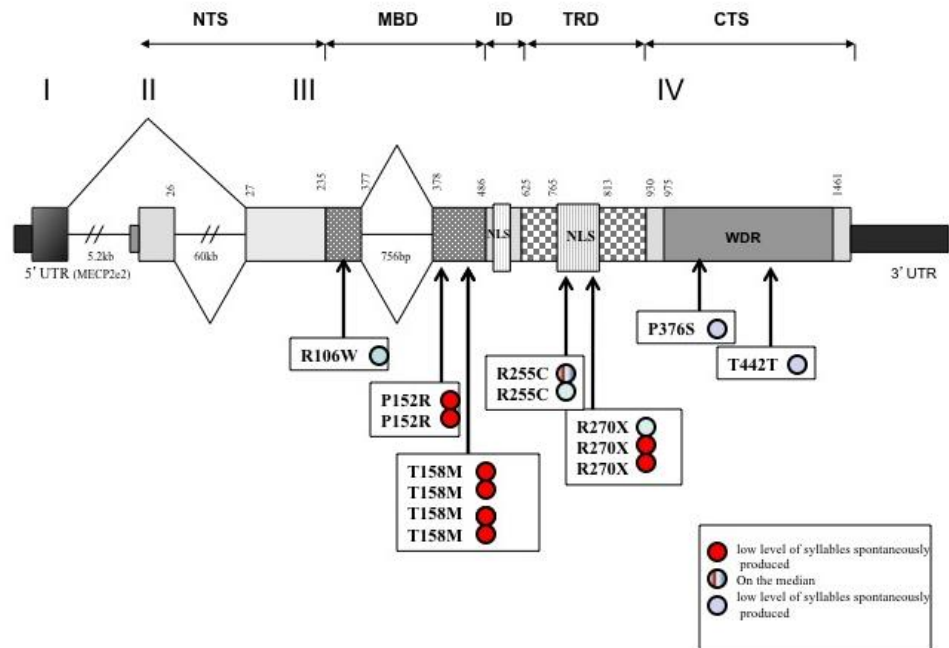


Figure 5. Genomic structure of the MECP2 gene and localization of syllables spontaneously produced in the coding regions.

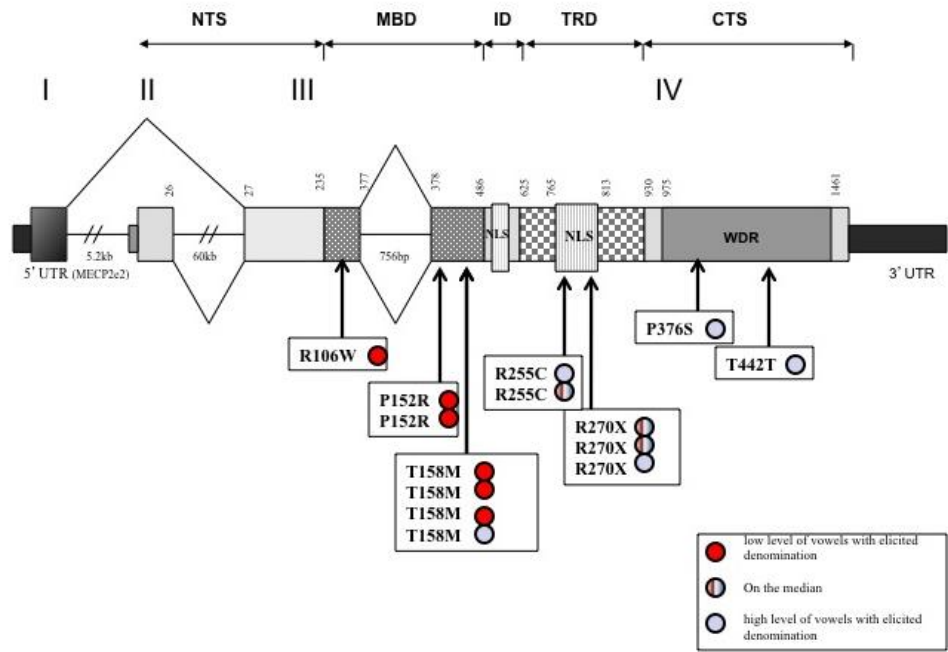


Figure 6. Genomic structure of the MECP2 gene and localization of vowels with elicited denomination.

CONCLUSION

In this study, disease-causing mutations have been related to the intensity of the parameters of the breathing problems and the Fanzago test, namely the number of vowels spontaneously produced, the number of syllables spontaneously produced, the number of vowels with elicited denomination and the number of consonants with elicited denomination. Comparisons between the truncating mutations differently affecting functional domains induce support for the idea that the crucial factor that leads to different phenotypes is the integrity of NLS.

As in the study of Fabio et al. (2014) this study supposes that if the protein can penetrate into the nucleus and link to Methylated CpG, it maintains a residual role causing milder clinical damage.

The differential phenotypic effects produced by the two kinds of mutations were clearly shown by the difference in the breathing scores and in the level of the vowels denomination with elicitation.

Our work analyzed samples with a limited number of patients, for this reason it is just a pilot study and more data has to be collected. The most important innovation introduced in the study was the use of the correlation between breathing and language in relation to the specific genotype.

REFERENCES

- Adler DA, Quaderi NA, Brown SD, Chapman VM, Moore J, Tate P, Distèche CM. (1995) The X-linked methylated DNA binding protein, Mecp2, is subject to X inactivation in the mouse. *Mamm Genome*; 6:491-2.
- American Psychiatric Association (2000) *Diagnostic and Statistical Manual of Mental Disorders*, American Psychiatric Press, Washington, DC, USA, 4th edition.
- American Psychiatric Association (2013) *Diagnostic and statistical manual of mental disorders, 5th ed.* American Psychiatric Publishing, Arlington.
- Antonietti, A., Castelli, I., Fabio, R. A., & Marchetti, A. (2003). *La sindrome di Rett. Prospettive e strumenti per l'intervento*. [Rett's syndrome. Perspectives and tools for intervention]. Roma: Carocci.
- Balmer D, Goldstine J, Rao YM and LaSalle JM. (2003) Elevated methyl-CpG binding protein 2 expression is acquired during postnatal human brain development and is correlated with alternative polyadenylation, *J. Mol. Med.* 81: 61–68.
- Budden, S. (1997) Rett syndrome: habilitation and management reviewed. “European Child Adolescent Psychiatry”, 6, Supplement 1., 103-107.
- Castelli, I., Antonietti, A., Fabio, R.A., Lucchini, B., Marchetti, A. (2013). Rett syndrome girls possess Theory of Mind. *Life Span and Disability XVI*, 2, 157-168.
- Cohen DR, Matarazzo V, Palmer AM, Tu Y, Jeon OH, Pevsner J and Ronnett GV. (2003) Expression of MeCP2 in olfactory receptor neurons is developmentally regulated and occurs before synaptogenesis, *Mol. Cell. Neurosci.* 22: 417–429.
- Coy JF, Sedlacek Z, Bachner D, Delius H, Poutska A. (1999) A complex pattern of evolutionary conservation and alternative polyadenylation within the long 3'-untranslated region of the methyl-CpG-binding protein 2 gene (MECP2) suggest a regulatory role in gene expression. *Hum Mutat*; 15:7-12.
- D'Esposito M, Quaderi NA, Ciccodicola A, Bruni P, Espodito T, D'Urso M, Brown SD. (1996) Isolation, physical mapping, and northern analysis of the Xlinked human gene encoding methyl CpG-binding protein MECP2. *Mamm Genome*; 7:533-35.
- De Bona, C.; Zappella, M.; Hayek G.; Meloni I.; Vitelli F.; Bruttini M.; Cusano R.; Loffredo P.; Longo I.; Ranieri A. (2000) Preserved speech variant is allelic of classic Rett syndrome, “*European Journal of Human Genetics*”, 8, 325–330.
- Deguchi, K., Antalffy, B.A., Twohill, L.J., Chakraborty, S., Glaze, D.G., Armstrong, D.D., 2000. Substance P immunoreactivity in Rett syndrome. *Pediatr. Neurol.* 22, 259–266.
- Elston, The genetic dissection of multifactorial traits. *Clin Exp Allergy*, 1995; 25:103–6.
- Fabio R.A., Antonietti A., Castelli I., Marchetti A. (2009) Attention and communication in Rett Syndrome, *Research in Autism Spectrum Disorders* 3, 329–335.
- Fabio, R.A., Castelli, I., Antonietti, A., Marchetti, A. (2013). Training communication abilities in Rett Syndrome through reading and writing. *Frontiers in Psychology for Clinical Settings*, 4, 911, 1-9. doi: 10.3389/fpsyg.2013.00911.
- Fabio R.A., Colombo B., Russo S., Cogliati F., Masciadri M., Foglia S. Antonietti A., Tavian D. (2014) Recent insights into genotype–phenotype relationships in patients with Rett syndrome using a fine grain scale, *Research in Developmental Disabilities* 35, 2976–2986.

- Fabio, R.A., Giannatiempo, S., Antonietti, A., Budden, S., (2009). The role of stereotypies in overselectivity processes in Rett Syndrome. *Research in Developmental Disabilities*, 30, 136–145.
- Fabio, R.A., Giannatiempo, S., Oliva, P., Murdaca, A.M. (2011). The increase of attention in Rett Syndrome. A pretest-post test research design. *Journal of Developmental and Physical Disability*, 23, 99-111.
- Fabio, R.A. Martinazzoli, C., Antonietti, A. (2005). Costruzione e standardizzazione dello strumento “R.A.R.S.” (Rett Assessment Rating Scale). *Ciclo Evolutivo e Disabilità*, 8, 257-281.
- Fabio, R.A., Oliva, P., Murdaca, A.M. (2011). Systematic and emotional contents in overselectivity processes in autism. *Research in Autism Spectrum Disorder*, 5, 575 – 583.
- Falzone A. (2014) Structural constraints on language, in "Reti, saperi, linguaggi. *Italian Journal of Cognitive Science*" 2/2014, pp. 247-266, doi: 10.12832/78928.
- Falzone A., Anastasi A., Pennisi A., (2014) Biological Constraints and Evolution of Language: a Hypothesis on Exaptation of Human Vocal Structures, in *Proceeding of European Society for the Study of Human Evolution*, pp. 66-67.
- Hanefeld F. The clinical pattern of the Rett syndrome. *Brain and Development* 1985; 7:320-325.
- Jones PL, Veenstra GJ, Wade PA, Vermaak D, Kass SU, Landsberger N, Strouboulis J, Wolffe AP. (1998) Methylated DNA and MeCP2 recruit histone deacetylase to repress transcription. *Nature Genet*; 19:187-191.
- Jung, B.P., Jugloff, D.G., Zhang, G., Logan, R., Brown, S. and Eubanks., J.H. (2003) The expression of methyl CpG binding factor MeCP2 correlates with cellular differentiation in the developing rat brain and in cultured cells, *J. Neurobiol.* 55: 86–96.
- Katz D.M., Dutschmann M., Ramirez J.M., Hilaire J. (2009) Breathing disorders in Rett syndrome: Progressive neurochemical dysfunction in the respiratory network after birth, *Respiratory Physiology & Neurobiology* 168, 101–108.
- Kishi and Macklis, 2004 N. Kishi and J.D. Macklis. (2004) MECP2 is progressively expressed in post-migratory neurons and is involved in neuronal maturation rather than cell fate decisions, *Mol. Cell. Neurosci.* 27: 306–321.
- Knudsen, G.P., Neilson, T.C., Pedersen, J., Kerr, A., Schwartz, M., Hulten, M., et al. (2006). Increased skewing of X chromosome inactivation in Rett syndrome patients and their mothers. *European Journal of Human Genetics*, 14, 1189–1194.
- Kriaucionis K, and Bird A. (2004) The major form of MeCP2 has a novel Nterminus generated by alternative splicing, *Nucleic Acids Res.* 32: 1818–1823.
- Lee SS, Wan M, Francke U. (2001) Spectrum of MECP2 mutations in Rett syndrome. *Brain Dev*; 23 Suppl 1:S138-S143.
- Lewis JD; Meehan RR, Henzel WJ, Maurer-Fogy I, Jeppesen P, Klein F, Bird A. (1992) Purification, sequence, and cellular localization of a novel chromosomal protein that binds to methylated DNA. *Cell*; 69:905-914.
- Marschik P.B., Sigafos J., Kaufmann W.E., Wolin T., Talisa V.B., Bartl-Pokorny K.D., Budimirovic b, Ralf Vollmann B.D., Einspieler C. (2012b) Peculiarities in the gestural repertoire: An early marker for Rett syndrome? “*Research in Developmental Disabilities*” 33, 1715–1721.
- Meehan RR, Lewis JD, Bird AP. Characterization of MeCP2, a vertebrate DNA binding protein with affinity for methylated DNA. *Nucleic Acids Res* 1992; 20(19):5085-92.

- Mnatzakanian GN, Lohi H, Munteanu I, Alfred SE, T. Yamada T, MacLeod PJ, J.R. Jones JR, Scherer SW, Schanen NC and Friez MJ et al., (2004) A previously unidentified MECP2 open reading frame defines a new protein isoform relevant to Rett syndrome, *Nat. Genet.* 36: 339–341.
- Mullaney BC, Johnston MV and Blue ME. (2004) Developmental expression of methyl-CpG binding protein 2 is dynamically regulated in the rodent brain, *Neuroscience* 123: 939–949.
- Nan X, Campoj FJ, Bird A. (1997) MeCP2 is a transcriptional repressor with abundant binding sites in genomic chromatin. *Cell*; 88:471–481.
- Neul JL, Kaufmann WE, Glaze DG, et al. Rett syndrome: revised diagnostic criteria and nomenclature. *Ann Neurol* 2010; 68: 944–50.
- Ogier M.; Katz D.M. (2008), Breathing dysfunction in Rett syndrome: Understanding epigenetic regulation of the respiratory network, “*Respiratory Physiology & Neurobiology*” 164, 55–63.
- Ramirez J.M., Ward C.S., Neul J.L. (2013) Breathing challenges in Rett Syndrome: Lessons learned from humans and animal models, “*Respiratory Physiology & Neurobiology*” 189, 280–287.
- Ranieri A.; Mari F., Mencarelli M.A.; Scala E., Ariani F., Longo I., Meloni I., Cevenini G., Pini G., Hayek G., Zappella M. (2009) Diagnostic criteria for the Zappella variant of Rett syndrome (the preserved speech variant), “*Brain & Development*” 31, 208–216.
- Saito, Y., Ito, M., Ozawa, Y., Matsuishi, T., Hamano, K., Takashima, S., 2001. Reduced expression of neuropeptides can be related to respiratory disturbances in Rett syndrome. *Brain Dev.* 23 (Suppl. 1), S122–S126.
- Scala E, Ariani F, Mari F, Caselli R, Pescucci C, Longo I, Meloni I, Giachino D, Bruttini M, Hayek G, Zappella M, Renieri A. (2005) CDKL5/STK9 is mutated in Rett syndrome variant with infantile spasms. *J Med Genet.* Feb; 42(2):103–107.
- Shahbazian MD, Antalffy B, Armstrong DL and Zoghbi HY. (2002) Insight into Rett syndrome: MeCP2 levels display tissue- and cell-specific differences and correlate with neuronal maturation, *Hum. Mol. Genet.* 11: 115–124.
- Sparrow, S. S., Balla, D. A., Cicchetti, D. V. (1984) *Vineland Adaptive Behavior Scales*, Circle Pines, MN: American Guidance Service.
- Tao J, Van Esch H, Hagedorn-Greiwe M, Hoffmann K, Moser B, Raynaud M, Sperner J, Fryns JP, Schwinger E, Gecz J, Ropers HH, Kalscheuer VM. (2004) Mutations in the X-linked cyclin-dependent kinase-like 5 (CDKL5/STK9) gene are associated with severe neurodevelopmental retardation. *Am J Hum Genet.* Dec; 75(6):1149–54.
- Trevarthen C., Daniel S., (2005) Disorganized rhythm and synchrony: Early signs of autism and Rett syndrome, *Brain & Development* 27, S25–S34.
- Uchino J., Suzuki M., Hoshino K., Nomura Y., Segawa M. (2001) Development of language in Rett syndrome, *Brain & Development* 23, S233–S235.
- Vignoli A, Fabio RA, La Briola F, Giannatiempo S, Antonietti A, Maggiolini S, Canevini MP. Correlations between neurophysiological, behavioral, and cognitive function in Rett syndrome. *Epilepsy Behav.* 2010 Apr; 17(4):489–96. doi: 10.1016/j.yebeh.2010.01.024.
- Vilain A, Apiou F, Vogt N, Dutrillaux B, Malfroy B. (1996) Assignment of the gene for methyl-CpG-binding protein 2 (MECP2) to human chromosome band Xq28 by in situ hybridization. *Cytogenet Cell Genet*; 74:293–94.
- Weaving LS, Christodoulou J, Williamson SL, Friend KL, McKenzie OL, Archer H, Evans J, Clarke A, Pelka GJ, Tam PP, Watson C, Lahooti H, Ellaway CJ, Bennetts B, Leonard H,

- Gecz J. (2004) Mutations of CDKL5 cause a severe neurodevelopmental disorder with infantile spasms and mental retardation. *Am J Hum Genet.* 2004 Dec; 75(6):1079-93. Epub Oct 18.
- Zappella M.; Meloni I.; Longo I.; Hayek G.; Ranieri A. (2001) Preserved Speech Variants of the Rett Syndrome: Molecular and Clinical Analysis, "*American Journal of Medical Genetics*", 104, 14-22.

Chapter 49

MARFAN SYNDROME: A RARE AND SOMETIME VERY DISABLING GENETIC DISORDER

Henry J. Mankin and Keith P. Mankin*

Massachusetts General Hospital, Harvard Medical School,
Orthopaedic Department, Boston, MA, US

ABSTRACT

Marfan Syndrome was originally described by Antoine Bernard-Jean Marfan in 1896 and is an uncommon inherited connective tissue abnormality which occurs as an autosomal dominant genetic disorder with frequent mutations.

The patients have normal mentation but have a characteristic increase in height, abnormally long limbs, arachnodactyly, joint hypermobility, distinctive facial features, scoliosis, ectopia lentis, dural ectasia and array of aortic and cardiac abnormalities that are often life threatening. The genetic cause of the disease has been identified. Although some medical and surgical treatments are currently in practice they are not always helpful.

The orthopaedic problems are often significant and sometimes require complex surgical interventions.

INTRODUCTION

Marfan Syndrome originally described by Antoine Bernard-Jean Marfan in 1896 is an uncommon inherited connective tissue abnormality which occurs as an autosomal dominant genetic disorder with frequent mutations. The patients have normal mentation but have as characteristic features excessive height, abnormally long limbs (dolichostenomelia), arachnodactyly (spider-like digits on hands and feet), joint hypermobility, distinctive facial abnormalities, scoliosis, dislocating lenses (ectopia lentis), enlargement of the dural sac (dural

* Dr. Keith Mankin was formerly an active Chair-person in the MGH Pediatric Orthopaedic Division.

ectasia) and an array of aortic and cardiac abnormalities that are frequently life threatening. The nature of the abnormality has been identified and the genetic cause has been discovered. Some medical treatments are currently available to somewhat protect against cardiac disasters and surgical approaches can reduce the risk considerably. The orthopaedic problems can be significant and sometimes require surgical intervention, particularly for spinal or hip abnormalities or fractures.

HISTORY OF MARFAN SYNDROME

Antoine Bernard-Jean Marfan (1858-1942) was born in France and after completing his medical training in Toulouse and some special training in children's diseases, became Chief of Pediatrics at the University of Paris and the Hopital des Enfants Maladies in 1914, a position he held until his retirement in 1928 [19].

He became very involved in defining infectious and other disorders in children and described features of syphilis (Dennie-Marfan syndrome), a red triangle on the tongue characteristic of typhus (Marfan's sign), rachitic epiphyseal swelling of the medial malleous (Marfan's symptom), and a prognostic rule related to tuberculosis of the throat (Marfan's Law) [19]. In 1896, Marfan presented a case of a 5 year old girl, Gabrielle P. to the Societe Medicale des Hopitaux de Paris [37].

He pointed out her disproportionately long limbs, asthenic physique, and slender and exceptionally long fingers and toes. She was studied again with X-rays by Henri Mery and Leon Babonneix in 1902 [41] and they noted scoliosis and thoracic asymmetry; and somewhat later she was discovered to have cardiovascular abnormalities and dislocations of her lenses. Subsequent reports on other patients by Achard in 1902 [1] defined the familial nature of the disease and Salle in 1912 reported necropsy findings of changes in the heart and aorta [62]. In 1914, Boerger first clearly defined ectopia lentis in the patients [10].

It was Weve who in 1931 [77] confirmed the hereditary characteristic of the disease and provided the name "dystrophia mesodermalis congenital, type Marfanis" and since then, the disorder has been known as Marfan syndrome. The cardiac and aortic abnormalities in the patients were further described by Baer, Taussig and Oppenheim in 1943 [5] and shortly thereafter by Etter and Gover (21). Victor A. McKusick became fascinated with the disease and wrote about it extensively beginning in 1955 [28, 39, 40].

Gordon [29] and Schwartz [64] separately in 1962 and 1964 attempted to establish the likelihood that Abraham Lincoln had had Marfan syndrome but despite attempts to seek the gene error in his remains, the issue still remains unclear [28].

GENETIC CAUSATION AND FREQUENCY OF MARFAN SYNDROME

Marfan syndrome is an inherited connective tissue disorder transmitted as an autosomal dominant trait with frequent mutations [7,11,16,25,28,32,40,51]. The disorder is equally distributed in ethnic groups, is slightly more frequent in females than males, and occurs worldwide at 1 per 5-10,000 births [7, 28, 32, 40, 51].

The disorder is now known to be caused by mutations in the fibrillin-1 (FBN1) gene located on chromosome 15q21.1 [11, 15, 16, 28, 56]. The gene encodes the glycoprotein fibrillin which is a major building block for microfibrils including those involved in the skeletal system, the spinal dura, the optic lens supporting system and the elastin in the aorta and cardiac valvular structures [15, 16, 25, 26, 28, 47].

It has also been proposed that there is a disturbance of tissue homeostasis of elastic fibers, increased susceptibility of fibrillin to proteolysis and dysregulation of TGF- β resulting in increased apoptosis [16, 25, 43, 46]. The gene error can be identified not only in patients with the disorder or their family but in uterine fluids.

It is also possible to identify the changes seen in the child in the uterus by appropriate imaging studies [13, 28, 32].

CLINICAL FINDINGS IN PATIENTS WITH MARFAN SYNDROME

Patients with Marfan syndrome may have multiple finding, some frequent and some less commonly. In 1986, a Committee established the Berlin Criteria, which includes diagnostic features which are commonly present and are required to make the diagnosis (major factors); and others which are less commonly present and may be associated with other disorders [7, 28, 58]. Of some importance in any discussion of Marfan syndrome is a listing of some of the other genetic disorders that may resemble aspects of Marfan syndrome and cause confusion in both diagnosis and treatment protocols. The similar lesions that have some of the characteristics of Marfan syndrome include homocystinuria [28], Ehler's Danlos syndrome [28, 40], Klinefelter syndrome [33, 54, 67], Lujan Fryns syndrome [50], van den Ende-Gupta disease [30, 65], Joubert syndrome [27, 44], Shprintzen-Goldberg syndrome [31, 57, 61], Beals disease [72], Marfanoid hypermobility syndrome [76], Stickler syndrome [8], PHACE syndrome [66] and some other disorders that affect the hands, feet, heart, aorta and bones [6, 21, 27, 28, 40]. The list of findings included in this chapter are considered important only for the diagnosis of Marfan syndrome and hence are known as "major factors" according to the Berlin Criteria [7].

General and Orthopaedic Characteristics of Marfan Syndrome

Most patients with Marfan syndrome have normal mentation [20, 28, 40, 49, 51]. Children born with Marfan syndrome are often quite tall and in a short time are taller than their normal siblings and peers [7, 20, 28, 40, 48] (Figure 1). Some of the patients as adults may grow to over seven-feet in height [20, 32, 48, 49]. An important finding is dolichostenomelia, which is defined as having excessively long limbs, far in excess of the normal relationship of the upper and lower extremities to the rest of the body [28, 40] (Figure 2). The bones are generally osteopenic with frequent fractures [7, 12, 18, 24, 35] and with an unusually frequent occurrence of protrusio acetabulae which can be very disabling [18, 68, 70, 73, 79]. Scoliosis is common (Figure 2). The limbs are excessively thin suggesting that not only are the bones diminished in circumference but that there is less soft tissue surrounding them. The fat content of the soft tissues of the extremities is reduced and there is often muscular underdevelopment so that the extremities are much thinner than those of unaffected individuals [7, 28, 40] (Figure 1).

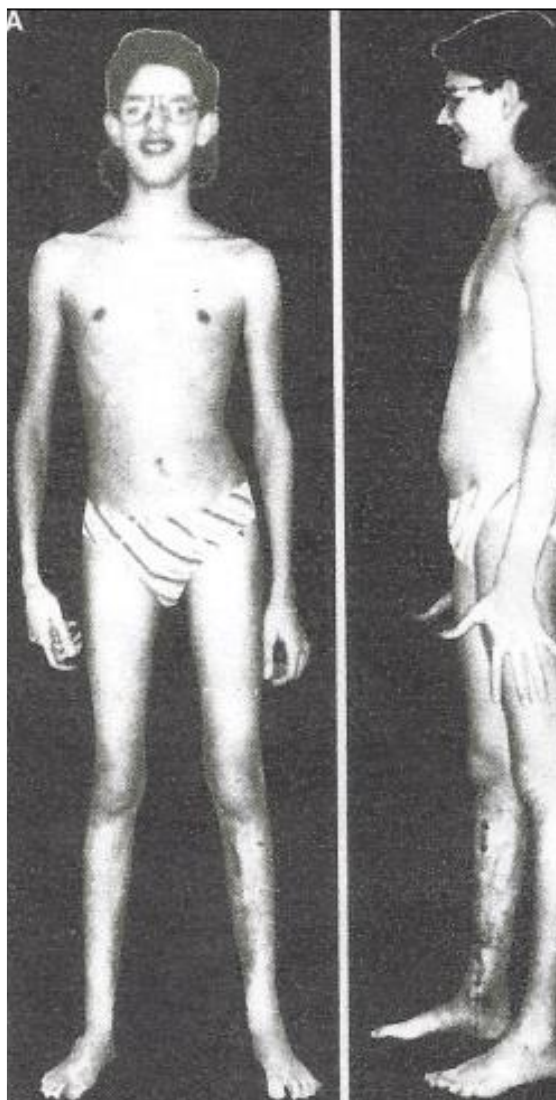


Figure 1. Patients with Marfan's syndrome are frequently excessively tall and sometimes very thin.

The hands and feet show very thin and much elongated digits which are described as “arachnodactyly”, suggesting that they resemble the limbs of spiders [7, 21, 25, 32, 51] (Figure 2).

Flat feet are common as are dislocations of joints such as the elbows, wrists, knees and especially hips [25, 28, 40] (Figure 2). The great toe is sometimes excessively elongated (Figure 3). The hand digits are excessively mobile and the flexed thumb may extend beyond the four fingers when the hand and digits are flexed (Steinberg Thumb Sign) [28] (Figure 4). In addition the combination of a thin wrist and long digits may allow an overlap of the thumb and first and fifth fingers when the wrist is grasped (Walker-Murdoch Sign) [28] (Figure 5).



Figure 2. Skeletal abnormalities in patients with Marfan's syndrome can be quite excessive and very disturbing to function. The calvarium can be thinner than normal and fingers elongated and distorted in shape. Scoliosis is common. Joint structures are disturbed and the patients may develop arthritic changes.

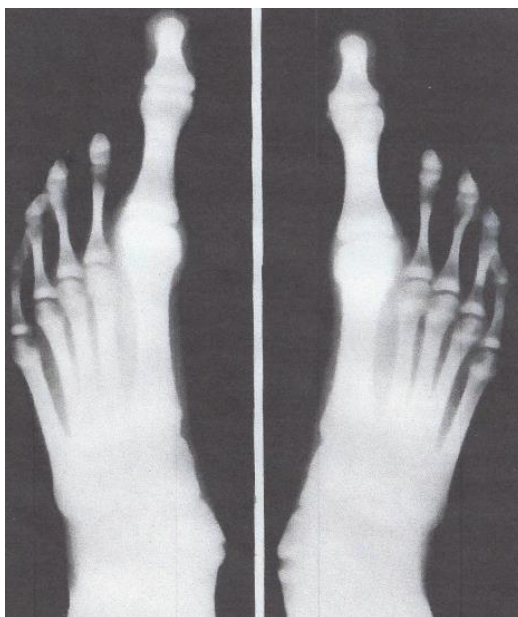


Figure 3. Elongation of the great toe sometimes occurs and may be very disabling.

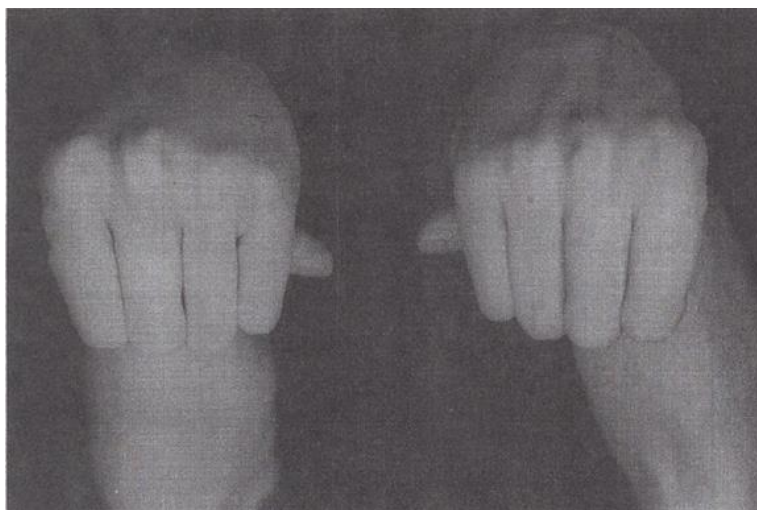


Figure 4. The Steinberg Thumb sign a combination of narrow hand, long digits and loose joints permits the thumb to extend well beyond the ulnar surface.

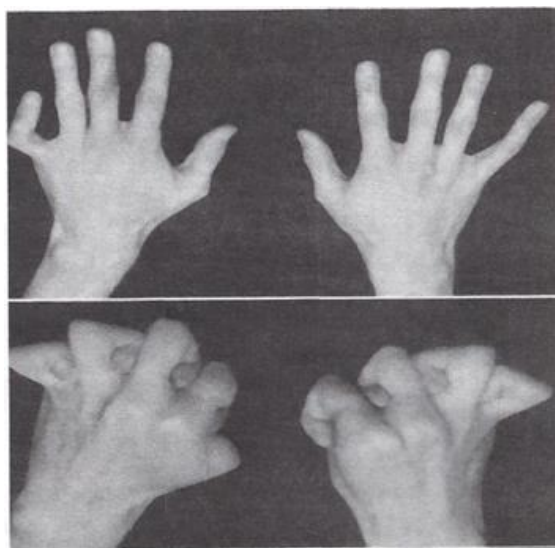


Figure 5. Walker Murdoch sign the digits are so flexible the tips of the fingers may appear in the spaces between the proximal structures in the same joint.

The ribs show abnormal growth and many of the patients show either a pectus excavatum or a pectus carinatum [4, 28, 40]. The limbs may show gross structural distortion resulting in functional loss (Figure 6).

Scoliosis or kyphoscoliosis occurs in 30-60% of the patients, is commonly present in even young children and progresses with advancing years [9, 25, 28, 32, 40, 89] (Figure 2). The lesions along with the pectus changes and facial deformities are such as to compromise pulmonary function in some of the patients and may be associated with sometimes severe sleep apnea [4, 14, 23, 40].



Figure 6. Bowing of the distal extremities may make it difficult for the patient to walk.

Additional findings include spina bifida occulta, hemivertebra, cleft or high-arched palate and dental crowding [9, 40] and a thin narrow face and prognathism is characteristic [9, 25, 28, 32] (Figure 1). One of the more common entities is dural ectasia in which the dura becomes moderately or markedly expanded causing widening of the spinal canal, thinning of the vertebral cortex and pedicles, dilatation of the neural foramina and protrusion of the dura outside the bony canal [2, 3, 7, 9, 22, 52, 71, 73, 75]. These findings may cause loss of nerve function and even paraplegia. Many of the patients are muscularly impaired so cannot perform exercises or participate in sports [12, 20, 40, 49, 74].

Ocular findings: Almost all patients with Marfan syndrome have ocular findings and often complaints related to visual alterations and loss. The most frequent problem is “ectopia lentis”, partial or complete dislocation of the lens, which occurs in 50 to 80% of the patients [6, 28, 38, 40, 51]. Tremor of the iris may occur and suggests the likelihood of dislocating lenses. Flattening of the cornea is common and blue sclerae have been reported in patients with the disease and visual loss is common [28, 38, 40]. Many of the patients have myopia and loss of visual acuity which requires glasses. Occasional cases of retinal detachment are encountered and blindness is not unusual in these patients.

Despite normal intelligence some of the patients have considerable difficulty in a school setting because of inability to read either small print or a distant blackboard [38, 49, 74].

Cardiovascular disorders: Seventy percent of the patients with Marfan syndrome have aortic root dilatation and aortic regurgitation [5, 27, 28, 32, 39, 40, 51, 53].

The process may be manifest at an early age and is more common in men than in women [6, 27, 28, 39, 40]. A diastolic murmur over the aortic valve may be present [28].

Aortic dissection involving the ascending aorta is a serious problem and can lead to the death of the patient. Mitral valve prolapse may occur in 55-70% of the patients and results in a high-pitched late-systolic murmur, shortness of breath and a rapid pulse rate [6, 28, 39, 40]. Pulmonary artery dilatation may also occur and even a dissecting aneurysm of the pulmonary artery which can be fatal [6, 28, 51, 53].

Most of these can be diagnosed by electrocardiography, echocardiography and MRI studies of the heart and the aorta and these studies should be done regularly.

In addition some patients have been found to have other cardiovascular manifestations, including coarctation of the aorta, patent ductus arteriosus, absent pulmonary valve and interatrial stenosis [28, 40, 51, 53]. Patients may develop spontaneous pneumothorax or apical blebs (23,28). Cardiac and aortic disorders are serious and can lead to the patient's death, often unexpectedly and at an early age or even in partum [13, 27, 28, 39, 45, 53, 63].

Other findings: These include unusual skin changes such as striae atrophicae (transverse striations often located in the back). Some of the patients may have spontaneous or incisional herniae. [28, 40].

DIAGNOSTIC STUDIES FOR THE STUDY OF PATIENTS SUSPECTED OF HAVING MARFAN SYNDROME

As indicated above, family history is an essential part of the evaluation of the patient. Studies for the molecular gene error seeking the FBN1 mutation are important to provide a competent system of diagnosis of the disease. Similar studies seeking the genetic error in the patient's parents and siblings may also be helpful [11, 15, 28, 40, 47].

All children or adults thought to have Marfan syndrome should be evaluated by a team of clinicians which includes individuals interested in genetic diseases, orthopaedic disorders, ocular problems and cardiac diseases. Both children and adults require imaging studies which includes the spine, pelvis, limbs, hands, feet, chest, skull, heart and lungs. Some of these should be done regularly. CT and MRI imaging are particularly important in the study of the spine for dural ectasia [2, 3, 22, 71, 75]. Bone densitometry should be performed at least yearly for all patients [24, 35]. Frequent evaluation of the heart sounds by a competent cardiologist and echocardiography, electrocardiography and MRI may be useful and life-saving [28, 32, 39, 40, 47, 51]. Ocular studies should include slit-lamp evaluation and keratometry [28, 38].

TREATMENT OF PATIENTS WITH MARFAN SYNDROME

Some of the patients with Marfan syndrome may require no treatment other than frequent observation for ocular, cardiac or orthopaedic complaints or findings. The majority however do require some attempts to alter the progress of the disease and not only keep the patients alive but try to improve their life status.

Medical treatment: Recently, beta-blockers have been introduced to reduce the cardiac and aortic stress and hopefully reduce some of the commonly developing problems [28, 32, 55, 59, 78]. The drugs include atenolol, propranolol hydrochloride and verapamil hydrochloride. Estrogen and androgen and somatostatin have been administered to children to slow their skeletal progression but at this point has only limited success and may produce additional problems [28, 48, 60].

Cardiac treatment: Supportive cardiac measures may be helpful but for some patients [42, 47]. For patients whose lives are threatened, surgery for aortic or mitral valve repair or replacement, aortic arch reconstruction or composite aortic graft insertion or replacement may

be required [28, 42, 47, 53]. A recent report suggests that cardiac transplantation is necessary for some patients [34].

Ocular treatment: Myopia is treatable with refraction and lens defects are best treated by surgical procedures [22, 38].

Orthopaedic treatment: Bracing of spinal curvature may be helpful and prevent progression of scoliotic and kyphotic changes. It may be necessary to do corrective surgery especially if the patient develops dural ectasia or progressive scoliotic deformity [17, 36, 42, 69, 71]. Pectus excavatum if severe may require surgery to prevent damage to the pulmonary tree and the aorta [28, 40]. Surgery to correct protrusion acetabulae is sometimes necessary but may not be effective [70, 73, 79]. Surgery for hands, feet, elbows, hips or knees are occasionally necessary for subluxations or fractures. The use of bisphosphonates in an attempt to decrease the osteopenia of bony segments and reduce the likelihood of fractures may be tried but as yet has not been reported to be successful [24, 28].

Psychological treatment: It should be apparent that the patients with Marfan syndrome and their families have a great burden in their lives [49, 74]. The children are tall and unusually structured and considered strange in appearance by normal persons. The development of eye problems sometimes limits their ability to be educated and their physical problems markedly limit their ability to participate in social and athletic activities. The patients and their parents may need psychiatric support and they should also consult with geneticists about issues related to having additional children.

DISCUSSION

Professor Marfan first described his patient Gabrielle P. in 1896 and we have over the years since then, gathered some additional information about the clinical characteristics of the disease and the sometimes life and limb threatening complications. We now know the genetic origin of the process and how it causes the disorder and this is very helpful in diagnosis and familial evaluation. We have cataloged the frequency of the various symptoms and signs and established their value in providing criteria for diagnosis. We have also begun a series of medical and surgical protocols for management of these patients but at least thus far, they really cannot reverse the process they can only treat the complications. It is hoped that over the next decade that we can define how best to treat the genetic error, possibly in utero after the diagnosis is made and thus reduce the effect of this sometimes devastating disorder.

REFERENCES

- [1] Achard C: *Arachnodactylie: Bull Mem Soc Med Hop* 19:834-843, 1902.
- [2] Ahn NU, Sponseller PD, Ahn UM et al: Dural ectasia in the Marfan syndrome: MR and CT findings and criteria. *Genet. Med.* 2: 173-179, 2000.
- [3] Ahn NU, Sponseller PD, Ahn UM et al: Dural ectasia is associated with back pain in Marfan syndrome. *Spine* 25:1562-1568, 2000.
- [4] Arn PH, Scherer LR, Haller JA Jr., Pyeritz RE: Outcome of pectus excavatum in patients with Marfan syndrome and in the general population. *J. Pediatr.* 115: 954-958, 1989.

- [5] Baer RW, Taussig HB, Oppenheimer EH: Congenital aneurysmal dilation of the aorta associated with arachnodactyly. *Bull Hopkins Hosp.* 72: 309-331, 1943.
- [6] Bawle E, Quigg MH: Ectopia lentis and aortic root dilatation in congenital contractural arachnodactyly. *Am. J. Med. Genet.* 42: 19-21, 1992.
- [7] Beighton p, dePaepe A, Danks D et al: International Nosology of Heritable Disorders of Connective Tissue Berlin 1986, *Am. J. Med. Genet.* 29: 581-594, 1988.
- [8] Bennett JT, McMurray SW: Stickler syndrome. *J. Pediatr. Orthop.* 10: 760-763, 1990.
- [9] Birch JG, Herring JA: Spinal deformity in Marfan syndrome. *J. Pediatr. Orthop.* 7: 546-552, 1987.
- [10] Boerger D: Ueber zwei Falle von arachnodaktyllie. [About two cases of arachnodactyly]. *Monatsschr Kinderheilkd* 12: 161-184, 1914.
- [11] Boileau C, Jondeau G, Mizuguchi T, Matsumoto N: Molecular genetics of Marfan syndrome. *Curr. Opin. Cardiol.* 20:194-200, 2005.
- [12] Braverman AC: Exercise and the Marfan syndrome. *Med. Sci. Sports Exerc.* 30:(Suppl): 387-395, 1998.
- [13] Buchanan R, Wyatt GP: Marfan's syndrome presenting as an intrapartum death. *Arch. Dis. Child* 60: 1074-1076, 1985.
- [14] Cistulli PA, Gotsopoulos H, Sullivan CE: Relationship between craniofacial abnormalities and sleep disordered breathing in Marfan's syndrome. *Chest.* 1455-1460, 2001.
- [15] Dietz HC, Cutting GR, Pyeritz RE et al: Marfan syndrome caused by a recurrent de novo missense mutation in the fibrillin gene. *Nature.* 353: 337-339, 1991.
- [16] Dietz HC, Pyeritz RE: Mutations in the human gene for fibrillin-1 (FBN1) in the Marfan syndrome and related disorders. *Hum. Genet.* 4: 1799-1809, 1995.
- [17] Di Sivestre M, Gregg T, Giacomini S et al: Surgical treatment for scoliosis in Marfan syndrome. *Spine* 30: 597-604, 2005.
- [18] Do T, Giampietro PF, Burke SW et al: The incidence of protrusio acetabuli in Marfan's syndrome and its relationship to bone mineral density. *J. Pediatr. Orthop.* 20: 718-721, 2000.
- [19] Enersen OD. *Antoine Bernard-Jean Marfan biography*: In <http://whonameedit.com/doctor.cfm/972.html>. 2001.
- [20] Erkula G, Jones KB, Sponseller PD et al: Growth and maturation in Marfan syndrome. *Am J Med Genet* 109: 100-115, 2002.
- [21] Etter LE, Glover LP: Arachnodactyly complicated by dislocated lens and death from rupture of dissecting aneurysm of the aorta. *JAMA* 123: 88-89, 1943.
- [22] Foran JR, Pyeritz RE, Dietz HC, Sponseller PD: Characterization of the symptoms associated with dural ectasia in the Marfan patient *Am. J. Med. Genet.* 134: 58-65, 2005.
- [23] Fuleihan FJ, Suh SK, Shepherd RH: Some aspects of pulmonary function in the Marfan syndrome. *Bull Hopkins Hosp.* 113: 320-329, 1963.
- [24] Giampietro PF, Peterson M, Schneider R et al: Assessment of bone mineral density in adults and children with Marfan syndrome. *Osteoporos Int.* 14: 559-563, 2003.
- [25] Giampietro PF, Raggio C, Davis JG: Marfan syndrome: orthopedic and genetic reviews. *Curr. Opin. Pediatr.* 14: 35-41, 2002.
- [26] Gigante A, Chillemi C, Greco F: Changes of elastic fibers in musculoskeletal tissues of Marfan syndrome: a possible mechanism of joint laxity and skeletal overgrowth. *J. Pediatr. Orthop.* 19: 283-288, 1999.

- [27] Gleason TG: Heritable disorders predisposing to aortic dissection. *Semin Thorac. Cardiovasc. Surg.* 17: 274-281, 2005.
- [28] Godfrey M: The Marfan syndrome: In *McKusick's Heritable Disorders of Connective Tissue*. Beighton P, editor. St. Louis Mosby-Year Book, Inc. 1993.
- [29] Gordon AM: Abraham Lincoln: a medical appraisal. *J. Ky. Med. Assoc.* 60: 249-253.
- [30] Guerra D, Sanchez O, Richieri-Costa A: van den Ende-Gupta syndrome of blepharophimosis, arachnodactyly and congenital contractures. *Am. J. Med. Genet.* 136: 377-380, 2005.
- [31] Hassad S, Shewmake K, Teo C et al: Shprintzen-Goldberg syndrome with osteopenia and progressive hydrocephalus. *Am. J. Med. Genet.* 70: 450-453, 1997.
- [32] Judge DP, Dietz HC: Marfan's syndrome. *Lancet* 366: 2965-1976, 2005.
- [33] Kamischke A, Baumgardt A, Horst J, Nieschlar E: Clinical and diagnostic features of patients with suspected Klinefelter syndrome. *Androl.* 24: 41-48, 2003.
- [34] Krasemann T, Kotthoff S, Kehl HG et al: Cardiac transplantation in neonatal Marfan syndrome—a life-saving approach. *Thorac. Cardiovasc. Surg.* 53 Suppl2: S146-148, 2005.
- [35] Le Parc JM, Plantin P, Jondeau G et al: Bone mineral density in 60 adult patients with Marfan syndrome *Osteoporos. Int.* 10: 475-479, 1999.
- [36] Lipton GE, Guille JT, Kumar SJ: Surgical treatment of scoliosis in Marfan syndrome: guidelines for a successful outcome. *J. Pediatr. Orthop.* 22: 302-307, 2002.
- [37] Marfan AB: Un cas de deformation congenitale des quatre membres plus prononcee aux extremités caracterisee par l'allongement des os ave un certain degre d'amincissement. *Bul. Mem. Soc. Med. Hop. Paris* 13:220-226, 1896.
- [38] Maumenee IH: The eye in the Marfan syndrome. *Trans. Am. Ophthalmol. Soc.* 79: 684-733, 1981.
- [39] McKusick VA: The cardiovascular aspects of Marfan's syndrome. *Circulation* 11: 321-342, 1955.
- [40] McKusick VA: The Marfan syndrome: In *Heritable Disorders of Connective Tissue*. Third Edition. St. Louis, CV Mosby Company, 1966. 38-149.
- [41] Mery H, Babonneix L: Un cas de deformation congenitale des quatre membres: hyperchondroplasie. *Bull. Mem. Soc. Med. Hop. Paris* 19: 671-676, 1902.
- [42] Milewicz DM, Dietz HC, Miller DC: Treatment of aortic disease in patients with Marfan syndrome *Circulation* 111: e150-157, 2005.
- [43] Mizuguchi T, Collod-Beroud G, Akiyama T et al: Heterozygous TGFBR2 mutations in Marfan syndrome. *Na. Genet.* 36: 855-860, 2004.
- [44] Morava E, Dinopoulos A, Kroes HY et al: Mitochondrial dysfunction in a patient with Joubert syndrome. *Neuropediatrics* 36: 214-217, 2005.
- [45] Murdoch JI, Walker BA, Halpern BL et al: Life expectancy and causes of death in the Marfan syndrome. *N. Engl. J. Med.* 286: 804-808, 1972.
- [46] Neptune ER, Frischmeyer PA, Arking DE et al: Dysregulation of TGF-beta activation contributes to the pathogenesis in Marfan syndrome. *Nat. Genet.* 33: 407-411, 2003.
- [47] Nollen GJ, Mulder BJ: What is new in the Marfan syndrome? *Int. J. Cardiol.* 97: Suppl 1: 103-108, 2004.
- [48] Noordam C, van Daalen S, Otten BJ: Treatment of tall stature in boys with somatostatin analogue 201-995: effect on final height. *Eur. J. Endocrinol.* 154: 253-257. 2006.

- [49] Peters K, Apse K, Blackford A et al: Living with Marfan syndrome: coping with stigma. *Clin. Genet.* 68: 6-14, 2005.
- [50] Purandare KN, Markar TN: Psychiatric symptomatology of Lujan-Fryns syndrome: an X-linked syndrome displaying Marfanoid symptoms with autistic features, hyperactivity, shyness and schizophreniform symptoms. *Psychiatr. Genet* 15: 229-231, 2005.
- [51] Pyeritz RE: The Marfan syndrome. *Annm Rev Med* 51: 481-510, 2000.
- [52] Pyeritz RE, Fishman EK, Bernhardt BA, Siegelman SS: Dural ectasia is a common feature of the Marfan syndrome. *Am. J. Hum. Genet.* 43: 726-732, 1988.
- [53] Ramirez F, Dietz HC: Therapy insight: aortic aneurysm and dissection in Marfan's syndrome. *Nat. Clin. Prac. Cardiovasc. Med.* 1: 31-36, 2004.
- [54] Ratcliffe SG, Bancroft J, Axworthy D: Klinefelter's syndrome in adolescence. *Arch. Dis. Child* 57: 6-12, 1982.
- [55] Rios AS, Silber EN, Bavishi N et al: Effect of long-term beta-blockade on aortic root compliance in patients with Marfan syndrome. *Am. Heart J.* 137: 1057-1061, 1999.
- [56] Robinson PN, Booms P, Katzke S et al: Mutations of FBN1 and genotype-phenotype correlations in Marfan syndrome and related fibrillinopathies. *Hum. Mutat.* 20: 153-161, 2002.
- [57] Robinson PN, Neumann LM, Demuth S et al: Shprintzen-Goldberg syndrome. Fourteen new patients and a clinical analysis. *Am. J. Med. Genet.* 135: 251-262, 2005.
- [58] Rose PS, Levy HP, Ahn NU et al: A comparison of the Berlin and Ghent nosologies and the influence of duraectasia in the diagnosis of Marfan syndrome. *Genet. Med.* 2: 278-282, 2000.
- [59] Rossi-Foulkes R, Roman MJ, Rosen SE et al: Phenotypic features and impact of beta blocker or calcium antagonist therapy on aortic lumen size in the Marfan syndrome. *Am. J. Cardiol.* 83: 1364-1368, 1999.
- [60] Rozendaal L, le Cessie S, Wit JM, Hennekam RC: Growth-reductive therapy in children with Marfan syndrome *J. Pediatr.* 147: 674-679, 2005.
- [61] Saal S, Eldadah ZA, Krause WL: Mutation in fibrillin-1 and the Marfanoid craniosynostosis (Shprintzen-Goldberg) syndrome. *Nat. Genet.* 12: 209-211, 1996.
- [62] Salle V: Ueber einen Fall von angeborener abnormen Grosse der Extremitäten mit einen Akronemegalia erinnernden Symptomenkomplex. [A case of congenital abnormal size of the extremata with an acronemegalia is reminiscent of the symptom complex]. *Jahrb. Kinderheilk* 75: 540-550, 1912.
- [63] Silverman DI, Burton KJ, Gray J, et al: Life expectancy in the Marfan syndrome. *Am. J. Cardiol.* 75: 157-160, 1995.
- [64] Schwartz H: Abraham Lincoln and the Marfan syndrome. *JAMA.* 187: 473-479, 1964.
- [65] Schweitzer DN, Lachman RS, Pressman BD, Graham JM Jr.: van den Ende-Gupta syndrome of blepharophimosis, arachnodactyly and congenital contractures: clinical delineation and recurrence in brothers. *Am. J. Med. Genet. A* 118: 267-273, 2003.
- [66] Slavotinek AM, Dubovsky E, Dietz HC, Lachawan F: Report of a child with aortic aneurysm, orofacial clefting, hemangioma, upper sternal defect and marfanoid features: possible PHACE syndrome. *Am. J. Med. Genet.* 110: 283-288, 2002.
- [67] Smyth CM, Bremner WJ: Klinefelter syndrome. *Arch. Intern. Med.* 158: 1309-1314, 1998.

- [68] Sponseller PD, Jones KB, Ahn NU et al: Protrusio acetabuli in Marfan syndrome: age-related prevalence and associated hip function. *J. Bone Joint. Surg.* 88A: 486-495, 2006.
- [69] Sponseller PD, Sethi N, Caameron DE, Pyeritz RE: Infantile scoliosis in Marfan syndrome. *Spine* 22: 509-515, 1997.
- [70] Steel HHJ: Protrusio acetabuli: its occurrence in the completely expressed Marfan syndrome and its musculoskeletal component and a procedure to arrest the course of protrusion in the growing pelvis. *J. Pediatr. Orthop.* 16: 704-718, 1996.
- [71] Stern WE: Dural ectasia and the Marfan syndrome. *J. Neurosurg.* 69: 221-227, 1988.
- [72] Takaesu-Miyagi S, Sakai H, Shiroma T et al: Ocular findings of Beals syndrome. *Jpn. J. Ophthalmol.* 48: 470-474, 2004.
- [73] Van de Velde S, Fillman R, Yandow S: Protrusio acetabuli in Marfan syndrome: history, diagnosis and treatment. *J. Bone Joint. Surg.* 88A: 639-646, 2006.
- [74] Van Tongerloo A, De Paepe A: Psychosocial adaptation in adolescents and young adults with Marfan syndrome: an exploratory study. *J. Med. Genet.* 35: 405-409, 1998.
- [75] Villeirs GM, Van Tongerloo AJ, Verstraete KL, et al: Widening of the spinal canal and dural ectasia in Marfan's syndrome: assessment by CT. *Neuroradiology* 41: 850-854, 1999.
- [76] Walker BA, Beighton PH, Murdoch JL: The Marfanoid hypermobility syndrome. *Ann. Intern. Med.* 71: 349-352, 1969.
- [77] Weve H: Ueber Archnodaktylie (dystrophia mesodermalis congenita, typus Marfanis). *Arch. F Ophth.* 104: 1-46, 1931.
- [78] Yetman AT, Bornemeier RA, McCrindle BW: Usefulness of Enalapril versus Propranolol or Atenolol for prevention of aortic dilation in patients with Marfan syndrome. *Am. J. Cardiol.* 95: 1125-1127, 2005.
- [79] Yule SR, Hobson EE, Dean JC, Gilbert FJ: Protrusio acetabuli in Marfan syndrome. *Clin. Radiol.* 54: 95-97, 1999.

Chapter 50

MARFAN SYNDROME AND PERIODONTITIS

Jun-ichi Suzuki^{1,*} and Yasunobu Hirata²

¹Department of Advanced Clinical Science and
Therapeutics, The University of Tokyo, Tokyo, Japan

²Tokyo Teishin Hospital, Tokyo, Japan

ABSTRACT

Marfan syndrome (MFS) is a systemic connective tissue disorder that is caused by mutations in the extracellular matrix protein fibrillin-1. While MFS is considered to be at high risk of dental disorders and cardiovascular disease (CVD), little causal relationship has been provided to date. In this article, we reviewed the prevalence of periodontitis in patients with MFS to assess the relationship between periodontal bacterial burden and CVD in MFS patients.

1. MARFAN SYNDROME: A SYSTEMIC CONNECTIVE TISSUE DISORDER

Marfan syndrome (MFS) is an autosomal dominant disorder affecting the connective tissues [1]. Mutations in the fibrillin-1 gene are responsible for alterations of the glycoprotein fibrillin-1, which is a component of the microfibrils in the connective tissue matrices [2]. These microfibrils were included in the suspensory ligament of the lens, skeletal system, lungs, blood vessels and skin [3]. Thus, MFS is a systemic disease where the localization and degree of symptoms are individually different [4].

Cardiovascular complications, especially aortic dissection or ruptures, are the major cause of morbidity and mortality [5].

* Corresponding Author's Email: junichisuzuki-circ@umin.ac.jp (The University of Tokyo, 7-3-1 Hongo, Bunkyo-ku, Tokyo 113-8655, Japan. Tel: 81-3-5800-9116; Fax: 81-3-5800-9182).

MFS relies on defined clinical criteria (Ghent nosology) that were outlined to facilitate accurate recognition of this genetic syndrome. The Ghent criteria comprise of a set of major and minor manifestations in different body systems.

Recently, an international expert panel established a revised Ghent nosology that puts more weight on cardiovascular manifestations [6].

In addition to the systemic manifestations, MFS sometimes exhibits characteristic oral features including maxillary protrusion, high palate, crowded teeth and fragility of the temporomandibular joint [7].

However, detailed characteristics of oral features including periodontitis in MFS are still to be elucidated.

2. PERIODONTITIS: AN ORAL CONNECTIVE TISSUE DISORDER

Periodontitis is one of the most common chronic infectious diseases in humans. Pathologically, periodontitis is characterized by gingival inflammation and the loss of periodontal support tissue [8]. Periodontopathic bacteria generate host immunological inflammatory responses, thus resulting in the secretion of cytokines and matrix metalloproteinases (MMPs) [9]. This leads to the extracellular matrix destruction of the periodontal tissues [10] resulting in connective tissue disorder.

In patients with periodontitis, several inflammatory factors increase [11], meaning that systemic inflammation can be caused by periodontal infection. A strong association between dental disease and cardiovascular diseases has been demonstrated [12, 13].

Especially, periodontal disease has been reported to be an independent risk factor for cardiovascular disease [14]. Previous studies also revealed a deep relationship between periodontal diseases and abdominal aortic aneurysm (AAA) [15, 16]. Clinical investigations demonstrated that some periodontal pathogens accelerated the progression of AAA [17].

Recently, we demonstrated the pathophysiological and epidemiological relationship between specific periodontal pathogens and AAA using experimental [18-20] and clinical studies [21, 22].

3. TGF- β IS A COMMON CRITICAL FACTOR IN MFS AND PERIODONTITIS

It is well known that the elevated level of the active TGF- β in the plasma is major manifestations of MFS. Chaudhry et al. [23] demonstrated that mutation of fibrillin 1 altered intercellular communication and significantly increased TGF- β protein level in the extracellular space. TGF- β is a paracrine regulatory molecule of several processes [24].

It also enhanced collagen production and extracellular matrix (ECM) remodeling. TGF- β is produced in dimer form in the cells and is being bound with latency-associated protein (LAP) to form small latent complex (SLC). This secreted SLC is bound extracellularly to latent TGF- β binding protein (LTBP) to form large latent complex (LLC). In MFS, fibrillin 1 mutation occurs, and LLC becomes unable to attach to microfibrils. Latent form is not generated,

resulting in elevated serum TGF- β level. TGF- β joins to its dimer receptor forming a complex that induces the phosphorylation cascade [25].

In periodontitis patients, TGF- β levels were also significantly elevated in serum and saliva compared to controls [26].

It was reported that a periodontal pathogen, *Porphyromonas gingivalis* (*P.g.*) invaded aortic smooth muscle cells (SMCs) and activated TGF- β and Notch signaling pathways. These findings support the association between periodontitis and cardiovascular diseases [27].

Thus, TGF- β in saliva may serve to predict the progression of periodontitis and associated systemic inflammatory diseases.

4. CLINICAL OBSERVATION OF PERIODONTITIS IN MFS PATIENTS

De Coster et al. revealed that severe and frequent oral manifestations were observed in patients with MFS. Local hypoplastic enamel spots, root deformity, abnormal pulp shape, pulpal inclusions, calculus and gingival indices were frequent findings in MFS patients [28].

Although severe periodontitis is sometimes observed in MFS patients, little information was provided to show its morbidity in MFS patients. Judge et al. showed that individuals with MFS were at high risk of developing dental disorders [29].

However, there has been no report to reveal the severity and frequency of periodontitis in MFS patients.

Thus, we revealed the incidence and severity of periodontitis in Marfan syndrome patients [30]. The subjects were patients with MFS (n=40) and age and gender matched healthy individuals (n=14) were employed as a control group. Full-mouth clinical measurements, including number of teeth, probing of pocket depth (PD), bleeding on probing (BOP) and community periodontal index (CPI) were recorded. We revealed that MFS patients had periodontitis (CPI grade 3 and 4) more frequently than the age and gender matched control subjects.

Furthermore, MFS patients had significantly more severe periodontitis and fewer remaining teeth compared to the controls. We concluded that a high incidence of periodontitis was observed in Japanese MFS patients.

Next, we analyzed periodontitis in cardiovascular disease (CVD) patients with or without Marfan syndrome [31]. In this clinical investigation, we analyzed periodontal condition and periodontopathic pathogens. The subjects were MFS patients with CVD (n=47); age and gender matched non-MFS CVD patients (n=48) were employed as controls. Full-mouth clinical measurements and the existence of three periodontal pathogens, *Porphyromonas gingivalis*, *Aggregatibacter actinomycetemcomitans*, and *Prevotella intermedia* were measured. We revealed that MFS patients had periodontitis more frequently than the age and gender matched non-MFS control subjects. MFS patients had significantly severer periodontitis, fewer remaining teeth and deeper PD compared to the non-MFS controls.

Furthermore, the serum antibody titer level against *Prevotella intermedia* was significantly lower in MFS patients compared to the non-MFS patients.

CONCLUSION

We concluded that periodontitis might influence the pathophysiology of MFS. Periodontal pathogens might be a therapeutic target to prevent CVD development in MFS patients.

ACKNOWLEDGMENTS

This work was supported by Mitsui Life Social Welfare Foundation, Daiwa Security Health Foundation, Mitsui Sumitomo Marine Welfare Foundation, Institute of Geriatric Dentistry Foundation, Institute of St. Luka Life Science Foundation, Health Welfare Foundation, Taiyo Life Welfare Foundation, the 8020 Foundation, Terumo Science Foundation, Pfizer Health Research Foundation, and General Health Promotion Foundation. We thank Dr. Yuichi Izumi, Dr. Mitsuaki Isobe, Dr. Naho Kobayashi, Dr. Tomoya Hanatani, Dr. Norihiko Ashigaki, Dr. Asuka Sekinishi, Dr. Makoto Kaneko, Dr. Yuka Shiheido, Dr. Hiroki Sato, Dr. Atsuhiko Akai, Dr. Chisato Takamura, Dr. Masahito Ogawa, Dr. Hirofumi Zempo, Mr. Ryo Watanabe and Ms. Noriko Tamura for their excellent contributions.

Disclosures: Conflict of Interest: None declared.

REFERENCES

- [1] Dietz, H. C., Cutting, G. R., Pyeritz, R. E., Maslen, C. L., Sakai, L. Y., Corson, G. M., Puffenberger, E. G., Hamosh, A., Nanthakumar, E. J., Currstin, S. M. Marfan syndrome caused by a recurrent de novo missense mutation in the fibrillin gene. *Nature*. 1991; 13: 337-339.
- [2] Maslen, C. L., Corson, G. M., Maddox, B. K., Glanville, R. W., Sakai, L. Y. Partial sequence of a candidate gene for the Marfan syndrome. *Nature*. 1991; 13: 334-337.
- [3] Sakai, L. Y., Keene, D. R., Engvall, E. Fibrillin, a new 350-kD glycoprotein, is a component of extracellular microfibrils. *J. Cell Biol.* 1986; 13: 2499-2509.
- [4] Rybczynski, M., Bernhardt, A. M., Rehder, U., Fuisting, B., Meiss, L., Voss, U., Habermann, C., Detter, C., Robinson, P. N., Arslan-Kirchner, M., Schmidtke, J., Mir, T. S., Berger, J., Meinertz, T., von Kodolitsch, Y. The spectrum of syndromes and manifestations in individuals screened for suspected Marfan syndrome. *Am. J. Med. Genet. A*. 2008; 13: 3157-3166.
- [5] Ammash, N. M., Sundt, T. M., Connolly, H. M. Marfan syndrome-diagnosis and management. *Curr. Probl. Cardiol.* 2008; 13: 7-39.
- [6] Loeys, B. L., Dietz, H. C., Braverman, A. C., Callewaert, B. L., De Backer, J., Devereux, R. B., Hilhorst-Hofstee, Y., Jondeau, G., Faivre, L., Milewicz, D. M., Pyeritz, R. E., Sponseller, P. D., Wordworth, P., De Paepe, A. M. The revised Ghent nosology for the Marfan syndrome. *J. Med. Genet.* 2010; 13: 476-485.

- [7] Westling, L., Mohlin, B., Bresin, A. Craniofacial manifestations in the Marfan syndrome: palatal dimensions and a comparative cephalometric analysis. *J. Craniofac. Genet. Dev. Biol.* 1998; 13: 211-218.
- [8] Choi, B. K., Jung, J. H., Suh, H. Y., Yoo, Y. J., Cho, K. S., Chai, J. K., Kim, C. K. Activation of matrix metalloproteinase-2 by a novel oral spirochetal species *Treponema lecithinolyticum*. *J. Periodontol.* 2001; 72: 1594-600.
- [9] Salvi, G. E., Lang, N. P. Host response modulation in the management of periodontal diseases. *J. Clin. Periodontol.* 2005; 32 Suppl. 6: 108-129.
- [10] Ashley, R. A. Clinical trials of a matrix metalloproteinase inhibitor in human periodontal disease. SDD Clinical Research Team. *Ann. N. Y. Acad. Sci.* 1999; 878: 335-346.
- [11] Reis, C., D. A. Costa, A. V., Guimarães, J. T., Tuna, D., Braga, A. C., Pacheco, J. J., Arosa, F. A., Salazar, F., Cardoso, E. M. Clinical improvement following therapy for periodontitis: Association with a decrease in IL-1 and IL-6. *Exp. Ther. Med.* 2014; 8: 323-327.
- [12] Mattila, K. J., Nieminen, M. S., Valtonen, V. V., Rasi, V. P., Kesäniemi, Y. A., Syrjälä, S. L., Jungell, P. S., Isoluoma, M., Hietaniemi, K., Jokinen, M. J. Association between dental health and acute myocardial infarction. *BMJ* 1989; 298: 779-781.
- [13] Sakurai, K., Wang, D., Suzuki, J., Umeda, M., Nagasawa, T., Izumi, Y., Ishikawa, I., Isobe, M. High incidence of *actinobacillus actinomycetemcomitans* infection in acute coronary syndrome. *Int. Heart J.* 2007; 48: 663-675.
- [14] Mendez, M. V., Scott, T., LaMorte, W., Vokonas, P., Menzoian, J. O., Garcia, R. An association between periodontal disease and peripheral vascular disease. *Am. J. Surg.* 1998; 176: 153-157.
- [15] Kurihara, N., Inoue, Y., Iwai, T., Umeda, M., Huang, Y., Ishikawa, I. Detection and localization of periodontopathic bacteria in abdominal aortic aneurysms. *Eur. J. Vasc. Endovasc. Surg.* 2004; 28: 553-558.
- [16] Paraskevas, K. I., Mikhailidis, D. P., Giannoukas, A. D. Periodontitis and abdominal aortic aneurysms: a random association or a pathogenetic link? *Int. Angiol.* 2009; 28: 431-433.
- [17] Delbosc, S., Alsac, J. M., Journe, C., Louedec, L., Castier, Y., Bonnaure-Mallet, M., Ruimy, R., Rossignol, P., Bouchard, P., Michel, J. B., Meilhac, O. *Porphyromonas gingivalis* participates in pathogenesis of human abdominal aortic aneurysm by neutrophil activation. Proof of concept in rats. *PLoS One.* 2011; 6: e18679.
- [18] Aoyama, N., Suzuki, J., Wang, D., Ogawa, M., Kobayashi, N., Hanatani, T., Takeuchi, Y., Izumi, Y., Isobe, M. *Porphyromonas gingivalis* promotes murine abdominal aortic aneurysms via matrix metalloproteinase-2 induction. *J. Periodontal Res.* 2011; 46: 176-183.
- [19] Aoyama, N., Suzuki, J., Ogawa, M., Watanabe, R., Kobayashi, N., Hanatani, T., Yoshida, A., Ashigaki, N., Izumi, Y., Isobe, M. Clarithromycin suppresses the periodontal bacteria-accelerated abdominal aortic aneurysms in mice. *J. Periodontal Res.* 2012; 47: 463-469.
- [20] Aoyama, N., Suzuki, J., Ogawa, M., Watanabe, R., Kobayashi, N., Hanatani, T., Ashigaki, N., Sekinishi, A., Izumi, Y., Isobe, M. Toll-like receptor-2 plays a fundamental role in periodontal bacteria-accelerated abdominal aortic aneurysms. *Circ. J.* 2013; 77: 1565-1573.

- [21] Suzuki, J., Aoyama, N., Aoki, M., Tada, Y., Wakayama, K., Akazawa, H., Shigematsu, K., Hoshina, K., Izumi, Y., Komuro, I., Miyata, T., Hirata, Y., Isobe, M. High incidence of periodontitis in Japanese patients with abdominal aortic aneurysm. *Int. Heart J.* 2014 May 7. [Epub. ahead of print].
- [22] Suzuki, J., Aoyama, N., Aoki, M., Tada, Y., Wakayama, K., Akazawa, H., Shigematsu, K., Hoshina, K., Izumi, Y., Komuro, I., Miyata, T., Hirata, Y., Isobe, M. Incidence of periodontitis in Japanese patients with cardiovascular diseases - A comparison between abdominal aortic aneurysm and arrhythmia. *Heart Vessels* 2014 Apr. 16. [Epub. ahead of print].
- [23] Chaudhry, S. S., Cain, S. A., Morgan, A., Dallas, S. L., Shuttleworth, C. A., Kielty, C. M. Fibrillin-1 regulates the bioavailability of TGFbeta1. *J. Cell Biol.* 2007; 176: 355-367.
- [24] Huang, F., Chen, Y. G. Regulation of TGF-b receptor activity. *Cell Biosci.* 2012; 2: 9.
- [25] Dietz, H. C. Marfan syndrome: From molecules to medicines. *Am. J. Hum. Genet.* 2007; 81: 662-667.
- [26] H. Lönn, J., Bengtsson, T. Cytokines and chemokines are differentially expressed in patients with periodontitis: possible role for TGF-β1 as a marker for disease progression. *Cytokine.* 2014; 67: 29-35.
- [27] Zhang, B., Elmabsout, A. A., Khalaf, H., Basic, V. T., Jayaprakash, K., Kruse, R., Bengtsson, T., Sirsjö, A. The periodontal pathogen *Porphyromonas gingivalis* changes the gene expression in vascular smooth muscle cells involving the TGFbeta/Notch signalling pathway and increased cell proliferation. *BMC Genomics.* 2013; 14: 770.
- [28] De Coster, P. J., Martens, L. C., De Paepe, A. Oral manifestations of patients with Marfan syndrome: a case-control study. *Oral Surg. Oral Med. Oral Pathol. Oral Radiol. Endod.* 2002; 93: 564-572.
- [29] Judge, D. P., Dietz, H. C. Therapy of Marfan syndrome. *Annu. Rev. Med.* 2008; 59: 43-59.
- [30] Suzuki, J., Imai, Y., Aoki, M., Fujita, D., Aoyama, N., Tada, Y., Akazawa, H., Izumi, Y., Isobe, M., Komuro, I., Nagai, R., Hirata, Y. High incidence and severity of periodontitis in Marfan syndrome patients in Japan. *Heart Vessels.* 2013 Nov. 8. [Epub. ahead of print].
- [31] Suzuki, J., Imai, Y., Aoki, M., Fujita, D., Aoyama, N., Tada, Y., Wakayama, K., Akazawa, H., Izumi, Y., Isobe, M., Komuro, I., Nagai, R., Hirata, Y. Periodontitis in cardiovascular disease patients with or without Marfan syndrome - A possible role of *Prevotella intermedia*. *PLoS One.* 2014; 9: e95521.

Chapter 51

COMBINED PECTUS CORRECTION AND AORTIC VALVE SPARING ROOT REPLACEMENT IN MARFAN PATIENTS

Jean-Philippe Verhoye* and Jacques Tomasi

Department of Thoracic and Cardio-Vascular Surgery,
University Hospital of Rennes, rue Henri Le-Guilloux, Rennes, France

ABSTRACT

Introduction: Pectus deformities can coexist with cardiovascular diseases. This association is well known in tissue conjonctive disorders such as Marfan's syndrom. Combined procedures can be performed safely and represent an interesting alternative in such situations.

Valve sparing aortic root replacement has excellent long term outcomes and has become an increasingly popular alternative to aortic root replacement especially in young marfan patients to avoid lifelong anticoagulation.

We present our serie of single-stage pectus correction and cardiac surgery, and emphasize the role of aortic valve sparing interventions in such situations by a review of the literature.

Methods: A retrospective review was conducted of patients who underwent chest deformity repair and cardiac surgery at the same time from January 2007 to May 2014. All datas were collected propestively in our data base.

A review of literature was conducted to collect all published cases of combined valve sparing root replacement and correction of a chest wall deformity.

Results: Including our serie (4 patients) 12 patients underwent a combined Tirone David and chest wall deformity correction. 10 patients underwent a Nuss procedure, 2 patients a modified Ravitch procedure.

Conclusion: Combined technique of valve sparing aortic root replacement and correction of a chest wall deformity especially by Nuss technique is safe and effective. This strategy has excellent mid-term results for both aortic and chest wall pathologies.

* Corresponding Author's Email: jean-phillipe.verhoye@chu-rennes.fr (Jean-Philippe Verhoye, MD, PhD).

Keywords: Marfan syndrom, pectus, aortic valve sparing

Chest wall deformities and congenital or acquired heart disease are frequently encountered in Marfan patients. Fifteen percent of Marfan patients with aortic root dilatation presents an associated chest wall deformity such as pectus excavatum (PE) [1].

The concomitant presence of cardiac disease and pectus that both necessitate surgical intervention create a dilemma for surgeons and many questions remain controversial.

MOST OF CHEST WALL DEFORMITIES ARE PECTUS EXCAVATUM (PE)

Pectus excavatum is caused by disturbances in the growth of the sternum and costal arches. The involved cartilages can be fused, deformed or rotated. Intrinsic abnormality of the costochondral cartilages is suggested by occurrence among patients with connective tissue disorder.

PE is most frequently recognized during early childhood. During adolescent growth the severity of the depression increases until full skeletal maturity is achieved. Worsening of symptoms and cardio-pulmonary function can be observed with increasing ages. PE is more than a cosmetrical problem. The severity of defect is not correlated to the severity of symptoms.

Surgery can be considered in patients close to the age of skeletal maturity, younger patients with cardiopulmonary compromise may also be candidates for repair but a too early repair can lead to improper growth of the chest wall or recurrences.

Criteria for surgical referral are listed as below (table 1)

Table 1. Criterias for surgical referral

Criteria for surgical referral (At least 2)
-Symptomatic
-Aggravation of deformation
-Paradoxal movement of chest wall
-Haller > 3,0
-Cardiac compression
-Pulmonary compression
-Restrictive syndrome
-Mitral valve prolapsus
-Cardio-pulmonary dysfunction in exertion (VO_2 /Heart Rate)
-Dysmorphism
-Failed reparation

OPERATIVE TECHNIQUES FOR PECTUS EXCAVATUM

The first breakthrough in management came in 1949 when Ravitch described costochondral osteotomy to repair PE. Robicsek made modifications to this procedure using sternal turnover and stabilizing mesh.

In 1998 Nuss introduced a minimally invasive repair by temporarily implanting metal bars.

Johnson et al. conducted a systematic review to help physicians to choose between Nuss and Ravitch in adult and pediatric patients [2]. In adult group 262 Nuss procedure patients were compared to 498 highly modified Ravitch procedure patients. The outcome were similar 91% of good or excellent in Ravitch versus 88% in Nuss. Difference were observed in operation time (191 min vs 94 min). Non displacement complication rate for Ravitch (8%) was much lower than Nuss (21%) suggesting a tendency for fewer relevant complications with the Ravitch procedure. Nuss in adults has greater tendency for bar displacements, the use of stabilizers, suture fixation, stronger bar and placement technique must be studied. Nuss was traditionally reserved to children but Park et al. [3] shown that adult patients can also have benefit from Nuss procedure.

In pediatric groups 1500 Nuss procedure patients were compared to 1186 modified Ravitch procedure patients. Outcomes were similar 96% of good or excellent result in Ravitch versus 95% in Nuss.

Pediatric patients will have greater potential benefit from the Nuss procedure rather than the Ravitch Procedure.

In conclusion Nuss and Ravitch are viable at all ages.

The involvement of the aortic root is correlated to prognosis in Marfan patients. In such situations a prophylactic replacement of the aortic root is indicated. This intervention should be realized ideally in any condition of acute aortic syndrome.

Historically the replacement was performed by a Bentall technique. The aortic root and valve were replaced. In young patients a mechanical prosthetic valve was usually chosen. The use of life oral anticoagulation is responsible of morbidity because of thrombotic or hemorrhagic event. Prosthetic valves are more susceptible for endocarditis than native valves.

Sometimes the mechanism of aortic leak does not involve the valvular tissue and is the consequence of ectasia of the aorta. The most used technique nowadays is the Tirone David technique that is also known as the inclusion technique.

The Tirone David procedure has excellent long-term outcomes and has become an increasingly popular alternative to aortic root replacement especially in young Marfan patient to avoid lifelong anticoagulation [4].

When there is both surgical indications for chest wall deformity and heart disease, the surgeon can choose between 2 main strategies: combined repair or staged repair.

The expected advantages of a staged repair are: less bleeding, infection and surgery time and less desvascularization of the sternum and cartilages [5,6].

Cardiopulmonary function may be compromised in chest wall deformities because of compression of right chambers, reduced ventricular filling [7] and reduced cardiac output during exercise [8]. Combined procedures can improve hemodynamic after surgery but is associated with higher pain [9]. In one case an emergent PE repair was conducted immediately after sternal closure because of serious hemodynamic compromise. Cardiac compression is known to result in postoperative hemodynamic instability and impairment of pulmonary function jeopardizing a successful outcome [10].

There is actually no consensus, historically staged approaches were conducted but sporadic cases of simultaneous repair were reported with great results [11-13] (Table 2).

Table 2. Summary of cases and series of simultaneous repair

Serie, year	n	M/F	Age	Correction	Cardiac procedure
Shamberger, 1988	10			Ravitch	
Kalangos, 1995	1	1		Ravitch	
DeLeon, 1997	2		2 ± 1.5	Ravitch	ASD 2
Willekes, 1999	9			Ravitch	A
Hasegawa, 2002	12	5/7	5.6 ± 3	Ravitch	VSD 6, ASD 2, TF 2, DORV 1, CAVC 1
Okamura, 2004	1	0/1	47	Ravitch	ASD
Javangula, 2006	1	1/0	23	Ravitch	Bentall
Okay, 2008	6			Ravitch	CAD 1, ASD+VSD 1, VSD 1, MVR 1, AA 2
Ryu, 2009	1	1/0	39	Ravitch	Bentall+MVR+TVF
Kao, 2010	1	1/0	25	Nuss	ASD
Kawamura, 2011	1	0/1	32	Ravitch	VSRR
Stephens, 2012	1	1/0		Ravitch	
Casamassima, 2013	9	7/2	17.6 ± 6.12	Nuss	VSRR 4, VSRR+PFO 1, redo VSRR+PFO 1, redo bentall 1, VSRR+PFO+MVP 1, MVP+PFO 1
Schmidt, 2014	10	8/2	43 ± 17.5	Ravitch	CABG 2, aortic tube 2, MVR 3, ASD 1, AVR 1, PVR+TVR+PFO 1
Kandakure, 2014	1	0/1	3	Nuss	DORV
Our serie, 2014	11	8/3	27.3 ± 17	Nuss 6 Ravitch 5	VSRR 4, MVP+VSRR 1, MVP 2

AA aortic aneurysm. ASD Atrial Septal Defect. DORV Double outlet right ventricle. CABG coronary artery bypass grafting. CAD coronary artery disease. CAVC complete atrio-ventricular canal defect. MVP Mitral valve plasty. MVR mitral valve replacement. PFO patent foramen ovale. VSD Ventricular septal defect, VSRR Valve sparing root replacement.

METHODS

A retrospective review was conducted of patients who underwent chest deformity repair and cardiac surgery at the same time from January 2007 to May 2014. All datas were collected prospectively in our data base.

We collected demographic, operative and post operative datas.

The evaluation of satisfaction was defined in excellent if the chest wall aspect was normal, good if a minimal deformation was still visible or failed in case of recurrence or fail in reparation.

All patients were approached by a mid-sternotomy. The cardiac intervention was performed under cardio-pulmonary bypass (CPB). CPB was conducted at 32 °C. Anterograde cardioplegia was induced by a cold cristalloid solution (St Thomas).

At the end of procedure, patients were warmed to 37°C. CPB was weaned, canulas were retired. Antagonisation of heparin was performed. The chest drainage was put in place after completion of hemostasis of the operative field.

After sternal closure modified Ravitch or Nuss procedure was performed.

For the Ravitch technique the chondral cartilages was largely dissected. Chondrotomies and osteotomies was performed step by step to ensure a good mobilization of the sternum. The reconstruction was secured by metal bares (Strasbourg Thorax Osteosyntheses System—STRATOS™, MedXpert GmbH, Heitersheim, Germany).

A review of English literature was conducted to collect all published cases of combined aortic valve sparing root replacement and correction of a pectus excavatum.

RESULTS

Patient demographics and comorbidities (Table 3).

In our serie 11 patients underwent a concomitant repair of heart disease and pectus.

Mean age was 27.3 ± 17 years and the sex ratio M/F was 8/3.

Ten patients were Marfan patients. Six patients followed for aortic root aneurysm underwent an aortic valve sparing root replacement (Tirone David procedure), 5 patients underwent a mitral valve repair and 1 patient a mitral valve replacement.

3 patients underwent pectus carinatum repair (modified Ravitch) and 8 patients underwent a PE repair (6 Nuss procedure, 2 modified Ravitch procedure). One patient underwent a concomitant Tirone David, mitral valve repair and modified Ravitch procedure for pectus carinatum.

Table 3. Patients demographics. The last 4 patients below the line are patients who underwent pectus excavatum repair and aortic valve sparing root replacement

Age	Sexe	Deformation	Aortic diameter	Cardiac intervention	Chest wall reparation	CPB time	Aortic clamp time	Blood loss	Complications	Satisfaction
17	M	PC	52	Tirone David	Ravitch	190	177	330	-	Excellent
25	M	PC	51	Tirone David + MVP	Ravitch	236	208	550	Atrio-ventricular block, superficial wound infection	Excellent
63	M	PE	-	MVP	Nuss	96	68	480	-	Excellent
52	F	PC	-	MVP	Ravitch	89	58	430	-	Excellent
27	F	PE	-	MVP	Nuss	85	49	380	-	Excellent
24	M	PE	-	MVP	Nuss	104	78	560	Chronic pain	Good
8	M	PE	-	MVR	Ravitch	83	56	220	-	Good
21	M	PE	54	Tirone David	Nuss	144	127	410	-	Excellent
19	M	PE	51	Tirone David	Nuss	215	202	2100	Superficial wound infection, transfusion	Good
19	F	PE	48	Tirone David	Ravitch	210	196	890	Transfusion	Good
17	M	PE	47	Tirone David	Nuss	195	165	670	Transfusion	Excellent

A total of 4 patients underwent an aortic valve sparing root replacement and correction of PE (3 Nuss procedure and 1 modified Ravitch procedure). All patients were white and 3 were male. All of these 4 patients were Marfan patients. Mean age was 19 ± 1.63 years and mean aortic diameter was 50 mm (range 47-54).

All patients reported 2 or more symptoms, all patients complained of exertional dyspnea, 3 of exercise intolerance. One patient presented upper respiratory infection.

In the post operative course, blood loss was 1017.5 mL (range 410-2100mL).

Results were good for 2 patients and excellent for 2 patients. One patient had superficial wound infection requiring oral antibiotherapy. 3 patients required transfusion in our serie. None patient required a revision for hemostasis.

Pain was controlled by Patient Controlled Analgesia (PCA). The mean length of stay was 14.5 days (11-21). No bar displacement was observed. One patient underwent removal of the material and at 1 year of follow-up, result was excellent.

DISCUSSION

Ideally pectus should be corrected before any cardiac intervention to diminish cardiac compression. In young Marfan patients a prophylactic replacement of the aortic root can be necessary in case of aortic ectasia or aneurysm because of high risk of rupture. The association of aortic valve sparing root replacement and correction of pectus excavatum represents a safe alternative to stage approaches.

Multiple approaches have been proposed for simultaneous repair of PE and cardiac disease [12]. The presence of PE is a technical challenge in cardiac surgery since the sternum is rotated and the heart posteriorly displaced. Alternatives to midsternotomy in these conditions include lateral thoracotomy or turnover techniques [14,15]. However only midline approach insure adequate exposure of the aortic root [13,16].

Concomitant repairs have been reported and are resumed in table 2.

Casamassima [16] reported 7 patients (M/F = 6/1) who underwent concomitant PE repair by Nuss procedure and aortic valve sparing root replacement for aortic aneurysms.

Kawamura [17] reported the case of a 32 year-old female who had PE and annulo-aortic ectasia treated successfully by modified Ravitch and aortic valve sparing root replacement.

Including our series (4 patients), 12 patients successfully underwent this strategy.

Sex ratio M/F is 9/3. In 10 patients PE was treated using a Nuss procedure, modified Ravitch in the last 2 patients.

Outcomes were excellent and no patient required revision for hemostasis. In our series 2 patients needed transfusion for postoperative bleeding through chest tubes without impairing clinical results.

Bleeding is a potential complication in combined procedures after cardiopulmonary bypass. By avoiding the need for early anticoagulation, the Tirone David repair is an interesting strategy for patients sharing aortic root aneurysm and chest wall deformity. Controversy however persists considering the choice between the Ravitch and the Nuss techniques for PE correction [18,19]. Cartilage resection associated with muscular flap detachment in the Ravitch technique may present a greater risk of major bleeding for patients requiring immediate anticoagulation therapy after aortic root surgery. When the aortic valve can be preserved, Tirone David procedure is preferred compared to Bentall. Preservation of the native valve avoids the need for anticoagulation therapy and is thus a more adapted approach for combined correction of chest wall deformities. Moreover surgeons remain divided considering the use of retrosternal devices, which may dramatically impair cardiopulmonary resuscitation in case of early postoperative tamponade and circulatory arrest. Picton elaborated recommendations for resuscitation after Nuss procedure including anteroposterior placement of defibrillation paddles, exclusion of tension pneumothorax, monitoring of capnography or invasive blood pressure monitoring [20]. We add to these recommendations a second sterilized handles set to be kept to quickly remove the bar and replace it after resternotomy for major postoperative bleeding.

Our approach for the concomitant repair of aortic root dilatation and PE using respectively Tirone David and Nuss procedures is safe and effective with no apparent detriments. It achieved excellent functional and cosmetic results. Valve sparing procedures combined with mini-invasive correction of chest wall deformity may represent an interesting one-stage surgical approach in Marfan patients if validated by larger case series.

CONCLUSION

A single stage repair of both chest wall deformity and heart disease is safe and feasible in Marfan patients suffering of PE and aortic aneurysm.

The aortic valve sparing root replacement (Tirone David technique) by avoiding anticoagulation can be very interesting in such patients.

The choice of Nuss or modified Ravitch procedure remains unclear. In this small serie both strategies were safe with good or excellent results.

REFERENCES

- [1] Pyeritz RE, McKusick VA. The Marfan syndrome: diagnosis and management. *N. Engl. J. Med.* 1979;300(14):772–7.
- [2] Johnson WR, Fedor D, Singhal S. Systematic review of surgical treatment techniques for adult and pediatric patients with pectus excavatum. *J. Cardiothorac. Surg.* 2014;9(1):1–13.
- [3] Park HJ, Jeong JY, Jo WM, et al. Minimally invasive repair of pectus excavatum: a novel morphology-tailored, patient-specific approach. *J. Thorac. Cardiovasc. Surg.* 2010;139(2):379–86.
- [4] David TE. Surgical Treatment of Ascending Aorta and Aortic Root Aneurysms. *Progress in Cardiovascular Diseases* 2010;52(5):438–44.
- [5] Jones WG, Hoffman L, Devereux RB, Isom OW, Gold JP. Staged approach to combined repair of pectus excavatum and lesions of the heart. *Ann. Thorac. Surg.* 1994;57(1):212–4.
- [6] Shamberger RC, Welch KJ, Castaneda AR, Keane JF, Fyler DC. Anterior chest wall deformities and congenital heart disease. *J. Thorac. Cardiovasc. Surg.* 1988;96(3):427–32.
- [7] Miller DR, Pugh DM. Repair of ascending aortic aneurysm and aortic regurgitation complicated by acute cardiac compression by pectus excavatum in Marfan's syndrome. *J. Thorac. Cardiovasc. Surg.* 1970;59(5):673–84.
- [8] Zhao L, Feinberg MS, Gaides M, Ben-Dov I. Why is exercise capacity reduced in subjects with pectus excavatum? *J. Pediatr.* 2000;136(2):163–7.
- [9] Kandakure PR, Dhannapuneni R, Venugopal P, Franks R, Alphonso N. Concomitant congenital heart defect repair and Nuss procedure for pectus excavatum. *Asian Cardiovasc. Thorac. Ann.* 2014;22(2):212–4.

- [10] Stephens EH, Preventza O, Sarateanu CS, LeMaire SA, Coselli JS. Emergent pectus excavatum repair after aortic root replacement in Marfan patient. *J. Card Surg.* 2012;27(2):222–4.
- [11] Okay T, Ketenci B, Imamoglu OU, et al. Simultaneous open-heart surgery and pectus deformity correction. *Surg Today* 2008;38(7):592–6.
- [12] Javangula KC, Batchelor TJP, Jaber O, Watterson KG, Papagiannopoulos K. Combined severe pectus excavatum correction and aortic root replacement in Marfan's syndrome. *Ann. Thorac. Surg.* 2006;81(5):1913–5.
- [13] Willekes CL, Backer CL, Mavroudis C. A 26-year review of pectus deformity repairs, including simultaneous intracardiac repair. *Ann. Thorac. Surg.* 1999;67(2):511–8.
- [14] Ryu YG, Baek M-J, Baek M-J, et al. Simultaneous repair for aortic incompetence with annuloaortic ectasia and pectus excavatum by modified Ravitch procedure with pectus bars in an adult patient with Marfan syndrome. *J. Thorac. Cardiovasc. Surg.* 2009;137(1):e34–6.
- [15] DeLeon MM, Magliato KE, Roughneen PT, Graham L, Kudukis TM, DeLeon SY. Simultaneous repair of pectus excavatum and congenital heart disease. *Ann. Thorac. Surg.* 1997;64(2):557–9.
- [16] Casamassima MGS, Wong LL, et al. Modified Nuss Procedure in Concurrent Repair of Pectus Excavatum and Open Heart Surgery. *Ann. Thorac. Surg.* 2013;95(3):1043–9.
- [17] Kawamura T, Fukuda I, Minakawa M, Fukui K. Open door approach by partial sternotomy and sterno-costo-chondroplasty for annuloaortic ectasia with pectus excavatum. *Interact Cardiovasc. Thorac. Surg.* 2011;12(4):645–7.
- [18] Kao C-H, Tsai C-S, Huang T-W, Lee S-C, Cheng Y-L. The Nuss procedure for pectus excavatum at the time of atrial septal closure. *Ann. Thorac. Surg.* 2010;89(3):985–6.
- [19] Jaroszewski D, Notrica D, McMahon L, Steidley DE, Deschamps C. Current management of pectus excavatum: a review and update of therapy and treatment recommendations. *J. Am. Board Fam. Med.* 2010;23(2):230–9.
- [20] Picton P, Walker D, White N, Deakin CD. Cardiopulmonary resuscitation following minimally invasive repair of pectus excavatum (Nuss technique). *Resuscitation* 2003;57(3):309–10.

Chapter 52

SEVERE PERIODONTITIS IN MARFAN SYNDROME

Naoto Suda

Division of Orthodontics, Department of Human Development and Fostering
Meikai University School of Dentistry, Sakado-shi, Saitama, Japan

ABSTRACT

Marfan syndrome frequently causes cardiac complications such as, aneurysm and dilatation of the aortic root. Many Marfan syndrome patients have these cardiovascular problems, and the surgical replacement of aortic and mitral valve, and aortic roots is frequently required. In addition, cases are associated with severe periodontitis, which is a chronic inflammation of the gingiva, periodontal ligament, and alveolar bone. Because of the surgical replacement, it is essential to prevent dental infection, such as infectious endocarditis caused by the periodontitis. In Marfan syndrome, an unfavorable oral hygiene due to the crowded teeth and narrow dental arch had been thought as a cause of severe periodontitis. However, clinical and basic studies have highlighted the genetic background as a pathogenesis of the severe periodontitis. It is suggested that cell alignment and tissue architecture of periodontal ligament are impaired in the model mice of Marfan syndrome. The model mice were more susceptible to alveolar bone resorption after the infection of *Porphyromonas gingivalis*, which is known to cause chronic periodontitis. It is likely that activated TGF- β signaling upregulates IL-17 and TNF- α levels, resulted in the increased alveolar bone resorption. In this review, the perspective of the dental management and the effect of angiotensin II receptor blocker are discussed.

1. CARDIAC PROBLEMS AND ANGIOTENSIN II RECEPTOR BLOCKER (ARB) IN MARFAN SYNDROME

Marfan syndrome is an autosomal dominant connective tissue disease that affects about one in 5,000 individuals [1]. The responsible gene of this syndrome is *FBN1* which encodes the extracellular matrix protein fibrillin-1 [1]. *FBN1* mutations lead to defects in multiple organs including skeletal, cardiovascular, and ocular systems [2]. Among them, the most serious problems are seen in the cardiovascular system, such as, aortic regurgitation, aneurysm

and dissection of the aortic root, mitral valve prolapse, and mitral regurgitation, causing a short life expectancy in patients [3].

It is reported that fibrillin-1 regulates the function of endogenous transforming growth factor (TGF)- β by targeting the respective complexes to the extracellular cell matrix [4]. Studies of animal [5] and human [6] reported that TGF- β signaling drives aneurysm progression in the aorta [7]. Since Marfan syndrome patients have cardiovascular problems, the surgical replacement of aortic and mitral valve, and aortic roots is often required [3].

It is known that effects of angiotensin II are mediated by two receptors, the AT1 receptor (AT1) and the angiotensin II type 2 (AT2) receptor [8]. AT1-receptor signaling can increase the production of TGF- β ligands and receptors [9]. Angiotensin II-receptor blockers (ARBs) selectively blocking the angiotensin II binding to its receptor within the renin-angiotensin system [10]. AT1-receptor blockade decrease the TGF- β signaling, resulting in the inhibition of phosphorylation of Smad2. Recently, losartan, one of the angiotensin II-receptor blockers (ARBs) have been reported to suppress the progression of aortic root dilation by inhibiting TGF- β signaling [5, 11]. Application of ARBs are now providing great benefit to Marfan syndrome patients by improving cardiovascular conditions.

2. PERIODONTAL DISEASE IN MARFAN SYNDROME

Oral manifestations are not included as diagnostic criteria of Marfan syndrome, but it is reported that this disease is frequently affected with severe periodontitis [12-14].

Periodontitis is an inflammatory condition affecting periodontal tissues, including gingiva, periodontal ligament (PDL), and alveolar bone [15]. It is reported that approximately 15% of the adult population has an advanced form of periodontitis, causing multiple negative impacts on quality of life [16, 17]. Consequences of periodontitis include negative esthetics, and functional problems in occlusion, chewing and speaking, and finally result in tooth loss [18, 19].

Periodontitis is initiated by chronic inflammation and immune reactions to bacterial pathogens [20]. It is known that several bacteria play important roles in the pathogenesis of periodontitis, but *Porphyromonas gingivalis* playing a central role in pathogenesis of periodontitis [21].

It is reported that 87.5% of Marfan syndrome patients had periodontitis with more than 4 mm of periodontal pocket depth, while only 35.7% of healthy volunteers were with this condition [22]. Interestingly, higher percentage of periodontitis with more than 4 mm of periodontal pocket depth was seen in patients with cardiovascular disease than those without cardiovascular disease [23]. Also, there was lower percentage of remaining teeth in the former than in the latter. Many Marfan syndrome patients have these cardiovascular problems, and the surgical replacement of aortic and mitral valve, and aortic roots is often required [3]. Because of this surgical replacement, it is essential to prevent dental infection, such as infectious endocarditis caused by the periodontitis and caries.

The reason of higher percentage of severe periodontitis in Marfan is not known. However, the lower number of caries has been reported in adult Marfan syndrome patients than in healthy volunteers [12]. This implies that periodontal tissues but not teeth have structural problems causing susceptible to severe periodontitis. The abnormal alignment of collagen fibers were

observed in one of the model mice of Marfan syndrome (MgR mice). Homozygous MgR mice show that 72% of reduction in *Fbn1* (encoding mouse fibrillin-1) expression as a result of transcriptional interference by insertion of the PGK*neo*-cassette [24], and resemble the phenotype of Marfan syndrome by showing 10% longer long bones than wild-type littermates. A comparable level of type I collagen, which is the most major collagen in periodontal ligaments, was expressed in PDL-cells of homozygous MgR mice as in wild type mice [25]. However, multi-oriented collagen fiber bundles with a thinner appearance were noted in homozygous mice, whereas well-organized definite collagen fiber bundles were seen in WT mice. This suggests that normal level of fibrillin-1 is indispensable for the normal architecture of periodontal ligament.

3. ARB AND PERIODONTAL DISEASE

Telmisartan is an ARB used in the management of hypertension [26, 27] and expected as an effective drug for the management of vascular condition in Marfan syndrome [28]. This drug has a binding affinity 3,000 times higher for AT1 than receptor type 2 (AT2) [29]. Telmisartan, was examined in a mouse model of Marfan syndrome (Mg Δ). Heterozygous Mg Δ mice, another mice model of Marfan syndrome, show half level of *Fbn1* as WT mice [24]. Six-week-old male heterozygous Mg Δ and WT mice were challenged with *Porphyromonas gingivalis* with and without telmisartan application [30]. Infection of *Porphyromonas gingivalis* induced alveolar bone resorption in both heterozygous Mg Δ and wild-type mice. The amount of alveolar bone resorption was significantly larger in the former than the latter (Figure 1A).

Interleukin (IL)-17 and tumor necrosis factor (TNF)- α levels were significantly higher in infected Mg Δ mice than infected wild-type mice. Telmisartan treatment significantly suppressed the alveolar bone resorption of infected Mg mice (Figure 1A). Telmisartan also significantly decreased levels of TGF- β IL-17 and TNF- α in infected Mg Δ mice to levels seen in infected wild-type mice (Figure 1B and C). This study suggests that ARB can prevent the severe periodontitis frequently seen in Marfan syndrome.

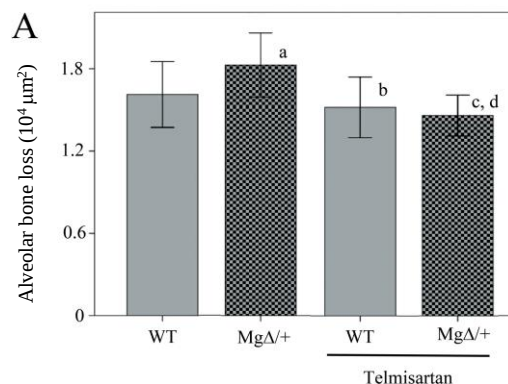


Figure 1. (Continued).

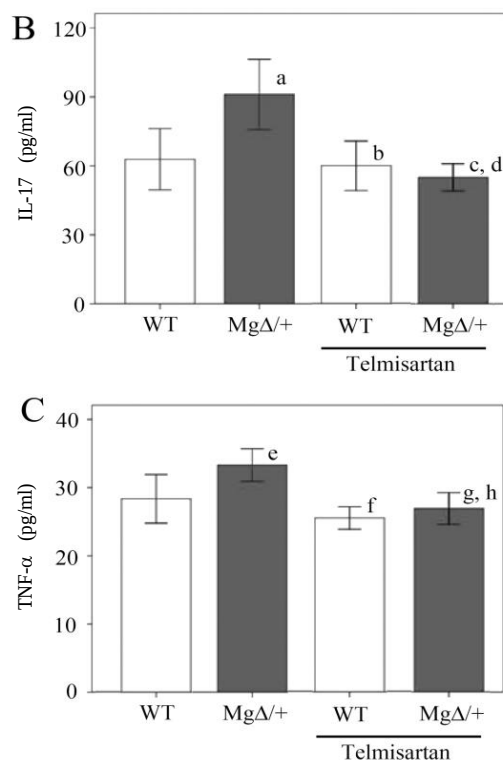


Figure 1. Alveolar bone resorption of *Porphyromonas gingivalis* infected heterozygous MgΔ (MgΔ/+) and wild-type (WT) mice with (denoted as Telmisartan) and without the application of telmisartan (A). a; significantly different ($P<0.05$) from infected WT mice without the telmisartan application, b; not significantly different from infected WT mice without the telmisartan application, c; significantly different ($P<0.05$) from infected MgΔ/+ mice without the telmisartan application, d; not significantly different from infected WT mice with the application of telmisartan. Interleukin (IL)-17 (B) and tumor necrosis factor (TNF)-α (C) in the serum of infected heterozygous MgΔ (MgΔ/+) and wild-type (WT) mice with (denoted as Telmisartan) and without the application of telmisartan. a; significantly different ($P<0.05$) from IL-17 levels of infected WT mice without telmisartan application, b; not significantly different from IL-17 levels of infected WT mice without telmisartan application, c; significantly different ($P<0.01$) from IL-17 levels of infected MgΔ/+ mice without telmisartan application, d; not significantly different from IL-17 levels of infected WT mice with the application of telmisartan, e; significantly different ($P<0.05$) from TNF-α levels of infected WT mice without telmisartan application, f; not significantly different from TNF-α level of infected WT mice without telmisartan application, g; significantly different ($P<0.01$) from TNF-α levels of infected MgΔ/+ mice without telmisartan application, h; not significantly different from TNF-α level of infected WT mice with the application of telmisartan.

REFERENCES

- [1] Dietz, H. C. (1991). Marfan syndrome caused by a recurrent de novo missense mutation in the fibrillin gene. *Nature* 352, 337-339.
- [2] McKusick, V. A. (1991). The defect in Marfan syndrome. *Nature* 352, 279-281.
- [3] Judge, D. P. & Dietz, H. C. (2005). Marfan's syndrome. *Lancet* 366, 1965-1976.

- [4] Ramirez, F. & Dietz, H. C. (2009). Extracellular microfibrils in vertebrate development and disease processes. *J. Biol. Chem.* 284, 14677-14681.
- [5] Habashi, J. P., et al. (2006). Losartan, an AT1 antagonist, prevents aortic aneurysm in a mouse model of Marfan syndrome. *Science* 312, 117-121.
- [6] Brooke, B. S. (2008). Angiotensin II blockade and aortic-root dilation in Marfan's syndrome. *N. Engl. J. Med.* 358, 2787-2795.
- [7] Neptune, E. R., et al. (2003). Dysregulation of TGF-beta activation contributes to pathogenesis in Marfan syndrome. *Nat. Genet.* 33, 407-411.
- [8] Summers, C., et al. (1991). Angiotensin II receptor subtypes are coupled with distinct signal-transduction mechanisms in neurons and astrocytes from rat brain. *Proc Natl Acad Sci USA.* 88, 7567-7571.
- [9] Habashi, J. P., et al. (2011). Angiotensin II type 2 receptor signaling attenuates aortic aneurysm in mice through ERK antagonism. *Science* 332: 361-365.
- [10] Holm, T. M., et al. (2011). Noncanonical TGF β signaling contributes to aortic aneurysm progression in Marfan syndrome mice. *Science* 332, 358-361.
- [11] Takagi, H., H. et al. (2012). An evidence-based hypothesis for beneficial effects of telmisartan on Marfan syndrome. *Int. J. Cardiol.* 158: 101-102.
- [12] DeCoster, P. J., et al. (2002). Oral manifestations of patients with Marfan syndrome: a case-control study. *Oral Surg. Oral Med. Oral Pathol. Oral Radiol. Endod.* 93, 564-572.
- [13] Shiga, M., et al. (2008). Characteristic phenotype of immortalized periodontal cells isolated from a Marfan syndrome type I patient. *Cell Tissue Res.* 331, 461-472.
- [14] Suda N, et al. (2009). Marfan syndrome and its disorder in periodontal tissues. *J Exp Zool B Mol Dev Evol.* 312B, 503-509.
- [15] Hujoel, P., et al. (2012). Historical perspectives on theories of periodontal disease etiology. *Periodontol.* 2000. 58, 153-160.
- [16] Fox, C. H. (1992). New considerations in the prevalence of periodontal disease. *Curr. Opin. Dent.* 2, 5-11.
- [17] Fox, C. H., et al. (1994). Periodontal disease among New England elders. *J. Periodontol.* 65, 676-684.
- [18] O'Dowd, L. K., et al. (2010). Patients' experiences of the impact of periodontal disease. *J. Clin. Periodontol.* 37, 334-339.
- [19] Pihlstrom, B. L. (2001). Periodontal risk assessment, diagnosis and treatment planning. *Periodontol.* 2000, 25, 37-58.
- [20] Page, R. C., & Schroeder, H. E. (1982). *Periodontitis in Man and Other Animals: A comparative review.* p. 5-41. Karger medical and scientific publishers, Basel, Switzerland.
- [21] Christersson, L. A., et al. (1991). Dental bacterial plaques. Nature and role in periodontal disease. *J. Clin. Periodontol.* 18, 441-446.
- [22] Suzuki, J. I., et al. (2014) High incidence and severity of periodontitis in patients with Marfan syndrome in Japan. *Heart Vessels.* in press.
- [23] Suzuki J., et al. (2014). Periodontitis in cardiovascular disease patients with or without Marfan syndrome--a possible role of *Prevotella intermedia*. *PLoS One* 9, e95521.
- [24] Pereira L., et al. (1999). Pathogenetic sequence for aneurysm revealed in mice underexpressing fibrillin-1. *Proc Natl Acad Sci USA.* 96, 3819-3823.

- [25] Ganburged, G., et al. (2010). Dilated capillaries, disorganized collagen fibers and differential gene expression in periodontal ligaments of hypomorphic fibrillin-1 mice. *Cell Tissue Res.* 341, 381-395.
- [26] Karlberg, B. E., et al. (1999). Efficacy and safety of telmisartan, a selective AT1 receptor antagonist, compared with enalapril in elderly patients with primary hypertension. *TEES Study Group. J. Hypertens.* 17, 293-302.
- [27] Benson, S. C., et al. (2004). *Identification of telmisartan as a unique angiotensin II receptor antagonist with selective PPARgamma-modulating activity.*
- [28] an prevents aneurysm progression *Hypertension* 43, 993-1002.
- [29] Kaschina, E., et al. (2008). Telmisartin the rat by inhibiting proteolysis, apoptosis and inflammation. *J. Hypertens.* 26, 2361-2373.
- [30] Ohno, K., et al. (2011). Unique "delta lock" structure of telmisartan is involved in its strongest binding affinity to angiotensin II type 1 receptor. *Biochem. Biophys. Res. Commun.* 404, 434-437.
- [31] Suda, N., et al. (2013) Effect of angiotensin II receptor blocker on experimental periodontitis in a mouse model of Marfan syndrome. *Infect Immun*, 81, 182-188.

Chapter 53

PREIMPLANTATION GENETIC DIAGNOSIS FOR MARFAN SYNDROME

Anver Kuliev* and Svetlana Rechitsky

Reproductive Genetics Innovations, Northbrook, IL, US

INTRODUCTION

Preimplantation genetic diagnosis (PGD) was introduced 24 years ago with the purpose of performing genetic testing before pregnancy, in order to establish only unaffected pregnancies and avoid the need for pregnancy termination, which is the major limitation of traditional prenatal diagnosis [1-2]. Despite the requirement for ovarian hyperstimulation and in vitro fertilization (IVF), needed to perform genetic testing of oocyte or embryo prior to transfer, PGD has been accepted in most parts of the world [3-5]. Thousands of PGD cycles have now been performed for single gene disorders, with PGD presently offered for some indications that have never been practiced in prenatal diagnosis, such as late onset diseases with genetic predisposition, and preimplantation HLA typing [6-9]. The present paper describes our experience on PGD for Marfan syndrome, caused by *FBN1* gene, which was performed in 38 cases, as part of our PGD experience of 2,860 cycles for single gene disorders, which is the world's largest PGD experience.

Similar to other autosomal dominant conditions, Marfan syndrome is important candidate for PGD, as couples have a 50% risk of producing an affected child.

Our experience of PGD for Marfan syndrome is presented in Table 1, summarizing the data of 38 PGD cycles performed for 20 couples at risk for producing offspring with Marfan syndrome. PGD was performed either by polar body testing, in cases of maternally derived mutations, or by embryo biopsy, involving blastomere or blastocyst biopsy [10-11]. Overall, 410 oocytes or embryos were biopsied, including 107 oocytes tested by polar bodies sampling, following oocytes maturation and fertilization, and 303 embryos tested by blastomere biopsy in 239 cases, and blastocyst biopsy in 64. Of 310 embryos with conclusive results, 158 (51%) were mutant, and 152 (49%) free of *FBN1* mutation, of which 59 were pre-selected for transfer

* Corresponding Author's Email: anverkuliev@hotmail.com (Tel: 847 400-1515; Fax: 847 400-1516).

in 30 cycles (1.9 embryos on the average), resulting in 16 (53%) clinical pregnancies and birth of 14 healthy children, confirmed to be free of Marfan syndrome.

Table 1. Preimplantation Genetic Diagnosis for Marfan Syndrome

Test Type	Patients	Cycle	ET***	#Embryos Transferred	Pregnancy	SAB#	Birth/Baby	Families with De Novo Mutations
MARFAN	9	21	20	45	11(55%)	4(25%)	7/9	7
MARFAN+A* (PCR)	6	11	9	13	4	0	4	3
MARFAN+24A**	5	6	1	1	1	0	1	1
SUBTOTAL	11	17	10	14 (1.4)	5 (50%)	0	5/5	4
TOTAL	20	38	30	59 (1.9)	16 (53.3%)	4 (25%)	12/14	11

*Aneuploidy Testing.

**24-Chromosome Aneuploidy Testing.

*** Embryo Transfer #Spontaneous abortions.

Because a half of the PGD couples were of advance reproductive age, concomitant aneuploidy testing was performed in 17 of 38 cycles, including 24-chromosome aneuploidy testing in 6 of them. It is of note that 4 of 16 pregnancies resulting in spontaneous abortions were from those cycles performed without aneuploidy testing.

It is also of interest that 11 of 20 PGD couples were with *de novo* mutations, and presented a particular challenge, as the family members were free of *FBN1* mutation to be able to establish the relevant haplotypes, required for performing PGD with sufficient level of accuracy.

Of these 11 couples, 4 were with paternally derived mutation, so we performed single sperm typing to establish the relevant haplotypes, while 7 were with maternally derived mutation, in which the polar body testing was method of choice. Irrespective of mutation origin, no misdiagnosis was observed, with confirmation of mutant gene in affected embryos avoided from transfer, and the mutation free status in all 14 children born as a result of PGD.

The presented data demonstrate the utility of PGD for Marfan syndrome, suggesting that the at risk couples should be informed about the availability of the PGD technology, which may be applied irrespective of mutation origin and also combined with aneuploidy testing to improve the outcome of PGD in at risk couples with advance reproductive age.

REFERENCES

- [1] Handyside AH, Kontogiani EH, Hardy K, Winston RML. Pregnancies from biopsied human preimplantation embryos sexed by Y-specific DNA amplification. *Nature* 1990; 344:768-770.
- [2] Verlinsky Y, Ginsberg N, Lifchez A, Valle J, Moise J, Strom CM. Analysis of the first polar body: Preconception genetic diagnosis. *Hum. Reprod* .1990; 5: 826-829.

- [3] Preimplantation Genetic Diagnosis International Society (PGDIS). Guidelines for good practice in PGD: program requirements and laboratory quality assurance. 2008. *Reprod. Bio. Med. Online* 16:134-147.
- [4] ESHRE Preimplantation Genetic Diagnosis (PGD) Consortium. Best practice guidelines for preimplantation genetic diagnosis/screening (PGD/PGS). *Hum. Reprod.* 2011; 26:14-46.
- [5] Verlinsky Y, Munne S, Cohen J. et al. (2004) Over a decade of preimplantation genetic diagnosis experience – A multi-center report. *Fertil Steril* 2004; 82:000-000.
- [6] Kuliev A.M. Expanding indications for Preimplantation Genetic Diagnosis. *Expert Rev. Obstet. Gynecol* 2011; 6:599-607.
- [7] Verlinsky Y, Rechitsky S, Schoolcraft W, Strom C, Kuliev A. Preimplantation diagnosis for Fanconi anemia combined with HLA matching. *JAMA* 2001; 285:3130-3133.
- [8] Verlinsky Y, Rechitsky S, Sharapova T, Morris R, Tharanissi M, Kuliev A. Preimplantation HLA typing. *JAMA* 2004; 291:2079-208.
- [9] Rechitsky S, Verlinsky O, Chistokhina A et al. Preimplantation Genetic Diagnosis for Cancer Predisposition. *Reproductive BioMedicine Online* 2002; 4:148-155.
- [10] Kuliev A, *Practical preimplantation genetic diagnosis*. Springer, New York, London, Heidelberg, 2012.
- [11] Kuliev A, Rechitsky S, Verlinsky O. *Atlas of Preimplantation Genetic Diagnosis*. 3rd Edition. CRS Press, Taylor and Francis, London, 2014.

Chapter 54

ANURAN CYTOGENETICS: AN OVERVIEW

***Cíntia Pelegrineti Targueta¹, Stenio Eder Vittorazzi²,
Kaleb Pretto Gatto¹, Daniel Pacheco Bruschi³,
Ana Cristina P. Veiga-Menoncello¹,
Shirlei M. Recco-Pimentel¹ and Luciana Bolsoni Lourenço^{1,*}***

¹Department of Structural and Functional Biology,
University of Campinas (Unicamp), Campinas, São Paulo, Brazil

²Department of Biological Sciences, Mato Grosso State University, Tangará da Serra,
Mato Grosso, Brazil

³Department of Genetics, Federal University of Paraná (UFPR),
Curitiba, Paraná, Brazil

ABSTRACT

The order Anura currently encompasses over 6,800 amphibian species distributed in 56 families and is an interesting group for cytogenetic studies. Whereas some groups of species present conservative karyotypes, others are highly variable in diploid number or number/location of diverse chromosomal markers, such as nucleolus organizer regions, heterochromatic bands and specific satellite DNA sites. In some cases, karyotypic variation overcomes morphological diversification, turning cytogenetics into a useful tool for taxonomy. Special variation is observed with respect to sex chromosomes and sex determination systems, with both female and male heterogameties observed. Although most species already karyotyped do not show sex chromosome heteromorphism, distinct levels of differentiation are observed between the sex chromosomes of several species, which makes this group particularly interesting for studies of sex chromosome evolution. In this chapter, we explore the use of cytogenetic data for studies of frogs as well as the insights that hypotheses of phylogenetic relationships have added to this issue. In addition, we provide a brief review of PcP190 satellite DNA (with new data for the genus *Engystomops*), sex chromosome systems and B chromosomes found in Anura.

* Corresponding Author's Email: bolsoni@unicamp.br.

1. CYTOGENETICS, TAXONOMY AND PHYLOGENETIC INFERENCES

In the last 15 years, knowledge of frog taxonomy and systematics has increased dramatically, boosted by the inclusion of DNA sequences in analyses. Phylogenetic relationships at the family level (Frost et al., 2006; Grant et al., 2006; Pyron & Wiens, 2011; Padial et al., 2014; Duellman et al., 2016; Feng et al., 2017) and especially in lower taxonomic levels (examples in Faivovich et al., 2012; Blotto et al., 2013; De Sá et al., 2014; Vacher et al., 2017; Chen et al., 2017) have been inferred, motivating several taxonomic revisions. The number of amphibian species descriptions has risen rapidly (Kohler et al., 2005), with over 100 species of frog described between January and November 2017 (based on to the list provided by Frost, 2017), but the species-level diversity of amphibians remains underestimated (see analysis provided by Giam et al., 2012; for examples of recent disclosure of still unnamed species, see Funk et al., 2012; Caminer et al., 2017; Chen et al., 2017, among others). In parallel, several species have been synonymized (examples in Guarnizo et al., 2012; Orrico et al., 2013, 2017).

In this scenario of frequent taxonomic changes and uncertainties, particular attention should be devoted to the identification of cytogenetic data available for frogs. There are several examples in the literature of species with cytogenetic data described or discussed under different names, which brings additional complexity when assembling specific cytogenetic data. That is the case, for example, for the craugastorid species currently known as *Strabomantis biporcatus*, whose karyotype was first described as belonging to *Eleutherodactylus maussi* (Schmid et al., 1992, 2002a), or the dendrobatid *Ranitomeya vanzolinii*, whose karyotype was first described by Bogart (1991) and assigned to *Dendrobates quinquevittatus* (a synonym junior of *Ameerega quinquevittatus*) (see comments in Grant et al., 2006 and Rodrigues et al., 2011). Taking into account the recent improvement in DNA barcoding of frogs (Vences et al., 2005a,b; Fouquet et al., 2007; Lyra et al., 2016), an approach that has become useful to minimize taxonomic problems is to obtain DNA sequences from the same individuals analyzed cytogenetically (Targueta et al., 2010; Medeiros et al., 2013; Veiga-Menoncello et al., 2014; Teixeira et al., 2016).

The increasing number of phylogenetic studies of frogs has also improved inferences of karyotypic evolution. The evolutionary comparison of karyotypes involves the inference of interspecific chromosomal homeologies and changes, which relies on the identification of many chromosomal markers. Therefore, the greater the number of chromosomal markers available for a representative sample of the taxonomic group of interest and its outgroup, the better the chances of obtaining informative data for the inference of karyotypical evolution. However, the chromosomal markers usually obtained in cytogenetic studies of frogs are scarce, with most of them being restricted to diploid number, chromosomal morphology (i.e., chromosome size and centromeric position), and location of heterochromatin and nucleolus organizer regions (NORs). Although some important phylogenetic groups were first recognized based on their diploid numbers, as the 30-chromosome hylids (currently the *Dendropsophus* genus – see Faivovich et al., 2005) and the *Sclerophrys regularis* group (known as the 20-chromosome toads – see Cunningham & Cherry, 2004 and references therein), the available chromosome markers have usually been insufficient for the proposition of substantial chromosomal evolutionary hypotheses. Accordingly, proper karyotypic evolutionary inferences have been facilitated by the tracking of chromosomal characters in cladograms inferred from other source

of data, mostly DNA sequence matrices (examples in Lourenço et al., 2008, 2015; Cardozo et al., 2016). This approach involves the optimization of chromosomal characters based on the parsimony criterion, and has allowed for the recognition of a number of putative chromosomal synapomorphies as well as homoplastic characters, either concerning diploid chromosome number, chromosome morphology, NOR location or heterochromatic bands, as exemplified below.

1.1. Inferences of Evolutionary Changes in Diploid Number

Diploid chromosome number is easily coded as a phenotypic character and can be readily included in data matrices used for phylogenetic inferences (see example in Grant et al., 2006). The amount of missing data regarding diploid number of frogs, however, is still extensive, and most available inferences of transformations related to this character have derived from its analysis in the light of phylogenetic hypotheses inferred previously or from other sources of data. Some examples of chromosome number transformations already inferred for anurans are listed here.

- (i) Progressive reduction in diploid number from 2n = 22 to 2n = 20, 18 and 16 in the leptodactylid genus *Pseudopaludicola* (Veiga-Menoncello et al., 2014; Cardozo et al., 2016), with a parallel origin of a 20-chromosome karyotype in *Pseudopaludicola boliviana* (Cardozo et al., 2016).
- (ii) Increase in chromosome number from 2n = 24 to 2n = 30 as a synapomorphy of the hylid genus *Dendropsophus* (Suárez et al., 2013; Faivovich et al., 2005).
- (iii) Three independent changes in diploid number from the plesiomorphic 2n = 26 to 2n = 30, 2n = 34, and 2n = 22 in the alsodid genus *Alsodes* (Blotto et al., 2013).
- (iv) Reduction from 2n = 24 to 2n = 22 due to a putative fusion of chromosomes 3 and 12 as a synapomorphy of the *Aplastodiscus albofrenatus* group (Hylidae), and independent change of the plesiomorphic condition 2n = 24 in the *Aplastodiscus albosignatus* group (sensu Berneck et al., 2016) (Gruber et al., 2012a; Berneck et al., 2016). Berneck et al. (2016) note that changes in diploid number in the *A. albosignatus* group are still ambiguous as the karyotypes of most species of this group are still unknown but consider the hypothesis of a former reduction to 2n = 20 and a sequential reduction to 2n = 18 in this group (as previously proposed by Gruber et al., 2012a).
- (v) Reduction from 2n = 22 to 2n = 20 as a synapomorphy of the Duovox clade of the leptodactylid genus *Engystomops* (Targueta et al., 2012).
- (vi) Reduction from 2n = 22 to 2n = 20 as a synapomorphy of the *Sclerophrys regularis* group (Bufonidae), with a reversion to 2n = 22 in *Sclerophrys pardalis*, was inferred in the phylogenetic study conducted by Cunningham & Cherry (2004). The authors, however, emphasize the need of further analysis to verify this hypothesis of reversion as they could not statistically discard the alternative arrangement of *S. pardalis* as the sister species of the group composed of the 20-chromosome toads.
- (vii) Increase from 2n = 24 to 2n = 28 in the hylid species *Pseudis cardosoi* as a result of two centric fission events (Busin et al., 2001; Aguiar-Jr. et al., 2007).
- (viii) Reduction from 2n = 24 to 2n = 22 in the hylids *Phyllodytes edelmi* and *Phyllodytes luteolus* (Gruber et al., 2012b). Because the interspecific relationships in *Phyllodytes*

remain unknown as well as the diploid number of other species of this genus, it is not possible to evaluate if the diploid numbers of *P. edelmoi* and *P. luteolus* have resulted from the same evolutionary event or if this condition is shared with other *Phyllodytes* species. However, the reduction in diploid number could be inferred considering that $2n = 24$ is a putative synapomorphy of the subfamily Hyliinae, as suggested by Faivovich et al. (2005) (Gruber et al., 2012b). Also based on this assumption, Gruber et al. (2007) inferred the reduction from $2n = 24$ to $2n = 22$ in the hylid species *Boana albopunctata*.

Although some rearrangements and chromosomes have been considered as possibly involved in some of the aforementioned cases (see discussions in Gruber et al., 2007, 2012a,b; Cardozo et al., 2016), only the diploid reduction observed in *Pseudis cardosoi* could be associated with an explicit proposal of chromosome rearrangements (Busin et al., 2001). In this case, based on chromosome morphology, C-bands and NOR site, Busin et al. (2001) inferred a reliable hypothesis of homeology between the chromosomes of *P. cardosoi* and its sister species *P. minuta* (see Aguiar-Jr. et al., 2007), according to which the bi-armed chromosomes 1 and 4 of *P. minuta* would be homeologous, respectively, to the telocentric chromosomes 6 and 7 and chromosomes 8 and 9 of *P. cardosoi*. Based on this hypothesis of homeology, and considering $2n = 24$ as the plesiomorphic state for the group in analysis, Busin et al. (2001) suggested that chromosomes 6, 7, 8 and 9 of *P. cardosoi* have arisen from centric fissions of the same ancestral chromosomes that have led to chromosomes 1 and 4 of *P. minuta*.

Phylogenetic inferences have also assisted the analysis of changes in chromosome number that are derived from polyploidization. In the study of the diploid *Dryophytes chrysoscelis* and the tetraploid *Dryophytes versicolor* (as *Hyla versicolor*), the phylogenetic analyses supported the hypothesis that *D. versicolor* is possibly polyphyletic, as tetraploids were inferred to have arisen multiple times in this group (Ptacek et al., 1994; Holloway et al., 2006). Another interesting example is found in the genus *Xenopus*, in which independent and recurrent polyploidization events have been inferred (Evans et al. 2004, Evans, 2007, 2008).

1.2. Inferences Related to Chromosomal Morphology or Chromosomal Markers

In contrast to coding of diploid chromosome number, coding of other chromosome characters (such as chromosome morphology, chromosome markers or even chromosome rearrangements) is more complex and depends on reliable hypotheses of interspecific chromosome homeology. Usually, the description of frog karyotypes does not provide detailed characterizations of each chromosome, as the available information is typically limited to size, centromeric position (which allows chromosome classification as metacentric, submetacentric, subtelocentric, and telocentric, following the criteria proposed by Green & Sessions, 1991), and presence of C-bands and NORs. The paucity of informative markers prevents proper comparisons among karyotypes and as a result many karyotypes were described without any discussion of primary hypotheses of interspecific chromosome homeology, with chromosomes arranged by size and numbered in decreasing order. In addition, as different chromosomes of the same karyotype may present similar lengths and centromeric positions, this criterion of chromosome classification and karyogram organization has proved to be ineffective as it may

cause errors in chromosome numbering, which adds further problems for the comparison of described karyotypes.

Measurement of chromosomes using specific software (as MicroMeasure – Reeves & Tear, 2000; IdeoKar – Mirzaghaderi & Marzangi, 2015; KaryoType – Altinordu et al., 2016; Drawid – Kirov et al., 2017) may help minimize huge discrepancies in karyotype descriptions, but it does not eliminate the problem. Taking measurement of centromeric regions or curved chromosome arms may be complicated and may vary from one research study to another. In addition, chromosome compaction degree varies amongst different metaphase cells, even in the same individual, and sometimes even between the homologues in a single cell, which may interfere in the calculation of chromosome relative length (RL) and arm ratio (AR) or centromeric index (CI). Special attention should be given to borderline values when ordering the chromosomes by size or when classifying the chromosomes with respect to their centromeric positions. The selection of good-quality metaphases, avoiding supercompacted, curved or crossed chromosomes, is of fundamental importance when measuring chromosomes and could minimize some measurement problems. Nevertheless, chromosome measurements are not accurate enough to allow for the usage of RL, AR and CI as powerful parameters for karyotypic comparisons. Although important, these parameters should not be the only or the first criterion for assessing chromosome pairing or organizing karyograms. It is not rare that the best hypothesis of arrangement of Giemsa-stained chromosome pairs in a karyogram proves to be wrong after the same chromosomes are C-banded or silver stained.

The sequential analysis of the same chromosomes submitted to different techniques is therefore indicated as a powerful approach for a better characterization of each chromosome of a karyotype. Sometimes different chromosomes that are very similar in morphology bear distinct markers and, in these cases, if the techniques that enable the detection of these markers are not applied sequentially, a proper interpretation of the karyotype is precluded. This is the case, for instance, for chromosomes 3 and 4 of the leptodactylid *Physalaemus ephippifer*. These chromosomes are very similar in size and centromeric position and may be easily misidentified in Giemsa-stained metaphases, but they may be differentiated by the presence of a pericentromeric DAPI-positive C-band in the short arm of chromosome 3, which is also detected by 5S rDNA and PcP190 satellite DNA probes (Nascimento et al., 2010; Vittorazzi et al., 2014a).

Therefore, when assessing interspecific chromosome homeologies for further identification of shared characters, chromosome morphology may help, especially when all (or most) chromosomes are similar between the karyotypes in comparison, but the analysis of specific chromosome markers is indispensable. Using this approach, some reliable primary homeologies may be recognized even between chromosomes identified by different numbers according to their sizes and, especially in these cases, the homeology hypotheses should be explained (for examples, see discussion about the NOR-bearing chromosomes found in *Paratelmatobius* and *Scythrophrys*, and in species of the hyliid subfamily Scinaxinae – Lourenço et al., 2008 and Cardozo et al., 2011, respectively). In addition, when describing new karyotypes, it may be easier if the arrangement of chromosome pairs in karyograms reflects the homology hypotheses considered by the authors, even if the chromosome sizes suggest a distinct numbering. This type of decision was made when karyotypes of some species of *Physalaemus* were described (Tomatis et al., 2009; Vittorazzi et al., 2014b).

Once reliable chromosomal characters are recognized, interesting chromosomal synapomorphies, homoplasies, and evolutionary changes may be inferred. Some illustrative

cases are found in the hylid Scinaxinae, which is currently composed of the genera *Sphaenorhynchus*, *Oloolygon*, *Julianus*, and *Scinax* (Duellman et al., 2016), and in the leptodactylid genus *Physalaemus*. Cardozo et al. (2011), taking into account the phylogenetic relationships proposed by Faivovich (2002) and Faivovich et al. (2005), inferred the submetacentric morphology of chromosome pairs 1 and 2 and the NOR in chromosome pair 6 as three synapomorphies of *Oloolygon* (referred as *Scinax catharinae* clade in Cardozo et al., 2011). In addition, Cardozo et al. (2011) also recognized the NOR in chromosome pair 11 of *Oloolygon canastrensis* as a reversion to the plesiomorphic state of this character.

With respect to *Physalaemus*, by tracing cytogenetic characters in a cladogram inferred from DNA sequences, a telocentric chromosome pair 11 was inferred as a synapomorphy of the *Physalaemus signifer* clade, and this character was confirmed as homoplastic in relation to the telocentric chromosome 11 of *P. fernandezae* (Lourenço et al., 2015 and references therein), as formerly hypothesized by Tomatis et al. (2009). The phylogenetic relationships inferred for this genus also allowed for the recognition of a large C-band of the short arm of chromosome 3 as a synapomorphy of the clade composed of *Physalaemus biligonigerus*, *P. marmoratus* and *P. santafecinus*, and as a homoplasy in relation to the large C-band found in the short arm of chromosome 3 of *P. nattereri* (Lourenço et al., 2015 and references therein), as previously suspected (Vittorazzi et al., 2014b). Additionally, the interstitial C-band in the metacentric chromosome 5 was confirmed as a synapomorphy of the *P. cuvieri* species group (Lourenço et al., 2015 and references therein).

2. CYTOGENETICS AS A HELPFUL TAXONOMIC TOOL

Some groups of species present a high degree of karyotypic similarity. The bufonid species *Rhinella crucifer*, *R. icterica*, *R. jimi*, *R. rubescens*, and *R. schneideri* provide an evident example of this, as their karyotypes could not be distinguished by chromosome morphology or position of NOR and C-bands (Kasahara et al., 1996, as species of *Bufo*; Amaro-Ghilardi et al., 2008, as species of *Chaunus*). On the other hand, in some cases notorious interspecific differences are noted, and sometimes karyotypic variation overcomes morphological diversification, turning cytogenetics into a useful tool for taxonomy.

The finding of a new diploid number ($2n = 30$) in the genus *Alsodes*, for example, helped in the description of *Alsodes norae* (Cuevas, 2008). In the case of the hylid species *Dendropsophus nanus* and *D. sanborni*, which are very similar morphologically, cytogenetic analyses assisted in taxonomic identification, as their karyotypes could be easily distinguished by the number of telocentric chromosomes (the fundamental number is 52 in *D. nanus* and 50 in *D. sanborni* – Medeiros et al., 2003).

In some cases, conspicuous karyotypic divergences are observed between specimens assigned to a single species, supporting suspicions of the presence of distinct unnamed taxa. For instance, in the leptodactylid genus *Engystomops*, cytogenetic signatures were found for *E. freibergi* and *E. petersi*, and two additional karyotypes were described for specimens that belongs to distinct clades as inferred from DNA sequence analyses, which likely correspond to still unnamed species (Targueta et al., 2010; Funk et al., 2012).

The *Pseudopaludicola* genus gives other striking examples of interspecific karyotypic divergences, with the first reports of diploid number variation dating from the 1960s and 70s (see Veiga-Menoncello et al., 2014 for references). Despite this, for long time the cytogenetics

of *Pseudopaludicola* was underexplored. After obtaining well-characterized samples, some of them coming from topotype specimens, the diploid chromosome number could be confirmed for *P. falcipes* ($2n = 22$), *P. mineira* ($2n = 22$), *P. saltica* ($2n = 22$), *P. ameghini* ($2n = 20$), *P. ternetzi* ($2n = 20$), *P. canga* ($2n = 18$), and *P. mystacalis* ($2n = 16$), and some karyotypes had to be assigned to unnamed species (Duarte et al., 2010; Fávero et al., 2011). Three of these unnamed species were described latter as *P. facureae* (Andrade & Carvalho, 2012), *P. atragula* (Pansonato et al., 2014), and *P. motorzinho* (Pansonato et al., 2016). In addition, based on difference of chromosome number, *P. ameghini* was removed from synonymy with *P. mystacalis* (Fávero et al., 2011), which contributed for solving recurrent taxonomic conflicts. Based on how informative is the diploid number for *Pseudopaludicola*, the inclusion of karyotypic information in the description of new species may be helpful and has been adopted by some authors (see descriptions of *P. murundu* – Toledo et al., 2010; *P. jaredi* – Andrade et al., 2016; and *P. ibisoroca* – Pansonato et al., 2016).

3. HOW TO STUDY FROG CHROMOSOMES

Giemsa staining, C-banding, silver staining and *in situ* hybridization to reveal NORs, and, to a lesser extent, staining with base-specific fluorochromes (especially 4',6-diamidino-2-phenylindole – DAPI, chromomycin A3 – CMA3, and mythramycin – MM) are techniques commonly applied to the study of frog karyotypes. Unfortunately, informative markers are not easily obtained throughout euchromatic regions of anuran chromosomes, although very interesting data have been generated by replication banding using BrdU incorporation in some species (example in Schmid & Steinlein, 2015).

Given the importance of chromosomal markers for reliable interspecific karyotypic comparisons (as discussed in item 1.2), the prospect of new chromosome markers may be very helpful. In this context, besides the chromosome mapping of satellite DNA (as the PcP190 satellite DNA – see item 4), the production of chromosome probes from microdissected or flow-sorted material arises as an interesting strategy. Although chromosome painting has been frequently used for cytogenetic studies of several animals and plants (Hassanane et al., 1998; Shibata et al., 1999; Karamysheva et al., 2002; Marchal et al., 2004; Rubtsov et al., 2004; Diniz et al., 2008; Teruel et al., 2009a; Teruel et al., 2009b; Wang et al., 2009; Henning et al., 2011; Vicari et al., 2011; Kawagoshi et al., 2012; Silva et al., 2016; Pita et al., 2017; Romanenko et al., 2017), few studies have employed this technique for anuran cytogenetics (Krylov et al., 2010; Gruber et al., 2014; Uno et al., 2015; Targueta et al., in press).

Comparative gene mapping is another valuable strategy as it may identify syntenic blocks that are conserved in different species. Uno et al. (2013), for example, based on the chromosome mapping of 60 genes, recognized homeologous chromosome groups between *Xenopus tropicalis* and the allotetraploid species *Xenopus laevis* and inferred the occurrence of inter- and intrachromosomal rearrangements during the evolution of these frogs that could not be inferred by other approaches.

Finally, the comparative analysis of genomes has emerged as a powerful tool, as it allows for the identification of specific regions (sex specific, for example), which can be further mapped onto chromosomes. This promising approach has greatly improved the studies of sex chromosomes, as discussed in item 5 of this chapter.

4. PCP190 SATELLITE DNA

4.1. Satellite DNA as Chromosome Markers

Satellite DNA (satDNA) comprises tandemly repetitive sequences in genomes and is mostly located in centromeric and telomeric regions (Charlesworth et al., 1994; Elder & Tuner, 1995; Richard et al., 2008; López-Flores & Garrido-Ramos, 2012). These sequences can be present in large quantity in genomes, and several distinct families of satDNA can coexist in the same genome (Slamovits & Rossi, 2002; Plohl et al., 2008, 2012). An important feature of satDNA repeats is their non-independent (concerted) evolution, which may lead to a high level of sequence homogeneity among the repeats of a satDNA family (Dover, 1982, 1986; review by Plohl et al., 2008). Repeat homogeneity is achieved by mechanisms of non-reciprocal sequence transfer (such as unequal crossover, gene conversion, rolling circle replication and reinsertion, and transposition), and higher sequence similarity is expected among adjacent repeats than among repeats from different arrays of the same chromosome, repeats positioned on homologous chromosomes and repeats on non-homologous chromosomes, consecutively (Plohl et al., 2008, 2012). Additionally, the size of the repetitive blocks may vary greatly in the same genome and between organisms, even those related to functional centromeres (Warburton et al., 2000; Henikoff et al., 2001; Plohl et al., 2012).

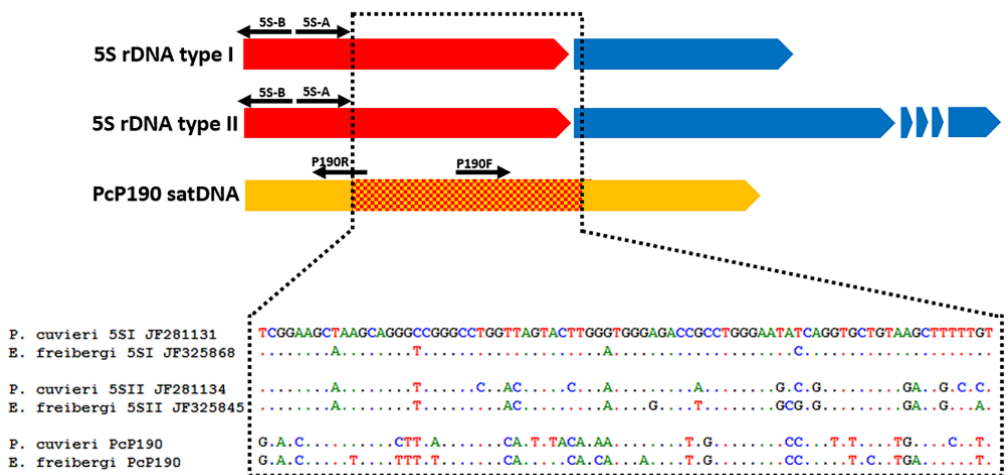


Figure 1. Comparison between type I and II 5S rDNA and PcP190 satDNA. Type I and type II 5S rDNA coding regions are shown in red, and non-transcribed regions (NTS) are shown in blue. Type I and type II 5S rDNA differ mostly in length and composition of NTS. An alignment of type I and type II 5S rDNA and PcP190 sequences from *P. cuvieri* and *E. freibergi* is also shown. GenBank accession numbers of the sequences are shown. Annealing regions of primers frequently used to isolate 5S rDNA (5S-A and 5S-B primers – Pendas et al., 1994) and PcP190 sequences are indicated (P190F and P190R – Vittorazzi et al., 2011).

Among evolutionarily closely related species, it is possible that the same satellite DNA family has different numbers of repeats, distributions and compositions, since the evolutionary dynamics of these repetitive sequences favors amplification and continuous deletion in the genome of organisms, so new families arise continuously by restructuring older sequences (Meštrović et al., 1998; Ugarkovic & Plohl 2002; Plohl et al., 2008).

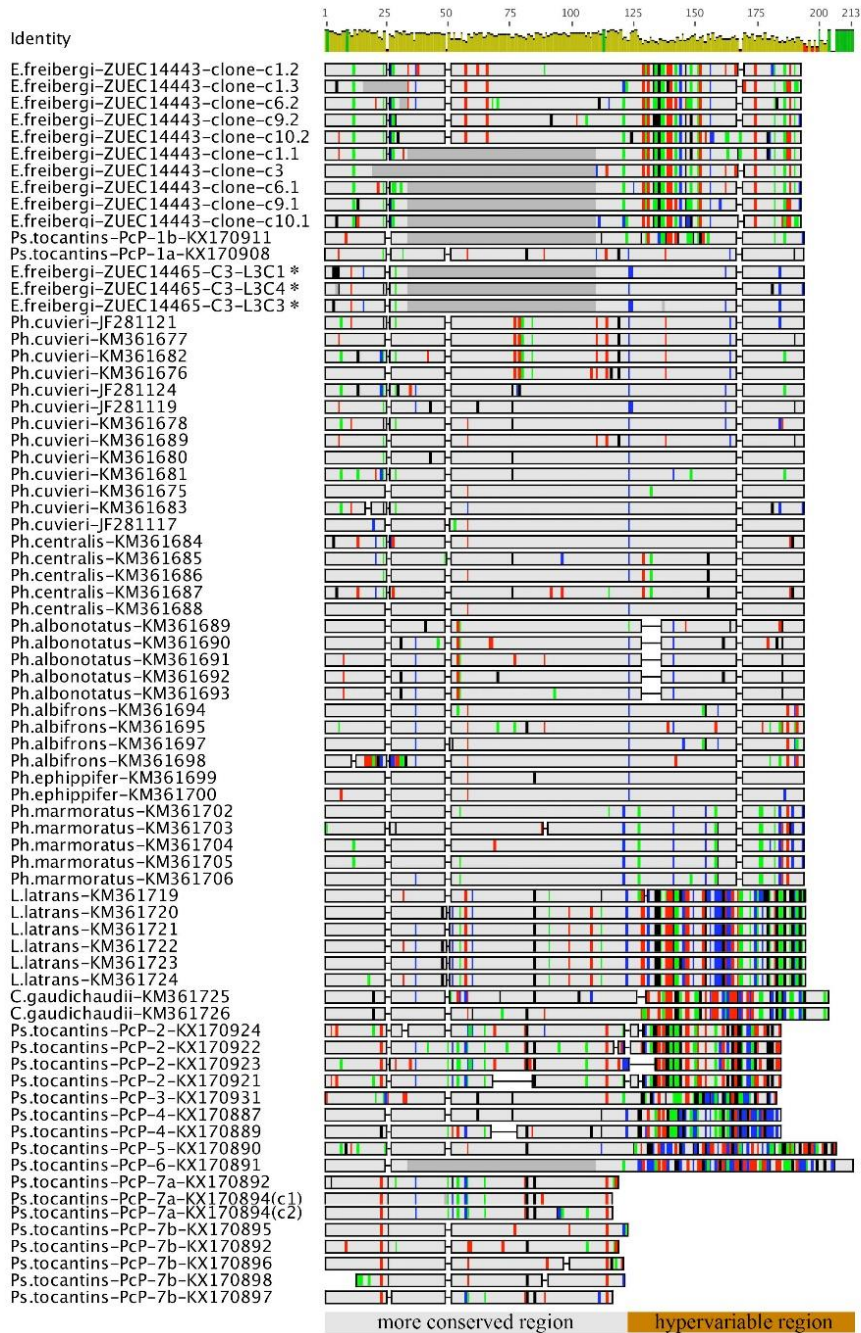


Figure 2. Alignment of all PcP190 sequences already described for anuran species. Only complete monomers are included, except for *Engystomops* and PcP-1b and PcP-6 sequences of *Pseudis tocantins*. The sequences are also identified by their GenBank accession numbers. Light gray boxes represent conservative sites and dark gray boxes are non-sequenced sites. Gaps are shown by white spaces and nucleotide polymorphisms, by colored boxes. Changes to thymine, cytosine, adenine, and guanine are represented by red, blue, green and black, respectively. Asterisks indicate the sequences obtained from microdissected chromosome 3 of *E. freibergi*. Note that these sequences are very similar to those obtained from *Physalaemus* species and to PcP-1a sequence of *Pseudis tocantins*.

Satellite DNA has been used as cytogenetic markers in comparative analysis of several organisms, such as insects, fish and mammals (Yamada et al., 2004; Saito et al., 2007; Acosta et al., 2007; Yoshimura et al., 2006; Cazaux et al., 2013; Matsubara et al., 2015; Ruiz-Ruano et al., 2016). The isolation and chromosome mapping of satellite DNA has also been employed in studies of some anurans (Table 1).

The satDNA PcP190 has proven to be widespread in frogs. It was first isolated from Leptodactylidae species (Vittorazzi et al., 2011) and has already been found in Hylodidae (Vittorazzi et al., 2014a) and Hylidae (Gatto et al., 2016) species to date. PcP190 satDNA is considered to be derived from 5S rDNA genes because of their high sequence similarity. Genetic similarities between the conserved region of Type I and II 5S rDNA and the corresponding regions in PcP190 were 70% and 66%, respectively, in *Physalaemus cuvieri* (Vittorazzi et al., 2011).

By comparing the nucleotide sequence of PcP190 fragments isolated from several frog species, two distinct regions can be recognized in the PcP190 repeats, one of which is more conserved, corresponding to the transcribing region of the 5S rDNA and one hypervariable region (Vittorazzi et al., 2014a, 2016; Gatto et al., 2016) (Figure 1). The hypervariable region varies both in size and nucleotide sequence and allows for the recognition of several different groups of sequences, including one group that is characterized by the absence of the hypervariable region (Gatto et al., 2016) (Figure 2).

While only one type of PcP190 sequence (Figure 2) was observed among the PcP190 fragments isolated from several species of *Physalaemus*, a great diversity of this satDNA was found in the hylid *Pseudis tocantins* (PcP-1 to PcP-7 sequence groups) (Figure 2; Table 1 and references therein). Different types of sequences were found adjacent to one another in the *P. tocantins* genome, supporting the occurrence of rearrangements between the distinct groups, although the homogenization process was less effective in this genome than in the *Physalaemus* genomes (Gatto et al., 2016). Based on the similarity between the coding region of 5S rDNA and the more conserved region of PcP190, and considering the great diversity among the non-transcribed regions of 5S rDNA, Gatto et al. (2016) discuss the possibility that the different groups of PcP190 sequences may have arisen from illegitimate recombination events between PcP190 satDNA and variants of the 5S rDNA. It should also be considered that the lower sequence variation in a specific region of the PcP190 satDNA repeats could be explained by differential selective pressures (Vittorazzi et al., 2014a; Gatto et al., 2016). In this context, it is worth noting that the same type of PcP190 sequences found in the *Physalaemus* species was observed in the *Pseudis* species as well (Figure 2; Gatto et al., 2016).

Clusters of PcP190 satDNA have been located at the centromeric/pericentromeric regions of some autosome pairs in seven *Physalaemus* species (*P. albifrons*, *P. albonotatus*, *P. centralis*, *P. cicada*, *P. cuvieri*, *P. ephippifer*, and *P. kroyeri*), raising the hypothesis that this satDNA could be involved in centromere biology in the studied species (Vittorazzi et al., 2014a, 2016); however, this has not been tested to date. In *Pseudis tocantins*, PcP190 satDNA was mapped exclusively to the large heterochromatic band in the long arm of the W chromosome, suggesting that this satDNA may have played an important role in the evolutionary differentiation of the sex chromosomes in this case. A differential distribution of PcP190 sequences was also observed between the sex chromosomes Z and W of *P. ephippifer*, although in this species chromosome 3 also bears a PcP190 site (Vittorazzi et al., 2014a).

Table 1. Satellite DNAs described and chromosome mapped in species of Anura

Family/Species	Satellite DNA (size of the repetitive unit)	Chromosome location	Reference
Alytidae <i>Discoglossus pictus</i>	pDS (288 bp) Dp-sat1 (510 bp) Dp-sat2 (143-148 bp) Dp-sat3 (170-177 bp)	1 to 14 int 8 to 14 cen 1 to 5 cen; 7 cen 1 to 5 cen; 7 cen	Odierna et al., 1999 Amor et al., 2009 Picariello et al., 2012 Picariello et al., 2012
Bufonidae <i>Bufotes viridis</i>	pBv (420 bp)	3p per	Odierna et al., 2004
Hylidae <i>Pseudis tocantins</i>	PcP190 (PcP-1a) (190 bp) PcP90 (PcP-1b) (?*) PcP190 (PcP-2) (166-189 bp) PcP190 (PcP-3) (181 bp) PcP190 (PcP-4) (173-181 bp) PcP190 (PcP-5) (204 bp) PcP190 (PcP-6) (?*) PcP190 (PcP-7) (107-121 bp)	- Wq int Wq int Wq int Wq int Wq int Wq int Wq int Wq int	Gatto et al., 2016 Gatto et al., 2016 Gatto et al., 2016 Gatto et al., 2016 Gatto et al., 2016 Gatto et al., 2016 Gatto et al., 2016 Gatto et al., 2016 Gatto et al., 2016
Hylodidae <i>Crossodactylus gaudichaudii</i>	PcP190 (201 bp)	-	Vittorazzi et al., 2014a
Leiopelmatidae <i>Leiopelma hochstetteri</i>	Lh1 (392 bp)	10q int (Great Barrier Island, Big Omaha)/11q int, B1, B2 per (Mt Moehau; Tapu)	Zeyl & Green, 1992
Leptodactylidae <i>Engystomops coloradorum</i> <i>Engystomops freibergi</i> <i>Engystomops guayaco</i> <i>Engystomops "magnus"</i> <i>Engystomops montubio</i> <i>Engystomops petersi</i>	PcP190 (unknown) PcP190 (189-190 bp) PcP190 (unknown) PcP190 (unknown) PcP190 (unknown) PcP190 (unknown) PcP190 (unknown)	5q per 3p per 3p per; 3q per; 5q per 5p per; 5q per 3p per; 3q per 3p per; 3q per	This work This work This work This work This work This work This work
Leptodactylidae (continued) <i>Engystomops pustulatus</i> <i>Engystomops randi</i> <i>Physalaemus albifrons</i> <i>Physalaemus albonotatus</i> <i>Physalaemus centralis</i> <i>Physalaemus cicada</i> <i>Physalaemus cuvieri</i> - BA <i>Physalaemus cuvieri</i> - MG <i>Physalaemus cuvieri</i> - MS <i>Physalaemus cuvieri</i> - RS	PcP190 (unknown) PcP190 (unknown) PcP190 (190 bp) PcP190 (183 bp) PcP190 (190 bp) PcP190 (200 bp) PcP190 (190 bp) PcP190 (190 bp) PcP190 (190 bp) PcP190 (190 bp) PcP190 (190 bp)	1p per; 1q per; 3p per; 4q per 3p per; 3q per; 5q per 3p per 3p per 1 to 5 cen; 8 cen; 10 cen; 3p; 4p per 1 cen 1 to 5cen; 3p per 1 to 7 cen; 9 to 11 cen 1 to 7 cen; 9p to 11p per 1 to 5 cen	This work This work Vittorazzi et al., 2014a Vittorazzi et al., 2014a Vittorazzi et al., 2014a Vittorazzi et al., 2016 Vittorazzi et al., 2011 Vittorazzi et al., 2014a Vittorazzi et al., 2014a Vittorazzi et al., 2014a

Table 1. (Continued)

Family/Species	Satellite DNA (size of the repetitive unit)	Chromosome location	Reference
<i>Physalaemus cuvieri</i> – PB <i>Physalaemus cuvieri</i> – TO <i>Physalaemus ephippifer</i> <i>Physalaemus kroyeri</i> <i>Physalaemus marmoratus</i> <i>Leptodactylus latrans</i>	PcP190 (190 bp) PcP190 (190 bp) PcP190 (190 bp) PcP190 (190 bp) PcP190 (190 bp) PcP190 (192 bp)	1 to 11cen 1 to 3 cen; 4p to 7p per; 10p per 3p per; Wq int 1 cen; 3p per - -	Vittorazzi et al., 2014a Vittorazzi et al., 2014a Vittorazzi et al., 2014a Vittorazzi et al., 2016 Vittorazzi et al., 2014a Vittorazzi et al., 2014a
Pipidae <i>Xenopus laevis</i>	XI32 (79 bp) Satellite 1 (741 bp) PTR-2 (388 bp) OAX (752 bp) REM 2 (487 bp) REM 3 (463 bp)	1 to 9Lpq ter, 1 to 9Spq ter - - - Disperse in almost all chromosomes 1Lp per	Spohr et al., 1981; Schmid & Steinlein, 2015 Lam & Carroll, 1983a Lam & Carroll, 1983b Ackerman, 1983 Hummel et al., 1984 Hummel et al., 1984
Ranidae <i>Glandirana rugosa</i> <i>Lithobates catesbeianus</i>	41-REL (41 bp) 31-REL (31 bp) RcS1 (360 bp) RcS2 (550 bp)	1pq to 13pq ter 1pq to 13pq ter; Wq per - -	Suda et al., 2011 Suda et al., 2011 Wu et al., 1986 Wu et al. 1986
Ranidae (continued) <i>Pelophylax esculentus</i> <i>Pelophylax lessonae</i> <i>Pelophylax ridibundus</i> <i>Rana dalmatina</i> <i>Rana graeca</i> <i>Rana holtzi</i> <i>Rana italica</i> <i>Rana macrocnemis</i> <i>Rana tavasensis</i> <i>Rana temporaria</i>	RrS1 (292 bp) Rana/PolII (227-235 bp) RrS1 (292 bp) Rana/PolII (227-235 bp) RrS1 (292 bp) Rana/PolII (227-235 bp) S1a (494 bp) S1b (332 bp) S1b (280 bp) S1a (476 bp) S1a (494 bp) S1b (280 bp) S1a (476 bp) S1a (476 bp) S1a (476 bp) S1a (476 bp)	1 to 5 cen 1 to 13 disp 1 to 13 cen 1 to 13 disp 1 to 13 cen 1 to 13 disp 1 per; 2 per, 3p to 5p per, 9q to 13q per 1 per; 2 per, 3p to 5p per, 9q to 13q per 1 per, 4p and 5p per, 8q per - 1 to 13 cen 1 to 13 cen - - -	Ragghianti et al. 1995 Ragghianti et al. 1999 Ragghianti et al. 1999 Ragghianti et al. 1999 Ragghianti et al. 1995 Ragghianti et al. 1999 Felicciello et al. 2006 Felicciello et al. 2006 Picariello et al. 2002 Picariello et al. 2016 Cardone et al. 1997 Cardone et al. 1997 Picariello et al. 2016 Picariello et al. 2016 Felicciello et al. 2005

q = long arm; p = short arm; cen = centromere; int = interstitial position; per = pericentromeric position; ter = terminal position in the chromosome; disp: disperse in the chromosome; * Only incomplete monomers were sequenced.

4.2. PcP190 satDNA in *Engystomops*: Sequences and Chromosome Locations

The Neotropical genus *Engystomops* encompasses two clades, Edentulus and Duovox, which comprise the species located in the east and west side of the Andean Mountain range, respectively (Ron et al., 2006; 2010). The Edentulus clade is composed of *Engystomops freibergi*, *E. petersi*, *E. pustulosus* and two Confirmed Candidate Species not yet formally described, *E. "magnus"* and *E. "selva"* (Funk et al., 2012; Trillo et al., 2017). The Duovox clade includes *E. randi*, *E. montubio*, *E. guayago*, *E. coloradum*, *E. pustulatus*, and *E. puyango*. A diploid number reduction from $2n = 22$ to $2n = 20$ was considered as a synapomorphy of the Duovox clade, although the chromosome rearrangements involved in this event are not known (Targueta et al., 2012).

PcP190 satDNA is present in the genus *Engystomops* and PcP190 sequences could be isolated from the genome and from microdissected chromosome 3 of *Engystomops freibergi* (Figure 2). When compared to the conserved regions of the PcP190 sequences found in the leptodactylids *Physalaemus*, *Leptodactylus latrans*, *Crossodactylus gaudichaudii*, and the hylid *Pseudis tocantins*, the conserved region of the complete PcP190 monomers of *E. freibergi* showed from 83% to 88% of similarity (Table 2). Three of the *E. freibergi* PcP190 sequences are highly similar to the PcP190 sequences of *Physalaemus* and PcP-1a sequences of *Pseudis*, even with respect to the hypervariable region (Figure 2). The remaining ten sequences, however, differ from those sequences in a 26 nucleotide-region at the beginning of the hypervariable region (Figure 2).

All species from the Edentulus clade showed a highly similar hybridization pattern with the PcP190 probe. PcP190 satDNA was mapped to a pericentromeric region of the the short arm of the chromosome identified as 3 in *E. freibergi* and *E. petersi* and identified as 5 in *Engystomops "magnus"* (Figure 3A-C). In addition, clusters of PcP190 were also detected near the centromere on the long arm of chromosomes 3 and 5 of *E. petersi* and *E. "magnus"*, respectively (Figure 3B,C).

Chromosome pairs 3 of *Engystomops freibergi* and *E. petersi* are very similar in morphology but they deeply differ from pair 3 of "*E. magnus*". Based on classical cytogenetic methods, the homeology between pair 3 of *E. freibergi* and *E. petersi* and pair 5 of "*E. magnus*" was proposed (Targueta et al., 2010), which was corroborated by the presence of a 5S rDNA site on the pericentromeric region of the short arm of these chromosomes (Rodrigues et al., 2012). Here we provide additional data to corroborate this homeology hypothesis, as chromosomes 3 of *E. freibergi* and *E. petersi* and chromosome 5 of *E. "magnus"* share the presence of PcP190 sites on both arms.

In contrast to the Edentulus clade, the karyotypes of the species of the Duovox clade are more similar to one another with respect to chromosome morphology, NOR position and heterochromatin distribution (Targueta et al., 2012) and they diverged more in relation to the location of PcP190 (Figure 3D-H). In the karyotypes of *E. montubio* (Figure 3D), *E. randi* (Figure 3F) and *E. guayaco* (Figure 3G), the PcP190 probe detected a pericentromeric region in both arms of chromosome 3. In the karyotypes of *E. randi* and *E. guayaco* (Figure 3F,G), the probe also mapped one pericentromeric region on the long arm of chromosome 5. In contrast, the accumulation of PcP190 sequences was detected only on a pericentromeric region on the long arm of chromosome 5 in *E. coloradum* (Figure 3E), whereas in *E. pustulatus* the PcP190 probe detected the pericentromeric region on both arms of chromosome 1, a

pericentromeric region on the short arm of chromosome 3 and a pericentromeric region on the long arm of chromosome 4 (Figure 3H).

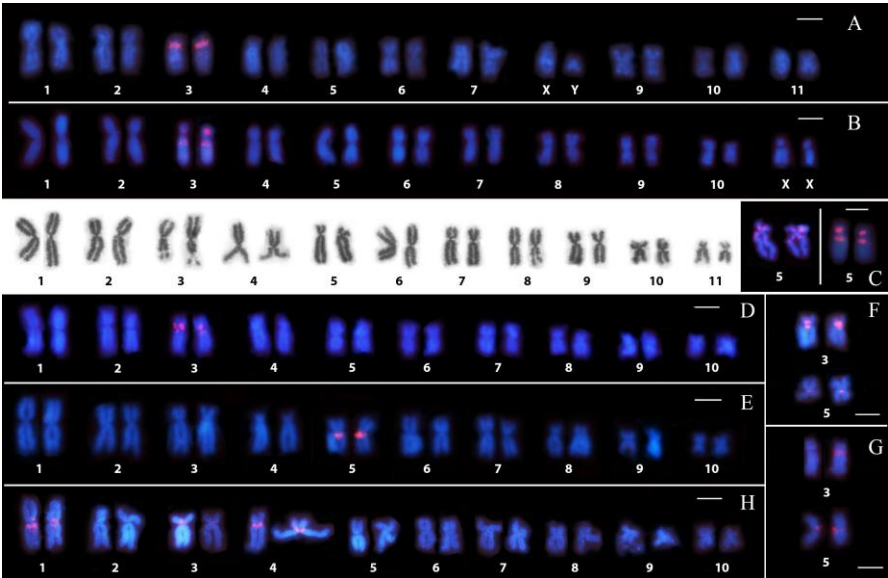


Figure 3. Chromosome mapping of PcP190 sequences in *Engystomops* species. In FISH essays, PcP190 probes were labeled with dUTP-digoxigenin and detected using anti-digoxigenin coupled to rhodamine, and the chromosomes were stained with DAPI. **A.** *E. freibergeri*. **B.** *E. petersi*. **C.** *E. "magnus"*. **D.** *E. montubio*. **E.** *E. coloradorum*. **F.** *E. randi*. **G.** *E. guayaco*. **H.** *E. pustulatus*. In **C**, the karyotype of *E. "magnus"* is stained with Giemsa for a better visualization of chromosome morphology, and chromosome pair 5 from two metaphases hybridized with PcP190 probe are shown. Note the similarity between chromosome 5 of *E. "magnus"* (**C**) and chromosome 3 of *E. freibergeri* (**A**) and *E. petersi* (**B**) with respect to morphology and location of PcP190 sites. See discussion about the possible homeology between pair 11 of *E. "magnus"* and sex chromosome pairs of *E. freibergeri* and *E. petersi* in Targueta et al. (2010). Bar: 4 μm.

Table 2. Genetic similarity (in percentage) between the conserved region of the complete PcP190 sequences isolated from *Engystomops freibergeri* (present work) and those obtained from species of *Physalaemus* (Vittorazzi et al., 2011, 2014a, 2016), *Crossodactylus gaudichaudii* (Vittorazzi et al., 2014a), *Leptodactylus latrans* (Vittorazzi et al., 2014a), and *Pseudis tocantins* (Gatto et al., 2016), grouped according to their hypervariable region. Only conserved regions of complete monomers were considered in this analysis. PcP-1b and PcP-6 sequences of *Pseudis tocantins* were not included because no complete monomer is available

	1	2	3	4	5	6	7	8
1. <i>E. freibergeri</i> sequences								
2. <i>Physalaemus</i> and <i>Pseudis</i> PcP-1 sequences	87.63							
3. <i>Pseudis</i> PcP-2 sequences	78.63	83.44						
4. <i>Pseudis</i> PcP-3 sequences	85.92	89.05	80.69					
5. <i>Pseudis</i> PcP-4 sequences	85.95	90.24	84.29	88.74				
6. <i>Pseudis</i> PcP-5 sequences	84.21	89.77	82.18	86.44	89.28			
7. <i>Pseudis</i> PcP-7 sequences	82.77	87.50	88.45	84.19	87.65	86.94		
8. <i>C. gaudichaudii</i> sequences	84.74	89.33	83.81	85.17	89.11	88.56	86.94	
9. <i>L. latrans</i> sequences	84.60	87.21	82.12	84.46	87.67	85.31	84.55	87.01

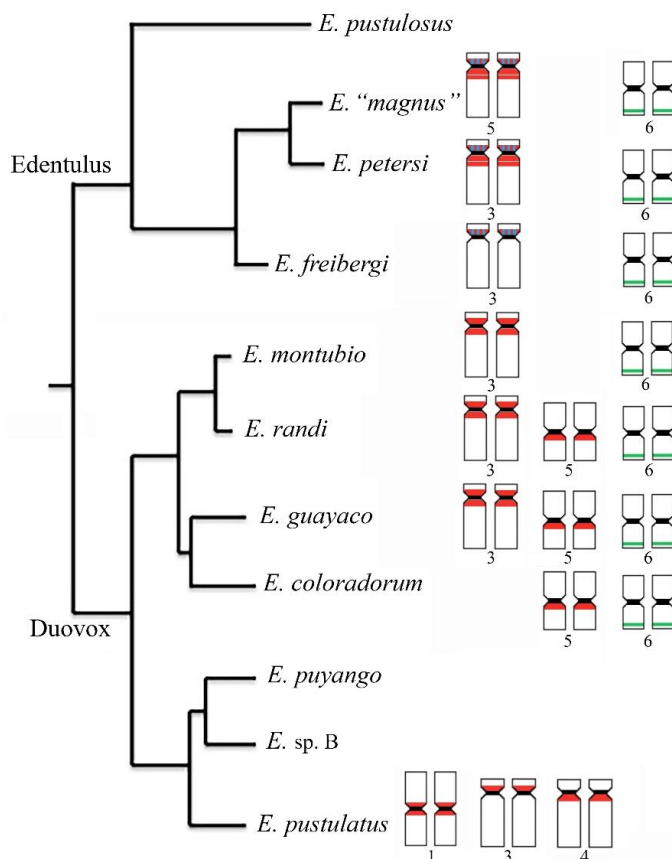


Figure 4. Chromosomal sites of PcP190 satDNA and 5S rDNA in *Engystomops*. The phylogenetic relationships in this genus are shown in the cladogram derived from the phylogenies inferred by Ron et al. (2010) and Funk et al. (2012) and ideograms show the hybridization sites of the PcP190 probe (red), type I 5S rDNA probe (blue; data obtained from Rodrigues et al., 2012) and type II 5S rDNA probe (green; data obtained from Rodrigues et al., 2012 and unpublished data) in each species. Note that type I 5S rDNA and PcP190 signals co-localize in *E. freibergi*, *E. petersi* and *E. "magnus"*. Because chromosomes 5 and 6 from species of the Duovox clade are very similar in size and centromeric position, simultaneous hybridization with the type II 5S rDNA probe and the PcP190 probe was needed to ensure that these probes detected distinct chromosomes. Based on a combined analysis of the phylogenetic and cytogenetic data, we may infer that the PcP190 cluster of chromosome 3 was lost in *E. coloradurum*. With respect to the PcP190 site of chromosome 5, one possible interpretation is that this PcP190 cluster was lost in *E. montubio* after having arisen in the common ancestor of *E. montubio*, *E. randi*, *E. guayaco*, and *E. coloradurum*. However, another hypothesis equally parsimonious is that a PcP190 cluster has arisen independently in chromosome 5 of *E. randi* and chromosome 5 of the common ancestor of *E. guayaco* and *E. coloradurum*.

Compiling all of the information regarding 5S rDNA and PcP190 sites on the karyotypes of *Engystomops* species (Figure 4), we can note that, at least in *E. freibergi*, *E. petersi*, and *E. "magnus"*, the type I 5S rDNA co-localizes with PcP190 clusters on the short arm of chromosome 3 (Rodrigues et al., 2012), as was also observed in *Physalaemus cuvieri* (Vittorazzi et al., 2011) and *P. ephippifer* (Nascimento et al., 2010; Vittorazzi et al., 2014a). This observation may contribute to the proposal that the origin of PcP190 satDNA occurred in an ancestral chromosome carrying 5S rDNA. Heteromorphic sex chromosomes X and Y are observed in *Engystomops freibergi* and *E. petersi*, and possibly heteromorphic Z and W chromosomes occur in *E. coloradurum* (Targueta et al., 2010; 2012), but none of them had

clusters of PcP190 satellite DNA revealed by FISH (fluorescent *in situ* hybridization). In these cases, sex chromosome differentiation did not involve this satDNA, in contrast to what could be inferred for *Pseudis tocantins* (Gatto et al., 2016) and *P. ehippifer* (Vittorazzi et al., 2014a).

5. SEX CHROMOSOMES IN ANURA

5.1. Sex Determination in Anurans: More Questions Than Answers

Genetic sex determination is a widely observed phenomenon in nature, mainly in animals and some plants, but only few genes are known for this function. In placental and marsupial mammals, the sex determination gene (SDG) is the gene *SRY* (sex determination region Y), a gene from the *SOX* gene family and a paralogue of the gene *SOX3* (Berta et al., 1990; Sinclair et al., 1990; Goodfellow & Lovell-Badge, 1993; Graves, 2008). In monotreme mammals, the candidate SDG is the gene *AMH*, which is located on the Y₅ chromosome (Cortez et al., 2014). For birds, characterized by female heterogamety, the SDG is the *DMRT1* (doublesex and mab-3-related transcription factor 1) gene, which is present on chromosome Z but absent on chromosome W, so that sex determination in avian species occurs by dosage of this gene (Nanda et al., 2000; Nanda et al., 2008; Smith et al., 2009). For non-avian reptiles that possess genetic sex determination, no candidate SDG is known to date. For fish species, there are two SDGs already characterized: the *DMY* gene, found in *Oryzias latipes*, a paralogue of the gene *DMRT1*; and the *AMHR2* gene, found in *Takifugu rubripes* (Paul-Prasanth et al., 2006; Matsuda et al., 2007; Graves & Peichel 2010).

For frogs, the only known SDG is the gene *DM-W*, a paralogue of *DMRT1* found on the W chromosome of *Xenopus laevis* (Yoshimoto et al., 2008; Mawaribuchi et al., 2017). *DM-W* probably arose from a duplication of *DMRT1* after the allotetraploidization event that occurred in the *X. laevis* lineage (Mawaribuchi et al., 2017). The gene *DM-W* is present in species closed related to *X. laevis*, such as *X. clivii*, *X. gilli*, *X. largeni* and *X. pygmaeus*, but it is absent in other species of the genus *Xenopus*, as in *X. tropicalis* and *X. muelleri* (Bewick et al., 2011). However, there is no evidence that *DM-W* is the SDG in all *Xenopus* species that possess it (Bewick et al., 2011). Recent studies identified sex-linked sequences in *X. borealis* that do not correspond to the *DM-W* gene and also inferred this species as the sister taxon of *X. clivii*. Because *DM-W* is present in *X. clivii*, and based on interspecific phylogenetic relationships, Furman & Evans (2016) suggested that *DM-W* arose as a sex-determining trigger in the ancestor of the group that includes *X. borealis*, *X. clivii*, *X. laevis*, *X. largeni*, and *X. allofraseri* and that *X. borealis* bears a new sex determination system that is derived with respect to the *DM-W*-based system.

In the ranid *Glandirana rugosa*, the *Sox3* gene was mapped to the sex chromosomes and has been considered as a candidate gene for sex determination, but its function is not yet elucidated (Uno et al., 2008; Miura & Ogata, 2013). Other candidate sex-determining gene is the *Dmrt1* found in some tree frog species (Dufresnes et al., 2015 and references therein).

These interesting findings concerning sex determination in frogs make evident how complex and diverse is this. Further studies on this issue will be certainly very elucidating as different sex determination systems, either with female and male heterogameties, are present in this group.

5.2. Sex Chromosome Differentiation

In organisms with genetic sex determination, the sex chromosomes harbor the sex determination gene that will trigger gonad development, in such a way that the heterogametic sex may be the male, such as in mammals (female: XX/male: XY), or female, such as in birds (male: ZZ/female: ZW) (for review, see Graves, 2008 and Schartl et al., 2016). Although sex chromosome differentiation has specific particularities in distinct lineages, some steps are common among all animals and plants that possess sex chromosomes, with the first one being the acquisition of a sex determination locus and the arrest of recombination in that region (Muller, 1964; Ohno, 1967; reviews in Graves, 2006, 2008). Suppression of recombination prevents breakage in sex determination linkage, allowing for the fixation of mutations on the chromosome exclusive of heterogametic sex (Y or W) (for review, see Graves, 2008 and Bergero & Charlesworth, 2009). Accumulation of heterochromatin by the amplification/invasion of repetitive DNA elements is a common consequence of this process, and it may also enhance the arrest of recombination between X/Z and Y/W, leading to higher levels of heteromorphism of these chromosomes (Singh et al., 1976, 1980; Charlesworth et al., 2005; Steinemann & Steinemann, 2005). In the final stages of sex chromosome differentiation, Y/W chromosomes may undergo a degeneration process, reducing in size and harboring only few genes involved with sex differentiation and/or reproduction (see Steinemann & Steinemann, 2005; Graves, 2008).

In anurans, both female and male heterogameties are present and the occurrence of heteromorphic sex chromosomes is rare. To date, among the already karyotyped anuran species, 21 species with XX/XY and 24 species with ZZ/ZW systems present heteromorphic sex chromosomes (see review by Schmid et al., 2012; Saba & Tripathi, 2014; Patawang et al., 2014; Sangpakdee et al., 2017). In addition, heteromorphic sex chromosomes were already found in multiple sex chromosome systems (derived from Y-autosome fusions) in species of the genera *Pristimantis* and *Strabomantis* (family Craugastoridae) and in *Eleutherodactylus cavernicola* (family Eleutherodactylidae) (Schmid et al., 1992, 2002a, 2003, 2010). Another uncommon sex chromosome system occurs in *Leiopelma hochstetteri*, as females has a supernumerary chromosome, denoting a 00/0W sex determination system (Green, 1988).

In addition to the diversity in sex chromosome systems, a high diversity is also observed with respect to the level of sex chromosome heteromorphism. In some cases, sex chromosome heteromorphism can be revealed only after employing specific techniques. In some anuran species, for example, the sex chromosomes can only be distinguished by differential replication patterns (Schempp & Schmid, 1981; Odierna et al., 2007). The sex chromosomes of *Xenopus laevis* can be differentiated cytogenetically only by mapping the *DM-W* gene onto chromosome W by *in situ* hybridization (Yoshimoto et al., 2008). The sex chromosomes of the bufonid *Rhinella marina* were identified by CGH (Comparative Genomic Hybridization) (Abramyan et al., 2009), and some species of the genus *Hyla* had a XX/XY sex determination system recognized only by linkage map analysis (Berset-Brändli et al., 2007). On the other hand, some species present highly differentiated sex chromosomes that differ both in morphology and molecular constitution. In *Pristimantis euphronides* and *Pristimantis shrevei*, for example, the Z chromosome is smaller than the W chromosome, which is almost entirely heterochromatic (Schmid et al., 2002a). In *Pseudis tocantins*, the submetacentric W chromosome is larger than the metacentric Z chromosome because of the presence of an amplified heterochromatin in Wq enriched for PcP190 satellite DNA (Busin et al., 2008; Gatto et al., 2016). In addition, these

sex chromosomes differ with respect to the NOR, which is interstitial in Zq and pericentromeric in Wq (Busin et al., 2008).

In most cases, however, intermediate stages of sex chromosome differentiation are noted. In some species, sex chromosomes are similar in size and centromeric position, but differ in heterochromatin accumulation and can be easily identified by C-banding (as in *Proceratophrys boiei* - Ananias et al., 2007, Amaro et al., 2012; in *Eupsophus roseus* - Iturra & Veloso, 1989; in *Gastrotheca pseutes* - Schmid et al., 1990; among several others). In other cases, sex chromosomes that are similar in size, centromeric position and heterochromatin amount can be distinguished by NOR location (as in *Dryophytes femoralis* – *Hyla femoralis* in Schmid & Steinlein, 2003; and possibly in *Engystomops coloradum* (Targueta et al., 2012). There are also cases in which sex chromosomes show similar sizes but differ with respect to centromeric position (as in *Eleutherodactylus johnstonei* – Schmid et al., 2010; and in *Pseudopaludicola saltica* – Duarte et al., 2010) or with respect to centromeric position and additional features, such as C-band distribution (as in *Eupsophus miguelli* and *Eupsophus insularis*, whose X chromosome is telocentric and possesses centromeric heterochromatin, while the Y chromosome is metacentric and lacks the centromeric positive C band – Iturra & Veloso, 1989; Cuevas & Formas, 1996). Differences in size between X/Z or Y/W have also been reported for several other species (reviewed by Schmid et al., 2012).

Because of the presence of different levels of sex chromosome heteromorphism, the frogs are an interesting group for studies on sex chromosome differentiation.

5.3. Sex Chromosome Turnover and Occasional Recombination between Sex Chromosomes

Transition between female and male heterogamety is very common in amphibians and may be observed even within the same genus. Evans et al. (2012), using tools for ancestral character state reconstruction and a molecular phylogeny proposed by Pyron & Wiens (2011), showed that there is no support to infer whether the ancestral heterogamety in frogs (or even in salamanders and amphibians) was female or male but they expanded the hypothesis first proposed by Hillis & Green (1990) that amphibians experienced several $XX/XY \leftrightarrow ZZ/ZW$ transitions. Over 30 cases of sex chromosome turnover in amphibians, including evolution of novel systems and reversals to ancestral conditions, were estimated by Evans and colleagues.

An interesting example of sex system transition is observed in *Glandirana rugosa*, as XX/XY and ZZ/ZW sex determination systems occur in geographically separated populations (Miura & Ogata, 2013 and references therein). The West-Japan and East-Japan geographic groups present homomorphic subtelocentric X and Y chromosomes, but they differ in morphology between the geographic groups, as the sex chromosomes of East-Japan specimens present a smaller short arm (Miura & Ogata, 2013 and references therein). Heteromorphic sex chromosomes are found in three other geographic groups, named the XY, ZW and Neo-ZW groups (Miura & Ogata, 2013 and references therein). Based on the analysis of mitochondrial DNA, allozymes, sex-linked genes and experimental reproductive crosses, two independent transitions from XX/XY to ZZ/ZW system could be inferred, one creating the ZW group and another giving rise to the Neo-ZW group. It is also very interesting to note that the metacentric X chromosome of the XY group and the metacentric W chromosome of both the ZW and neo-ZW groups are homeologous and probably originated from a pericentromeric inversion of the

sex chromosome of the East-Japan group. The XY group and the ZW group were created by hybridization events between West-Japan and East Japan groups, whereas the neo-ZW group was derived from re-hybridization of the XY group with the West-Japan group (Miura & Ogata, 2013 and references therein).

Transition between female and male heterogamety and translocation of the sex-determining locus to an autosome (which gives rise to a new sex chromosome pair without changing the heterogametic sex) have been frequently invoked to explain the prevalence of homomorphic sex chromosomes in some groups, such as anurans (see discussion in Volff et al., 2007; Perrin, 2009; Evans et al., 2012). Such events of sex chromosome turnover interrupt the differentiation between the previous sex chromosomes as new sex chromosome systems are achieved.

In addition to sex chromosome turnover, the fountain of youth model (Perrin, 2009) could also explain the prevalence of homomorphic sex chromosomes in anurans. According to this model, occasional recombination events between X/Z and Y/W near the sex-specific region would avoid accumulation of deleterious mutations (Muller ratchet) and arrest the degeneration process. Besides the evidence of recent events of recombination between the X and Y chromosomes of *Hyla arborea* (Berset-Brändli et al., 2008), which was already considered by Perrin (2009), new evidence of occasional sex chromosome recombination has been reported to species closely related to *H. arborea* and to *H. meridionalis* and *H. suweonensis* (Stöck et al., 2011, 2013a; Guerrero et al., 2012; Dufresnes et al., 2015). Species from other groups, such as the *Bufotes viridis* group, also agree with fountain of youth model, since sex-linked markers are not grouped by gametolog in phylogenetic analysis, but by species (Stöck et al., 2013b).

Both models (sex determining transitions and fountain of youth) are supported by strong empirical evidence, which suggests that they may act together, as argued by Dufresnes et al. (2015) for the Palearctic tree frog lineages. In this hylid group, both male heterogamety (as in *Hyla arborea* and closely related species, *Hyla meridionalis*, and *Hyla japonica*¹) and female (as in *Hyla suweonensis*) heterogamety are present, and evidence of occasional recombination events between sex chromosomes was found (Dufresnes et al., 2015 and references therein).

5.4. New Technologies in the Study of Sex Chromosomes in Anurans

Recently, next generation sequencing (NGS) has brought great progress to the study of sex chromosomes of several organisms. Different approaches using NGS have been employed and we summarize some of these as follows: (a) genome resequencing of heterogametic sex individuals and read mapping against a reference genome; (b) genome sequencing of males and females separately and comparing their *de novo* assemblies; (c) sequencing of isolated sex chromosomes obtained by laser capture microdissection or flow sorting; (d) transcriptome characterization of male and female gonads in several stages of embryonic development; and

¹ In *Hyla arborea* and its closely related species *H. intermedia*, *H. molleri*, and *H. orientalis*, the linkage group 1 (LG1) is sex linked. LG1 is also sex linked in *H. meridionalis* and in *H. suweonensis*, which presents female heterogamety. In contrast, in *H. japonica*, which is inferred to be the sister species of *H. suweonensis* and presents male heterogamety, LG1 is not sex linked. This linkage group is not sex linked either in *H. sarda*, *H. savignyi*, or *H. felixarabica* (Dufresnes et al., 2015 and references therein). An evolutionary proposal for sex chromosome turnovers in this group was presented by Dufresnes et al. (2015), who also discussed the fact that LG1 was co-opted for sex determination several times independently.

(e) genotyping by sequencing for genetic linkage map construction or identification of sex-linked markers.

Genome resequencing and read mapping against a reference genome have proven to be a reliable approach for investigating sex chromosomes, as in chicken (Chen et al., 2012) and in *Anopheles* mosquitoes (Hall et al., 2013). However, there are only three anuran genomes available to date: (a) *Xenopus tropicalis* (Hellsten et al., 2010); (b) *X. laevis* (Session et al., 2016) and (c) *Nanorana parkeri* (Sun et al., 2015). Since read mapping requires at minimum a good reference genome from a closely related species, currently this is not the best approach for anurans.

On the other hand, genotyping by sequencing (GBS; Elshire et al., 2011), which involves the sequencing of libraries of restriction fragments, has already shown interesting results in studies of sex chromosomes of frogs when data obtained from males and females are compared. Such an approach was satisfactorily used for the identification of sex-linked SNPs (single nucleotide polymorphisms) in *Xenopus borealis*, which helped Furman & Evans (2016) to propose that the DM-W gene is not the sex determination trigger in this species, although this gene is present in the other species of *Xenopus* closely related to *X. borealis*. In *Xenopus tropicalis* the data generated by restriction site-associated DNA sequencing (RAD-seq) together with the sanger sequencing of 65 amplicons supposedly linked to the sex-determining region suggested that the sex-specific region is small in this species, with large portion of its sex chromosomes corresponding to the pseudoautosomal region (Bewick et al., 2013). More recently, Brelsford et al. (2016) proposed a method for identifying sex chromosomes based on high-density linkage maps constructed from RAD-seq of individuals of a family without knowledge of offspring sexes. This method involves the identification of linkage groups (which correspond to the distinct chromosomes of the karyotype) and the construction of sex-specific maps based on the data obtained from the parents. The linkage group corresponding to the sex chromosome pair is inferred by comparing the number of informative markers per linkage group between the male- and female-specific maps. Because higher divergence is expected between the gametologs than between two chromosomes X (or Z), for the sex-specific linkage group a strong difference in the number of heterozygous markers between the male and female maps is expected, while for each autosome the number of informative markers is supposed to be similar between the two parental maps.

6. B CHROMOSOMES IN ANURA: GENERAL PROPERTIES AND EXCEPTIONS

The presence of supernumerary chromosomes, known as B chromosomes, is a characteristic found in the genomes of many eukaryote lineages, having been reported in all principal evolutionary branches, with many examples found in animals, plants, and fungi (for review, see Camacho, 2000). While we are still a long way from understanding the mechanisms involved in the origin, maintenance, and function of these chromosomes in genomes, they are characterized by three unifying features: (i) B chromosomes are dispensable to the principal genomic complement (A complement) and, when present, they may be limited to only a few populations of the species, and normally at low individual frequencies. Given this, new organisms in which these elements can be found are discovered sporadically. (ii) B

chromosomes are inherited irregularly, that is, in a non-Mendelian way. (iii) During meiosis, B chromosomes do not pair with any chromosome of the A complement. In anurans, 22 species belonging to twelve different families (Table 3) are known to have B chromosomes, but this represents only a very small proportion of the total number of anuran species karyotyped up until now.

As in other eukaryotes, the B chromosomes of anurans are typically minor genomic elements in comparison with the A chromosomes (Nur & Nevo, 1969; Schmid, 1978; Schmid et al., 1987; Baldisera et al., 1993; Rosset et al., 2006; Medeiros et al., 2006; Schmid et al., 2010; Hernández-Guzmán et al., 2011; Ferro et al., 2016). There are exceptions, however, such as in *Physalaemus feioi* (Milani et al., 2010), *Oreobates discoidalis* (Schmid et al., 2010), and *Hymenochirus boettgeri* (Mezzasalma et al., 2015), in which the B chromosomes are at least as large as the chromosomes of the A complement. Rarely, as observed in *Megaelosia massarti* (Rosa et al., 2003), the B chromosome is the largest chromosome of the karyotype.

These supernumerary elements behave independently of the A complement and at a local population level, as indicated by the fact that different populations of the same species may present distinctive morphologies or numbers of B chromosomes as observed among populations of *Boana albopunctata* (Gruber et al., 2007; Ferro et al., 2012 – as *Hypsiboas albopunctatus*) and *Odontophrynus americanus* (Rosset et al., 2006; Lanzone et al., 2008). Distinct chromosome morphotypes may coexist within a given population, and may even be found in the same individual in some cases (Ferro et al., 2016). Individuals may also present different combinations of these elements (Green, 1988; Green et al., 1993; Schmid et al., 2010). Schmid et al. (2010) documented, for example, four types of B chromosome in *Eleutherodactylus gundlachi*, all of which were telocentric, with two large elements similar in size to pairs 6 and 12 of the A complement of this species, while the other two B chromosomes are much smaller than those of the A complement. These authors documented specimens of *E. gundlachi* with up to 10 additional copies of the B chromosomes, but, when present, the large B chromosomes were only found in one copy in cells, whereas different random combinations of the small B chromosomes were found.

Given all of the different potential combinations of B chromosomes, some authors have attempted to define the limit for the number of copies of these supernumerary elements that a cell would be able to support. This type of analysis is complex, given that, in extreme cases, as many as nine or ten B chromosomes may be found in the same genome, as in the cases of *Gastrotheca espelitia* (Schmid et al., 2002b, Schmid, 2012) and *Eleutherodactylus gundlachi* (Schmid et al., 2010), respectively, reaching the unexpected number of 15 in *Leiopelma hochstetteri* (Green, 1988; Green et al., 1993). What number of B chromosomes a eukaryote cell can support and how the chromatin would be rearranged to accommodate this additional genomic content are questions that demand further attention, and these animals provide potentially interesting research models. The fact that segregation is non-Mendelian results in a spectrum of combinations of B chromosomes in a population, ranging from individuals who have no supernumerary elements to those that have a number of copies (for example, see Green, 1988; Green et al., 1993).

Table 3. Summary of the B chromosome systems described in Anura

Family/Species	B chromosomes			Chromosome markers		Reference
	Number*	Morphology**	Type	Heterocromatin	rDNA sites	
Alytidae <i>Discoglossus pictus</i>	2	S	t	n.e.	-	Schmid et al., 1987
Hemiphractidae <i>Gastrotheca espelitia</i>	9	D	m/t	C+	+	Schmid et al., 2002; Schmid et al., 2012
Hylidae <i>Acris crepitans</i>	5	S	sm	n.e.	n.e.	Nur & Nevo, 1969
<i>Bokermannohyla luctuosa</i>	1	S	m	C+ (?)	-	Baldiseria et al., 1993
<i>Dendropsophus nanus</i>	3	S	t	C-	-	Medeiros et al., 2006
<i>Boana albopunctata</i>	2	D	m	q+ptn/ q+p per	-	Cardoso et al., 1986; Gruber et al., 2007; Ferro et al., 2012 (as <i>Hypsiboas albopunctatus</i>)
<i>Smilisca baudinii</i>	2	S	t	n.e.	n.e.	Hernández-Guzmán et al., 2011
<i>Sphoerohyla dorsalis</i>	2	S	t	C+	-	Suaréz et al., 2013
Hylodidae <i>Megaelosia massarti</i>	1	D (2)	m/sm	C+/ q+pter	-	Rosa et al., 2003
Leiopelmatidae <i>Leiopelma hochstetteri</i>	15	D (3)	t	cen	-	Green et al., 1987; Green et al., 1993; Sharbel et al., 1998
Ranidae <i>Amolops liangshanensis</i>	1	S	t	cen	-	Wu & Zhao, 1985
<i>Sanguirana everetti</i>	1	S	t	cen	-	Kuramoto et al., 1989
<i>Rana temporaria</i>	4	S	sm	qint	-	Schmid et al., 1978
Scaphiopodidae <i>Spea hammondi</i>	1	S	t	C+	+	Green & Sessions, 1991
Craugastoridae <i>Craugastor</i> sp. B	2	S	m	C+	-	Schmid et al., 2010
<i>Oreobates discoidalis</i>	1	D (2)	t	n.e.	n.e.	Schmid et al., 2010
<i>Oreobates barituensis</i>	2	D	t/st	cen	+	Ferro et al., 2016
Eleutherodactylidae <i>Eleutherodactylus gundlachi</i>	10	D	t	C+	+	Schmid et al., 2010
Leptodactylidae <i>Physalaemus feioi</i>	1	S	st	qint	+ (qper)	Milani et al., 2010; this work
Odontophrynidae <i>Odontophrynus americanus</i>	5	S/D	st	qint	+ (qper)	Rosset et al., 2006; Lanzone et al., 2008
Pipidae <i>Hymenochirus boettgeri</i>	1	S	m	C+	-	Mezzasalma et al., 2015

* maximum number of B chromosomes per cell. ** (S) B chromosomes undifferentiated; (D) B chromosomes vary in size and morphology; ¹ Polyploid species with inter-individual and inter-population variation;

Presence of NORs is indicated by a “+”, while the absence of NORs is indicated by a “-”. q = long arm; p = short arm; C+ = completely heterochromatic; C- = absence of heterochromatin blocks; cen = centromere; int = interstitial position; per = pericentromeric position; ter = terminal position in the chromosome; n.e = not examine.

The meiotic behavior of these supernumerary chromosomes is the key to their irregular segregation. By definition, B chromosomes do not pair with the chromosomes from complement A during meiotic prophase I, and the observation of univalent chromosomes in a species is one of the criteria used to confirm the presence of supernumerary elements in the genome. When more than one copy is present in the nucleus, these elements may pair with one another, presenting a bivalent (example in Ferro et al., 2016) or multivalent (example in Nur & Nevo, 1969) configuration. In other cases, however, these copies may remain univalent during meiosis I, as observed in male *Dendropsophus nanus*, which have three copies of euchromatic B chromosomes (Medeiros et al., 2006). The majority of B chromosomes, despite being unstable during meiosis, are stable during mitosis (example in Gruber et al., 2007; Ferro et al., 2016), although there are exceptions (as B_d chromosome described by Ferro et al., 2016).

In practically all cases of B chromosomes in anurans, C-banding reveals elements either partially (see, for example, Schmid, 1978; Gruber et al., 2007; Milani et al., 2010; Ferro et al., 2012) or completely heterochromatic (Green & Sessions, 1991; Schmid et al., 2002b; Schmid et al., 2010; Suaréz et al., 2013; Mezzasalma et al., 2015), a characteristic also observed in the B chromosomes of other vertebrate groups. This would appear to be related to the evolutionary dynamics of these chromosomes because as dispensable elements in the genome, they are under weak selection pressure, which may lead to the accumulation of repetitive sequences (Camacho, 2000). In some organisms, the analysis of the heterochromatin content of B chromosomes has revealed the accumulation of transposons and retrotransposons, and the presence of different families of repetitive DNA. In anurans, unfortunately, the description of the heterochromatin content of the B chromosomes has been limited to conventional banding methods, while their nucleotide content has been analyzed only using base-specific fluorochromes (Gruber et al., 2007; Milani et al., 2010; Ferro et al., 2012; Mezzasalma et al., 2015; Ferro et al., 2016). However, small amounts of heterochromatin were already reported for supernumerary elements of *Dendropsophus nanus* (Medeiros et al., 2006) and in *Megaelasia massarti* (Rosa et al., 2003), for example, which makes these species interesting models for the study of the DNA content of these elements.

The fact that the B chromosomes accumulate large heterochromatic segments does not necessarily mean that they are transcriptionally inert. With the ongoing advances in genomic methods, this has been shown empirically in the B chromosomes of plants, grasshoppers, and fish. The B chromosomes of some anuran species contain rDNA sites which are transcriptionally active in the genome and can be detected through the silver Ag-NOR method (Green & Sessions, 1991; Milani et al., 2010; Schmid et al., 2010; Schmid et al., 2012; Ferro et al., 2016). Two copies of rDNA were detected by FISH in one of the B chromosome types of *Oreobates barituensis* (Ferro et al., 2016), while only one copy was detected by the Ag-NOR method, providing direct evidence of transcriptional activity prior to the last cell cycle.

The molecular composition of the B chromosomes is the principal source of evidence for the formulation of hypotheses on their origin. Three principal hypotheses have been proposed to account for the appearance of these chromosomes: (i) intraspecific origin, (ii) byproducts of the process of differentiation of the sex chromosomes, and (iii) interspecific origin, as vestiges of hybridization events. In the case of the intraspecific origin, the principal assumption is the sharing of DNA sequences with the A chromosomes genome. As mentioned above, the fact that B chromosomes do not recombine with the A genome means that these elements are inherited clonally, which means that we would expect young B chromosomes of intraspecific origin to present greater similarities with the A chromosomes. This hypothesis has been used

to explain the origin of B chromosomes of a number of different groups of organisms. The supernumerary chromosome of *Sphaenohynchus dorsiae*, for example, contains a large block of telomeric repeats that occupy almost the whole of the B chromosome, which Suárez et al. (2013) found to be highly similar to the segment present in the homologs of pair 2 of *Sphaenohynchus lacteus*, which is absent from homeolog 2 of *S. dorsiae*. This evidence led the authors to suggest that the B chromosome of this species may have originated from this region of chromosome 2.

In the B chromosome system of *L. hochstetteri*, however, the comparison of the DNA sequences of the B chromosomes with those of the W chromosome revealed a high degree of similarity, which is consistent with the intraspecific origin of B elements from the diversification of the W sex chromosome of this species (Sharbel et al., 1998). Similar evidence of sex-chromosome derived B chromosomes has been found in *Characidium gomesi*, another species with a ZZ/ZW system (Pansonato-Alves et al., 2014).

The third hypothesis postulates that the B chromosomes are remnants of genomic hybridization events that have persisted in the genome. This hypothesis is normally the most difficult to validate, because it requires the confirmation of two fundamental aspects: (i) the sharing of specific DNA sequences between the B chromosome of one species and the A chromosomes from putative other species interbreed, and (ii) the occurrence of the two species within the same zone of sympatry. While a number of studies have attempted to prove the role of this mechanism in the origin of B chromosomes in anurans, no conclusive evidence of the process has been presented in any case.

While the origin of B chromosomes has been identified conclusively in some cases, most attempts to assess their homologies with the A chromosomes have failed, which may be a consequence of the independent evolution of B and A complements. In *Physalaemus feioi*, for example, the supernumerary chromosome has a large heterochromatin block on its long arm, which was exclusively detected by a probe derived from microdissected B chromosomes (Figure 5). No hybridization signal of this specific B probe was detected in the A complement, suggesting that some repetitive sequences are accumulated or exclusively present in this supernumerary chromosome. A similar situation has also been reported to the B chromosome of a Brazilian population of *Boana albopunctata*, which was entirely detected by a probe that did not hybridize to the A complement of this species (Gruber et al., 2014). In this case, the B probe detected small scattered regions on all chromosomes of another species, *Boana raniceps* (Gruber et al., 2014), which may correspond to sites of repetitive DNA sequences. Given the dynamic evolution of repetitive DNA (for review, see Feschotte & Pritham, 2007 and Plohl et al., 2008), the sharing of repetitive sequences should be analyzed with caution when assessing homology hypotheses. For this reason, the results observed by Gruber and colleagues should not be interpreted as a direct evidence of interspecific origin of the B chromosome of *B. albopunctata* (see discussion in Gruber et al., 2014).

Another interesting finding on the B chromosomes of anurans is the presence of these elements in the genomes of natural polyploid species. While B chromosomes are relatively common in polyploid plants (for review, see Palestis et al., 2004), up until now, *Odontophrynus americanus*, a South American frog, is the only polyploid vertebrate known to have B chromosomes (Rosset et al., 2006; Lanzone et al., 2008). Although these elements were rare in the *O. americanus* populations analyzed by Rosset et al. (2006) and Lanzone et al. (2008), this organism provides a potentially interesting model for the analysis of the origin and maintenance of these elements in the genome.

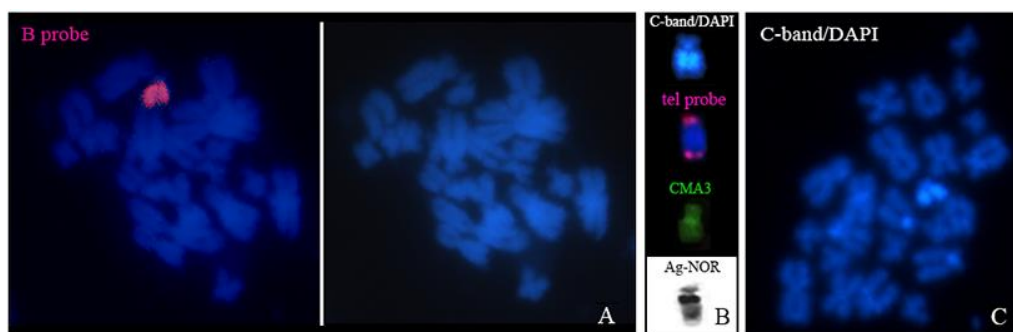


Figure 5. B chromosome of *Physalaemus feioi*. **A.** *In situ* hybridization of a probe derived from microdissected B chromosomes. On the right, DAPI image; on the left, merged image of DAPI and B probe. Note that no hybridization signal is seen in the A complement, whereas the long arm of the B chromosome is detected by the probe. **B.** B chromosome submitted to C-banding and DAPI staining (C-band+DAPI), *in situ* hybridized with telomeric probe (tel probe), stained with chromomycin A3 (CMA3), and silver stained (Ag-NOR). **C.** C-banded metaphase stained with DAPI. The chromosome preparations in **A** and **B** were obtained from the female specimen SMRP 371.28, whereas the metaphase shown in **C** was obtained from the male ZUEC 16249 (one of the specimens also analyzed by Milani et al., 2010). The B chromosome found in the specimen SMRP 371.28 differs from those described previously (Milani et al., 2010) by the presence of a pericentromeric NOR, although the specimens were all collected from Viçosa-MG, Brazil. ZUEC: Museum of Zoology, University of Campinas, Brazil. SMRP: Collection “Shirlei Maria Recco Pimentel”, University of Campinas, Brazil.

In conclusion, despite the general properties recognized among B chromosomes of the different anuran species, the evolutionary profiles of these enigmatic chromosomes are far from being completely understood.

REFERENCES

- Abramyan, J., Ezaz T., Graves, J. A. M. & Koopman, P. (2009). Z and W sex chromosomes in the cane toad (*Bufo marinus*). *Chromosome Res.*, 17, 1015–1024.
- Acosta, M. J., Marchal, J. A., Martínez, S. & Sánchez, A. (2007) Characterization of the satellite DNA Msat-160 from the species *Chionomys nivalis* (Rodentia, Arvicolinae). *Genetica*, 130, 43
- Aguiar-Jr, O., Bacci, Jr. M., Lima, A. P., Rossa-Feres, D. C., Haddad, C. F. B. & Recco-Pimentel, S. M. (2007). Phylogenetic relationships of *Pseudis* and *Lysapsus* (Anura, Hylidae, Hylinae) inferred from mitochondrial and nuclear gene sequences. *Cladistics*, 23, 455–463.
- Altinordu, F., Peruzzi, L., Yu, Y. & He, X. (2016). A tool for the analysis of chromosomes: KaryoType. *Taxon*, 65, 586–592.
- Amaro, R. C., Rodrigues, M. T., Yonenaga-Yassuda, Y. & Carnaval, A. C. (2012). Demographic processes in the montane Atlantic rainforest: Molecular and cytogenetic evidence from the endemic frog *Proceratophrys boiei*. *Mol. Phylogenet. Evol.*, 62, 880–888.
- Amaro-Ghilardi, R. C., Silva, M. J. J., Rodrigues, M. T. & Yonenaga-Yassuda, Y. (2007) Chromosomal studies in four species of genus *Chaunus* (Bufonidae, Anura): localization

- of telomeric and ribosomal sequences after fluorescence *in situ* hybridization (FISH). *Genetica*, 134, 159-168.
- Ananias, F., Modesto, A. D. S., Mendes, S. C. & Napoli, M. F. (2007). Unusual primitive heteromorphic ZZ/ZW sex chromosomes in *Proceratophrys boiei* (Anura, Cycloramphidae, Alsodinae), with description of C-Band interpopulational polymorphism. *Hereditas*, 144, 206-212.
- Andrade, F. S. & Carvalho, T. R. (2013). A new species of *Pseudopaludicola* Miranda-Ribeiro (Leiuperinae: Leptodactylidae: Anura) from the Cerrado of southeastern Brazil. *Zootaxa*, 3608, 389-397.
- Andrade, F. S., Magalhães, F. M., Nunes-de-Almeida, C. H. L., Veiga-Menoncello, A. C. P., Santana, D. J., Garda, A. A., Loebmann, D., Recco-Pimentel, S. M., Giaretta, A. A. & Toledo, L. F. (2016). A new species of long-legged *Pseudopaludicola* from northeastern Brazil (Anura, Leptodactylidae, Leiuperinae). *Salamandra*, 52, 107-124.
- Baldissera, F. A., Deoliveira, P. S. L. & Kasahara, S. (1993). Cytogenetics of 4 Brazilian *Hyla* species (Amphibia- Anura) and description of a case with a supernumerary chromosome. *Rev. Bras. Genet.*, 16, 335-345.
- Bergero, R. & Charlesworth, D. (2009). The evolution of restricted recombination in sex chromosomes. *Trends Ecol. Evol.*, 24, 94-102.
- Berneck, B. V. M., Haddad, C. F. B., Lyra, M. K., Cruz, C. A. G. & Faivovich, J. (2016). The green clade grows: A phylogenetic analysis of *Aplastodiscus* (Anura, Hylidae). *Mol. Phylogen. Evol.*, 97, 213-223.
- Berset-Brändli, L., Jaquier, J. & Perrin, N. (2007). Recombination is suppressed and variability reduced in a nascent Y chromosome. *J. Evol. Biol.*, 20, 1182-1188.
- Berset-Brändli, L., Jaquier, J., Broquet, T., Ulrich, Y. & Perrin, N. (2008). Extreme heterochiasmy and nascent sex chromosomes in European tree frogs. *Proc. R. Soc. B*, 275, 1577-1585.
- Berta, P., Hawkins, J. R., Sinclair, A. H., Taylor, A., Griffiths, B. L., Goodfellow, P. N. & Fellous, M. (1990). Genetic evidence equating SRY and the testis-determining factor. *Nature*, 348, 448-450.
- Bewick, A. J., Anderson, D. W. & Evans, B. J. (2011). Evolution of the closely related, sex-related genes DM-W and DMRT1 in African clawed frogs (*Xenopus*). *Evolution*, 65, 698-712.
- Bewick, A. J., Chain, F. J. J., Zimmerman, L. B., Sesay, A., Gilchrist, M. J., Owens, N. D. L., Seifertova, E., Krylov, V., Macha, J., Tlapakova, T., Kubickova, S., Cernohorska, H., Zarsky, V. & Evans, B. J. (2013). A large pseudoautosomal region on the sex chromosomes of the frog *Silurana tropicalis*. *Genome Biol. Evol.*, 5, 1087-1098.
- Blotto, B. L., Nuñez, J. J., Basso, N. G., Úbeda, C. A., Wheeler, W. C. & Faivovich, J. (2013). Phylogenetic relationships of a Patagonian frog radiation, the *Alsodes* + *Eupsophus* clade (Anura: Alsodidae), with comments on the supposed para-phyly of *Eupsophus*. *Cladistics*, 29, 113-131.
- Borgart, J. P. (1991). The influence of life history on karyotypic evolution in frogs. In *Amphibian cytogenetics and evolution* (ed. D. M. Green & S.K. Sessions), pp. 233-257. Academic Press, San Diego, USA.
- Brelsford, A., Dufresnes, C. & Perrin, N. (2016). High-density sex-specific linkage maps of a European tree frog (*Hyla arborea*) identify the sex chromosome without information on offspring sex. *Heredity*, 116, 177-181.

- Busin, C. S., Viciprova, G. & Recco-Pimentel, S. M. (2011). Chromosomal rearrangements as the source of variation in the number of chromosomes in *Pseudis* (Amphibia, Anura). *Genetica*, 110, 131-141.
- Busin, C. S., Andrade, G. V., Bertoldo, J., Del-Grande, M. L., Uetanabaro, M. & Recco-Pimentel, S. M. (2008). Cytogenetic analysis of four species of *Pseudis* (Anura, Hylidae), with the description of ZZ/ZW sex chromosomes in *P. tocantins*. *Genetica*, 133, 119-127.
- Camacho, J. P. M., Sharbel, T. F. & Beukeboom, L. W. (2000) B chromosome evolution. *Philos. T. R. Soc. B*, 355, 163-178.
- Caminer, M. A., Mila, B., Jansen, M., Fouquet, A., Venegas, P. J., Chávez, G., Loughheed, S. C. & Ron, S. R. (2017). Systematics of the *Dendropsophus leucophyllatus* species complex (Anura: Hylidae): Cryptic diversity and the description of two new species. *PLoS ONE*, 12(3), e0171785.
- Cardozo, D. E., Leme, D. M., Bortoleto, J. F., Catroli, G. F., Baldo, D., Faivovich, J., Kolenc, F., Silva, A. P. Z., Borteiro, C., Haddad, C. F. B. & Kasahara, S. (2011). Karyotypic data on 28 species of *Scinax* (Amphibia: Anura: Hylidae): diversity and informative variation. *Copeia*, 2, 251-263.
- Cazaux, B., Catalan, J., Justy, F., Escudé, C., Desmarais, E. & Britton-Davidian, J. (2013). Evolution of the structure and composition of house mouse satellite DNA sequences in the subgenus *Mus* (Rodentia: Muridea): a cytogenomic approach. *Chromosoma*, 122, 209-220.
- Charlesworth, D., Charlesworth, B. & Marais, G. (2005). Steps in the evolution of heteromorphic sex chromosomes. *Heredity*, 95, 118-128.
- Chen, J. M., Zhou, W. W., Poyarkov, Jr. N. A., Stuart, B. L., Brown, R. M., Lathrop, A., Wang, Y. Y., Yuan, Z. Y., Jiang, K., Hou, M., Chen, H. M., Suwannapoom, C., Nguyen, S. N., Duong, T. V., Papenfuss, T. J., Murphy, R. W., Zhang, Y. P. & Che, J. (2017). A novel multilocus phylogenetic estimation reveals unrecognized diversity in Asian horned toads, genus *Megophrys* sensu lato (Anura: Megophryidae). *Mol. Phylogenet. Evol.*, 106, 28-43.
- Cortez, D., Marin, R., Toledo-Flores, D., Froidevaux, L., Liechti, A., Waters, P. D., et al. (2014). Origins and functional evolution of Y chromosomes across mammals. *Nature*, 508, 488-493.
- Cuevas, C. C. & Formas, J. R. (1996). Heteromorphic sex chromosomes in *Eupsophus insularis* (Amphibia: Anura: Leptodactylidae). *Chromosome Res.*, 4, 467-70.
- Cuevas, C. C. (2008). A new species of the genus *Alsodes* (Anura: Neobatrachia) from the *Nothofagus* forest, Coastal Range, Southern Chile, identified by its karyotype. *Zootaxa*, 1771, 43-53.
- Cunningham, M. J. & Cherry, M. I. (2004). Molecular systematics of African 20-chromosome toads (Anura: Bufonidae). *Mol. Phylogenet. Evol.*, 32, 671-685.
- De Sá, R. O., Grant, T., Camargo, A., Heyer, W. R., Ponsa, M. L. & Stanley, E. (2014). Systematics of the Neotropical genus *Leptodactylus* Fitzinger, 1826 (Anura: Leptodactylidae): phylogeny, the relevance of non-molecular evidence, and species accounts. *S. Am. J. Herpetol.*, 9, S1-S100.
- Diniz, D., Laudicina, A., Cioffi, M. B. & Bertollo, L. A. C. (2008). Microdissection and whole chromosome painting. Improving sex chromosome analysis in *Triportheus* (Teleostei, Characiformes). *Cytogenet. Genome Res.*, 122, 163-168.
- Dover, G. (1982). Molecular drive: a cohesive mode of species evolution. *Nature*, 299, 111-117.

- Dover, G. A. (1986). Molecular drive in a multigene families: how biological novelties arise, spread and are assimilates. *Trends in Genet.*, 2, 159-165.
- Duarte, T. C., Veiga-Menoncello, A. C. P., Lima, J. F. R., Strüssmann, C., Del-Grande, M. L., Giaretta, A. A., Pereira, E. G., Rossa-Feres, D. C. & Recco-Pimentel, S. M. (2010). Chromosome analysis in *Pseudopaludicola* (Anura, Leiuperidae), with description of sex chromosomes XX/XY in *P. saltica*. *Hereditas*, 147, 43-52.
- Duellman, W. E., Marion, A. B. & Hedges, S. B. (2016). Phylogenetics, classification, and biogeography of the treefrogs (Amphibia: Anura: Arboranae). *Zootaxa.*, 4104, 001-109.
- Dufresnes, C., Borzée, A., Horn, A., Stöck, M., Ostini, M., Sermier, R., Wassef, J., Litvinchuck, S. N., Kosch, T. A., Waldman, B., Jang, Y., Brelsford, A. & Perrin, N. (2015). Sex-chromosome homomorphy in Palearctic tree frogs results from both turnovers and X - Y recombination. *Mol. Biol. Evol.*, 32, 2328-2337.
- Elder, J. F. & Turner, B. J. (1995). Concerted evolution of repetitive DNA sequences in eukaryotes. *Q. Rev. Biol.*, 70, 297-320.
- Elshire, R. J., Glaubitz, J. C., Sun, Q., Poland, J. A., Kawamoto, K., Buckler, E. S. & Mitchell, S. E. (2011). A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. *PLoS ONE*, 6(5), e19379.
- Evans, B. J., Kelley, D. B., Tinsley, R. C., Melnick, D. J. & Cannatella, D. C. (2004). A mitochondrial DNA phylogeny of clawed frogs: phylogeography on sub-Saharan Africa and implications for polyploid evolution. *Mol. Phylogenet. Evol.*, 33, 197-213.
- Evans, B. J. (2007). Ancestry influences the fate of duplicated genes millions of years after polyploidization of clawed frogs (*Xenopus*). *Genetics*, 176, 1119-1130.
- Evans, B. J. (2008). Genome evolution and speciation genetics of clawed frogs (*Xenopus* and *Silurana*). *Front. Biosci.*, 13, 4687-4706.
- Evans, B. J., Pyron, R. A. & Wiens, J. J. (2012) Polyploidization and Sex Chromosome Evolution in Amphibians. In: P. S. Soltis & D. E. Soltis (Eds.), *Polyploidy and Genome Evolution*. Springer-Verlag, Berlin, Heidelberg, pp. 385-410.
- Evans, B. J., Carter, T. F., Greenbaum, E., Gvoždík, V., Kelley, D. B., McLaughlin, P. J., Pauwels, O. S. G., Portik, D. M., Stanley, E. L., Tinsley, R. C., Tobias, M. L. & Blackburn, D. C. (2015). Genetics, morphology, advertisement calls, and historical records distinguish six new polyploid species of African clawed frog (*Xenopus*, Pipidae) from West and Central Africa. *PLoS ONE*, 10(12), e0142823.
- Faivovich, J. (2002). A cladistic analysis of *Scinax* (Anura: Hylidae). *Cladistics*, 18, 367-393.
- Faivovich, J., Haddad, C. F. B., Garcia, P. C. A., Frost, D. R., Campbell, J. R. & Wheeler, W. C. (2005). Systematic review of the frog family Hylidae, with special reference to Hylinae: phylogenetic analysis and taxonomic revision. *B. Am. Mus. Nat. Hist.*, 294, 1-240.
- Faivovich, J., Ferraro, D. P., Basso, N. G., Haddad, C. F. B., Rodrigues, M. T.; Wheeler, W. C. & Lavilla, E. O. (2012). A phylogenetic analysis of *Pleurodema* (Anura: Leptodactylidae: Leiuperinae) based on mitochondrial and nuclear gene sequences, with comments on the evolution of anuran foam nests. *Cladistics*, 28, 460-482.
- Fávero, E. R., Veiga-Menoncello, A. C. P., Rossa-Feres, D. C., Strüssmann, C., Giaretta, A. A., Andrade, G. V., Colombo, P. & Recco-Pimentel, S. M. (2011). Intrageneric karyotypic variation in *Pseudopaludicola* (Anura: Leiuperidae) and its taxonomic relatedness. *Zool. Stud.*, 50, 826-836.
- Feng, Y. J., Blackburn, D. C., Liang, D., Hillis, D. M., Wake, D. B., Cannatella, D. C. & Zhang, P. (2017). Phylogenomics reveals rapid, simultaneous diversification of three major clades

- of Gondwanan frogs at the Cretaceous-Paleogene boundary. *P. Natl. Acad. Sci. USA*, **114**, E5864-E5870.
- Ferro, M. J., Taffarel, A., Cardozo, D., Grosso, J., Puig, M. O., Suárez, P., Akmentins, M. S. & Baldo, D. (2016). Cytogenetic characterization and B chromosome diversity in direct-developing frogs of the genus *Oreobates* (Brachycephaloidea, Craugastoridae). *Comp. Cytogenet.*, **10**, 141-156.
- Feschotte, C. & Pritham, E. J. (2007). DNA transposons and the evolution of eukaryotic genomes. *Annu. Rev. Genet.*, **41**, 331-368.
- Fouquet, A., Gilles, A., Vences, M., Marty, C., Blanc, M. & Gemmell, N. J. (2007). Underestimation of species richness in Neotropical frogs revealed by rDNA analyses. *PLoS ONE*, **2**(10), e1109.
- Frost, D. R. *Amphibian species of the world: an online reference*. Version 6.0 (Date of access: 15/nov/2017). Available from: URL: [http:// research.amnh.org/herpetology/ amphibia/ index.html](http://research.amnh.org/herpetology/amphibia/index.html).
- Funk, W. C., Caminer, M. & Ron, S. R. (2012). High levels of cryptic species diversity uncovered in Amazonian frogs. *Proc. R. Soc. B.*, **279**, 1806-1814.
- Furman, B. L. & Evans, B. J. (2016). Sequential turnovers of sex chromosomes in African clawed frogs (*Xenopus*) suggest some genomic regions are good at sex determination. *G3-Genes Genom. Genet.*, **6**, 3625-3633.
- Gatto, K. P., Busin, C. S. & Lourenço, L. B. (2016). Unraveling the sex chromosome heteromorphism of the paradoxical frog *Pseudis tocantins*. *PLoS ONE*, **11**, e0156176.
- Giam, X., Scheffers, B. R., Sodhi, N. S., Wilcove, D. S., Ceballos, G. & Ehrlich, P. R. (2013). Reservoirs of richness: least disturbed tropical forests are centers of undescribed species diversity. *Proc. R. Soc. B*, **279**, 67-76.
- Goodfellow, P. N. & Lovell-Badge, R. (1993). SRY and sex determination in mammals. *Annu. Rev. Genet.*, **27**, 71-92.
- Grant, T., Frost, D. R., Caldwell, J. P., Galiardo, R., Haddad, C. F. B., Kok, P. J. R., Means, D. B., Noonan, B. P., Schargel, W. E. & Wheeler, W. W. (2006). Phylogenetic systematic of dart-poison frogs and their relatives (Amphibia, Athesphatanura, Dendrobatidae). *B. Am. Mus. Nat. Hist.*, **299**, 1-262.
- Graves, J. A. M. (2006). Sex chromosome specialization and degeneration in mammals. *Cell*, **124**, 901-914.
- Graves, J. A. M. (2008). Weird animal genomes and the evolution of vertebrate sex and sex chromosomes. *Annu. Rev. Genet.*, **42**, 565-586.
- Graves, J. A. M. & Peichel, C. L. (2010). Are homologies in vertebrate sex determination due to shared ancestry or to limited options? *Genome Biol.*, **11**, 205.
- Green, D. M., Kezer, J. & Nussbaum, R. A. (1987). Supernumerary chromosome variation and heterochromatin distribution in the endemic New Zealand frog, *Leiopelma hochstetteri*. *Chromosoma*, **95**, 339-344.
- Green, D. M. (1988). Cytogenetics of the endemic New Zealand frog, *Leiopelma hochstetteri*: extraordinary supernumerary chromosome variation and a unique sex-chromosome system. *Chromosoma*, **97**, 55-70.
- Green, D. M. & Sessions, S. K. (1991). (eds): *Amphibian Cytogenetics and Evolution*. Academic Press, San Diego.

- Green, D. M., Zeyl, C. W. & Sharbel, T. F. (1993). The evolution of hypervariable sex and supernumerary chromosomes in the relict New Zealand frog, *Leiopelma hochstetteri*. *J. Evol. Biol.*, 6, 417-441.
- Gruber, S. L., Haddad, C. F. B. & Kasahara, S. (2007). Chromosome banding in three species of *Hypsiboas* (Hylidae, Hyliinae), with special reference to a new case of a B-chromosome in anuran frogs and to the reduction of the diploid number of $2n = 24$ to $2n = 22$ in the genus. *Genetica*, 130, 281-291.
- Gruber, S. L., Zina, J., Narimatsu, H., Haddad, C. F. B. & Kasahara, S. (2012a). Comparative karyotype analysis and chromosome evolution in the genus *Aplastodiscus* (Cophomantini, Hyliinae, Hylidae). *BMC Genet.*, 13, 28.
- Gruber, S., Haddad, C. F. B. & Kasahara, S. (2012b). Karyotype analysis of seven species of the tribe Lophiohylini (Hyliinae, Hylidae, Anura), with conventional and molecular cytogenetic techniques. *Comp. Cytogenet.*, 6, 409-423.
- Gruber, S. L., Diniz, D., Sobrinho-Scudeler, P. E., Foresti, F., Haddad, C. F. B. & Kasahara, S. (2014). Possible interspecific origin of the B chromosome of *Hypsiboas albopunctatus* (Spix, 1824) (Anura, Hylidae), revealed by microdissection, chromosome painting, and reverse hybridization. *Comp. Cytogenet.*, 8, 185-197.
- Guarnizo, C. E., Escallón, C., Cannatella, D. C. & Amésquita, A. (2012). Congruence between acoustic traits and genealogical history reveals a new species of *Dendropsophus* (Anura: Hylidae) in the high Andes of Colombia. *Herpetologica*, 68, 523-540.
- Guerrero, R. F., Kirkpatrick, M. & Perrin, N. (2012). Cryptic recombination in the ever-young sex chromosomes of Hylid frogs. *J. Evol. Biol.*, 25, 1947-1954.
- Hall, A. B., Qi, Y., Timoshevskiy, V., Sharakhova, M. V., Sharakhov, I. V. & Tu, Z. (2013). Six novel Y chromosome genes in *Anopheles* mosquitoes discovered by independently sequencing males and females. *BMC Genomics*, 14, 273.
- Hassanane, M. S., Chaudhary, R. & Chowdhary, B. P. (1998). Microdissected bovine X chromosome segment delineates homoeologous chromosomal regions in sheep, goat and buffalo. *Chromosome Res.*, 6, 213-217.
- Hellsten, U., Harland, R. M., Gilchrist, M. J., Hendrix, D., Jurka, J., Kapitonov, V., Ovcharenko, I., Putnam, N. H., Shu, S., Taher, L., Blitz, I. L., Blumberg, B., Dichmann, D. S., Dubchak, I., Amaya, E., Detter, J. C., Fletcher, R., Gerhard, D. S., Goodstein, D., Graves, T., Grigoriev, I. V., Grimwood, J., Kawashima, T., Lindquist, E., Lucas, S. M., Mead, P. E., Mitros, T., Ogino, H., Ohta, Y., Poliakov, A. V., Pollet, N., Robert, J., Salamov, A., Sater, A. K., Schmutz, J., Terry, A., Vize, P. D., Warren, W. C., Wells, D., Wills, A., Wilson, R. K., Zimmerman, L. B., Zorn, A. M., Grainer, R., Grammer, T., Khokha, M. K., Richardson, P. M. & Rokhsar, D. S. (2010). The genome of the Western clawed frog *Xenopus tropicalis*. *Science*, 328, 633-636.
- Henikoff, S., Ahmad, K. & Malik, H. S. (2001). The centromere paradox: stable inheritance with rapidly evolving DNA. *Science*, 293, 1098-1102.
- Henning, F., Moysés, C. B., Calcagnotto, D., Meyer, A. & de Almeida-Toledo, L. F. (2011). Independent fusions and recente origins of sex chromosomes in the evolution and diversification of glass knife fishes (*Eigenmannia*). *Heredity*, 106, 391-400.
- Hernández-Guzmán, J., Arias-Rodríguez, L. & Indy, J. R. (2011). Los cromosomas meióticos de la rana arborícola *Smilisca baudinii* (Anura: Hylidae). *Rev. Biol. Trop.*, 59, 355-362. [Meiotic chromosomes of the tree frog *Smilisca baudinii* (Anura: Hylidae)].

- Hillis, D. M. & Green, D. M. (1990). Evolutionary phylogenetic changes of heterogametic history of amphibians. *J. Evol. Biol.*, 3, 49-64.
- Holloway, A. K., Cannatella, D. C., Gerhardt, H. C. & Hillis, D. M. (2006). Polyploids with different origins and ancestors form a single sexual polyploid species. *Am. Nat.*, 167, E88-E101.
- Iturra, P. & Veloso, A. (1989) Further evidence for early sex chromosome differentiation of anuran species. *Genetica*, 78, 25-31.
- Karamysheva, T. V., Andreenkova, O. V., Bochkarev, M. N., Borissov, Y. M., Bogdanchikova, N, Borodin, P. M. & Rubtsov, N. B. (2002). B chromosomes of Korean field mouse *Apodemus peninsulae* (Rodentia, Murinae) analysed by microdissection and FISH. *Cytogenet. Genome Res.*, 96, 154-160.
- Kasahara, S., Silva, A. P. & Haddad, C. F. B. (1996). Chromosome banding in three species of Brazilian toads (Amphibia - Bufonidae). *Braz J Gen*, 19, 237-242.
- Kawagoshi, T., Nishida, C. & Matsuda, Y. (2012). The origin and differentiation process of X and Y chromosomes of the black marsh turtle (*Siebenrockiella crassicollis*, Geoemydidae, Testudines). *Chromosome Res.*, 20, 95-110.
- Kirov, I., Khrustaleva, L., Van Laere, K., Soloviev, A., Meeus, S., Romanov, D. & Fesenko, I. (2017). DRAWID: user-friendly java software for chromosome measurements and ideogram drawing. *Comp. Cytogenet.*, 11, 747-757.
- Kohler, J., Vieites, D. R., Bonett, R. M., Garcia, F. H., Glaw, F., Steinke, D. & Vences, M. (2005). New amphibians and global conservation: A boost in species discoveries in a highly endangered vertebrate group. *Bioscience*, 55, 693-696.
- Krylov, V., Kubickova, S., Rubes, J., Macha, J., Tlapakova, T., Seifertova, E. & Sebkova N. (2010). Preparation of *Xenopus tropicalis* whole chromosome painting probes using laser microdissection and reconstruction of *X. laevis* tetraploid karyotype by Zoo-FISH. *Chromosomes Res.*, 1, 431-439.
- Kuramoto, M. (1989). Karyological studies on some Philippine frogs. In: M. Matsui, T. Kikida, R. C. Goris (Eds.), *Current Herpetology in East Asia.*, (pp 115–121) Kyoto, Herpetological Society of Japan.
- Lanzone, C., Baldo, D. & Rosset, S. D. (2008). Meiotic differentiation in two allopatric population groups of the tetraploid frog *Odontophrynus americanus* from Argentina. *Herpetol. J.*, 18, 213-222.
- López-Flores, I. & Garrido-Ramos, M. A. (2012). The repetitive DNA content of eukaryotic genomes. In: Garrido-Ramos, MA. *Repetitive DNA*. Basel, 1-28.
- Lourenço, L. B., Bacci, Jr. M., Martins, V. G., Recco-Pimentel, S. M. & Haddad, C. F. B. (2008). Molecular phylogeny and karyotype differentiation in *Paratelmatobius* and *Scythrophrys* (Anura, Leptodactylidae). *Genetica*, 132, 255-266.
- Lourenço, L. B., Targueta, C. P., Baldo, D., Nascimento, J., Garcia, P. C. A., Andrade, G. V., Haddad, C. F. B. & Recco-Pimentel, S. M. (2015). Phylogeny of frogs from the genus *Physalaemus* (Anura, Leptodactylidae) inferred from mitochondrial and nuclear gene sequences. *Mol. Phylogenet. Evol.*, 92, 204-216.
- Lyra, M. L., Haddad, C. F. & Azeredo-Espin, A. M. L. (2016). Meeting the challenge of DNA barcoding Neotropical amphibians: Polymerase chain reaction optimization and new COI primers. *Mol. Ecol. Resour.*, 17, 966-980.

- Mahony, M. J. (1991). Heteromorphic sex chromosomes in the Australian frog *Crinia bilineata* (Anura, Myobatrachidae). *Genome*, 34(3), 334-337.
- Marchal, J. A., Acosta, M. J., Nietzel, H., Sperling, K., Bullejos, M., Díaz-de-la-Guardia, R. & Sánchez, A. (2004). X chromosome painting in *Microtus*: Origin and evolution of the giant sex chromosomes. *Chromosome Res.*, 12, 767-776.
- Matsuda, M., Shinomiya, A., Kinoshita, M., Suzuki, A., Kobayashi, T., Paul-Prasanth, B., Lau, E., Hamaguchi, S., Sakaizumi, M. & Nagahama, Y. (2007). DMY gene induces male development in genetically female (XX) medaka fish. *P. Natl. Acad. Sci. USA*, 104, 3865-3870.
- Matsubara, K., Uno, Y., Srikulnath, K., Seki, R., Nishida, C. & Matsuda, Y. (2015). Molecular cloning and characterization of satellite DNA sequences from constitutive heterochromatin of the habu snake (*Potobothrops flavoviridis*, Viperidae) and the Burmese python (*Python bivittatus*, Pythonidae). *Chromosoma*, 124, 529-539.
- Mawaribuchi, S., Takahashi, S., Wada, M., Uno, Y., Matsuda, Y., Kondo, M., Fukui, A., Takamatsu, N., Taira, M. & Ito, M. (2017). Sex chromosome differentiation and the W- and Z-specific loci in *Xenopus laevis*. *Dev. Biol.*, 426, 393-400.
- Medeiros, L. R., Rossa-Feres, D. C. & Recco-Pimentel, S. M. (2003). Chromosomal differentiation of *Hyla nana* and *Hyla sanborni* (Anura, Hylidae) with a description of NOR polymorphism in *H. nana*. *J. Hered.*, 94, 149-154.
- Medeiros, L. R., Rossa-Feres, D. C., Jim, J. & Recco-Pimentel, S. M. (2006). B-chromosomes in two Brazilian populations of *Dendropsophus nanus* (Anura, Hylidae). *Genet. Mol. Biol.*, 29, 257-262.
- Medeiros, L. R., Lourenço, L. B., Rossa-Feres, D. C., Lima, A. P., Andrade, G. V., Giaretta A., Egito, G. T. B. T. & Recco-Pimentel, S. M. (2013). Comparative cytogenetic analysis of some species of the *Dendropsophus microcephalus* group (Anura, Hylidae) in the light of phylogenetic inferences. *BMC Genetics*, 14, 59.
- Mestrovic, N., Plohl, M., Mravinac, B. & Ugarkovic, D. (1998). Evolution of satellite DNAs from the genus *Palorus*— experimental evidence for the ‘library’ hypothesis. *Mol. Biol. Evol.*, 15, 1062-1068.
- Mezzasalma, M., Gaetano, F. G., Petraccioli, O. A. & Guarino, F. M. (2005). Chromosome analyses of *Pseudhymenochirus merlini* and *Hymenochirus boettgeri* provide new insights into the chromosome evolution in the anuran family Pipidae. *Zool. Anz.*, 258, 47-53.
- Milani, M., Cassini, C. S., Recco-Pimentel, S. M. & Lourenço, L. B. (2011). Karyotypic data detect interpopulational variation in *Physalaemus olfersii* and the first case of supernumerary chromosome in the genus. *Animal Biol. J.*, 2, 21-28.7
- Mirzaghaderi, G. & Marzangi, K. (2015). IdeoKar: an ideogram constructing and karyotype analyzing software. *Caryologia*, 68, 31-35.
- Miura, I. & Ogata, M. (2013). Change of heterogametic sex from male to female: Why so easy in the frog? *Chrom. Sci.*, 16, 3-9.
- Muller, H. J. (1964). The relation of recombination to mutational advance. *Mutat. Res.*, 1, 2-9.
- Nanda, I., Zend-Ajusich, E., Shan, Z., Grützner, F., Scharl, M., Burt, D. W., Koehler, M., Fowler, V. M., Goodwin, G., Schneider, W. J., Mizuno, S., Dechant, G., Haaf, T. & Schmid, M. (2000). Conserved synteny between the chicken Z sex chromosome and human chromosome 9 includes the male regulatory gene DMRT1: a comparative (re)view on avian sex determination. *Cytogenet. Cell Genet.*, 89, 67-78.

- Nanda, I., Schlegelmilch, K., Haaf, T., Scharl, M. & Schmid, M. (2008). Synteny conservation of the Z chromosome in 14 avian species (11 families) supports a role for Z dosage in avian sex determination. *Cytogenet. Genome Res.*, 122, 150-156.
- Nascimento, J., Quinderé, Y. R. S. D., Recco-Pimentel, S. M., Lima, J. R. F. & Lourenço, L. B. (2010). Heteromorphic Z and W sex chromosomes in *Physalaemus ephippifer* (Steindachner, 1864) (Anura, Leiuperidae). *Genetica*, 138, 1127-1132.
- Nur, U. & Nevo, E. (1969). Supernumerary chromosomes in the cricket frog *Acris crepitans*. *Caryologia*, 22, 97-102.
- Odierna, G., Aprea, G., Capriglione, T., Castellano, S. & Balletto, E. (2007). Cytological evidence for population-specific sex chromosome heteromorphism in Palaearctic green toads (Amphibia, Anura). *J. Biosci.*, 32, 763-768.
- Ohno, S. (1967). *Sex chromosomes and sex-linked genes*. New York, Springer-Verlag.
- Orrico, V. G. D., Nunes, I., Mattedi, C., Fouquet, A., Lemos, A. W., Rivera-Correa, M., Lyra, M., Loebmann, D., Pimenta, B. V. S., Caramaschi, U., Rodrigues, M. T. & Haddad, C. F. B. (2017). Integrative taxonomy supports the existence of two distinct species within *Hypsiboas crepitans* (Anura: Hylidae). *Salamandra*, 53, 99-113.
- Orrico, V. G. D., Duellman, W. E., Souza, M. B. & Haddad, C. F. B. (2013). The taxonomic status of *Dendropsophus allenorum* and *D. timbeba* (Anura: Hylidae). *J. Herpetol.*, 46, 615-618.
- Ohta, S., Sumida, M. & Nishioka, M. (1999). Sex-determining mechanism in *Buergeria buergeri* (Anura, Rhacophoridae). III. Does the ZZW triploid frog become female or male? *J. Exp. Zool.*, 283, 295-306.
- Padial, J. M., Grant, T. & Frost, D. R. (2014). Molecular Systematics of terraranas (Anura: Brachycephaloidea) with an assessment of the effects of alignment and optimality criteria. *Zootaxa*, 3825, 1-132.
- Palestis, B. G., Trivers, R., Burt, A. & Jones, R. N. (2004). The distribution of B chromosomes across species. *Cytogenet. Genome Res.*, 106, 151-158.
- Pansonato, A., Veiga-Menoncello, A. C. P., Mudrek, J. R., Martin, J., Recco-Pimentel, S. M. & Strussmann, C. (2016). Two New Species of *Pseudopaludicola* (Anura: Leptodactylidae: Leiuperinae) from Eastern Bolivia and Western Brazil. *Herpetologica*, 72, 235-255.
- Pansonato-Alves, J. C., Serrano, E. A., Utsunomia, R., Camacho, J. P. M., Costa-Silva, G. J., Vicari, M. R., Artoni, R. F., Oliveira, C. & Foresti, F. (2014). Single origin of sex chromosomes and multiple origins of B chromosomes in fish genus *Characidium*. *PLoS ONE*, 9, e107169.
- Patawang, I., Phimphan S., Tanomtong, A. & Nithikulworawong, N. (2014). Rice frog, *Fejervarya limnocharis* (Anura, Ranidae) from Northeast Thailand. *Cytologia*, 79, 141-150.
- Paul-Prasanth, B., Matsuda, M., Lau, E. L., Suzuki, A., Sakai, F., Kobayashi, T. & Nagahama, Y. (2006). Knock-down of DMY initiates female pathway in the genetic male medaka, *Oryzias latipes*. *Biochem. Biophys. Res. Commun.*, 351, 815-819.
- Pendas, A. M., Moran, P., Freije, J. P. & Garcia-Vazquez, E. (1994). Chromosomal mapping and nucleotide sequence of two tandem repeats of Atlantic salmon 5S rDNA. *Cytogenet. Cell Genet.*, 67, 31-36.
- Perrin, N. (2009). Sex reversal: a fountain of youth for sex chromosomes? *Evolution*, 63, 3043-3049.

- Pita, S., Panzera, F., Sánchez, A., Palomeque, T. & Lorite, P. (2017). Chromosome painting in Triatomine insects reveals shared sequences between X chromosomes and autosomes. *J. Med. Entomol.*, 54, 44-49.
- Plohl, M., Luchetti, A., Mestrovic, N. & Mantovani, B. (2008). Satellite DNAs between selfishness and functionality: structure, genomics and evolution of tandem repeats in centromeric (hetero)chromatin. *Gene*, 409, 72-82.
- Plohl, M., Meštrović, N. & Mravinac, B. (2012). Satellite DNA evolution. *Genome Dyn.*, 7, 126-152.
- Ptacek, M. B., Gerhardt, H. C. & Sage, R. D. (1994). Speciation by polyploidy in tree frogs: multiple origins of the tetraploid, *Hyla versicolor*. *Evolution*, 48, 898-908.
- Pyron, R. A. & Wiens, J. J. (2011). A large-scale phylogeny of Amphibia including over 2800 species, and a revised classification of extant frogs, salamanders, and caecilians. *Mol. Phylogenet. Evol.*, 61, 543-583.
- Reeves, A. & Tear, J. (2000). *Micromasure*. Version 3.3. Available from: URL: <http://www.colostate.edu/Depts/Biology/MicroMeasure>.
- Richard, G. F., Kerrest, A. & Dujon, B. (2008). Comparative genomics and molecular dynamics of DNA repeats in Eukaryotes. *Microbiol. Mol. Biol. Rev.*, 72, 686-727.
- Rodrigues, D. S., Rivera, M. & Lourenço, L. B. (2012). Molecular organization and chromosomal localization of 5S rDNA in Amazonian *Engystomops* (Anura, Leiuperidae). *BMC Genet.*, 13, 17.
- Rodrigues, P. C., Aguiar, Jr. O., Serpieri, F., Lima, A. P., Uetanabaro, M. & Recco-Pimentel, S. M. (2011). Chromosome analysis of five Brazilian species of poison frogs (Anura: Dendrobatidae). *J. Genetics*, 90, 31-37.
- Romanenko, S. A., Serdyukova, N. A., Perelman, P. L., Pavlova, S. V., Bulatova, N. S., Golenishchev, F. N., Stanyon, R. & Graphodatsky, A. S. (2017). Intrachromosomal rearrangements in rodents from the perspective of comparative region-specific painting. *Genes*, 8, 215.
- Ron, S. R., Santos, J. C. & Cannatella, D. C. (2006). Phylogeny of the túngara frog genus *Engystomops* (= *Physalaemus pustulosus* species group; Anura: Leptodactylidae). *Mol. Phylogenet. Evol.*, 39, 392-402.
- Ron, S. R., Toral, E., Rivera, M. & Terán-Valdez, A. (2010). A new species of *Engystomops* (Anura: Leiuperidae) from southwestern Ecuador. *Zootaxa*, 2606, 25-49.
- Rosa, C., Aguiar, O., Giaretta, A. A. & Recco-Pimentel, S. M. (2003). Karyotypic variation in *Megaelosia* (Anura, Hylodinae) with the first description of a B-chromosome in a leptodactylid frog. *Copeia*, 2003, 166-174.
- Rosset, S. D., Baldo, D., Lanzone, C. & Basso, N. G. (2006). Review of the geographic distribution of diploid and tetraploid populations of the *Odontophrynus americanus* species complex (Anura: Leptodactylidae). *J. Herpetol.*, 40, 465-477.
- Rubtsov, N. B., Karamysheva, T. V., Andreenkova, O. V., Bochkarev, M. N., Kartavtseva, I. V., Roslik, G. V. & Borissov, Y. M. (2004). Comparative analysis of micron and macro B chromosomes in the Korean field mouse *Apodemus peninsulae* (Rodentia, Murinae) performed by chromosome microdissection and FISH. *Cytogenet. Genome Res.*, 106, 289-294.
- Ruiz-Ruano, F. J., Cabrero, J., López-Léon, M. D. & Camacho, J. P. M. (2016). Satellite DNA content illuminates the ancestry of supernumerary (B) chromosome. *Chromosoma*, 126, 487-500.

- Saba, N. & Tripathi, N. K. (2014). Preliminary cytogenetic study and report of ZZ/ZW sex chromosomes in the bullfrog, *Hoplobatrachus tigerinus* (Anura, Amphibia) from high altitude area of Jammu and Kashmir, India. *The Nucleus*, 57, 55-59.
- Saito, Y., Edpalina, R. R. & Abe, S. (2007). Isolation and characterization of salmonid telomeric and centromeric satellite DNA sequences. *Genetica*, 131, 157-166.
- Sangpakdee, W., Pimphan, S., Tengjaroenkul, B., Pinthong, K., Neeratanaphan, L. & Tanomtong, A. (2017). Cytogenetic study of three microhylid species (Anura, Microhylidae) from Thailand. *Cytologia*, 82, 67-74.
- Schartl, M., Schmid, M. & Nanda, I. (2016). Dynamics of vertebrate sex chromosome evolution: from equal size to giants and dwarfs. *Chromosoma*, 125, 553-571.
- Schempp, W. & Schmid, M. (1981) Chromosome banding in Amphibia. VI. BrdU-replication patterns in Anura and demonstration of XX/XY sex chromosomes in *Rana esculenta*. *Chromosoma*, 83, 697-710.
- Schmid, M. (1978). Chromosome banding in Amphibia. II. Constitutive heterochromatin and nucleolus organizer regions in Ranidae, Microhylidae and Rhacophoridae. *Chromosoma*, 68, 131-148.
- Schmid, M., Vitelli, L. & Batistoni, R. (1987). Chromosome banding in Amphibia. XI. Constitutive heterochromatin, nucleolus organizers, 18S+28S and 5S ribosomal RNA genes in Ascaphidae, Pipidae, Discoglossidae and Pelobatidae. *Chromosoma*, 95, p. 271-284.
- Schmid, M., Steinlein, C., Friedl, R., de Almeida, C. G., Haaf, T., Hillis, D. M. & Duellman, W. E. (1990). Chromosome banding in Amphibia. XV. Two types of Y chromosomes and heterochromatin hypervariability in *Gastrotheca pseustes* (Anura, Hylidae). *Chromosoma*, 99, 413-423.
- Schmid, M., Steinlein, C. & Feichtinger, W. (1992). Chromosome banding in Amphibia. XVII. First demonstration of multiple sex chromosomes in amphibians: *Eleutherodactylus maussi* (Anura, Leptodactylidae). *Chromosoma*, 101, 284-292.
- Schmid, M., Feichtinger, W., Steilein, C., Haaf, T., Schartl, M., Visbal-García, R., Manzanilla-Pupo, J. & Fernández-Badillo, A. (2002a). Chromosome banding in Amphibia. XXVI. Coexistence of homomorphic XY sex chromosomes and a derived Y-autosome translocation in *Eleutherodactylus maussi* (Anura, Leptodactylidae). *Cytogenet. Genome Res.*, 99, 330-340.
- Schmid, M., Feichtinger, W., Steinlein, C., Nanda, I., Mais, C., Haaf, T., Visbal-Garcia, R. & Fernández-Badillo, A. (2002b). Chromosome banding in Amphibia. XXII. Atypical Y chromosomes in *Gastrotheca walkeri* and *Gastrotheca ovifera* (Anura, Hylidae). *Cytogenet. Genome Res.*, 96, 228-238.
- Schmid, M., Feichtinger, W., Steilein, C., Visbal-García, R. & Fernández-Badillo, A. (2003). Chromosome banding in Amphibia. XXVIII. Homomorphic XY sex chromosomes and a derived Y-autosome translocation in *Eleutherodactylus riveroi* (Anura, Leptodactylidae). *Cytogenet. Genome Res.*, 101, 62-73.
- Schmid, M. & Steinlein, C. (2003). Chromosome banding in Amphibia. XXIX. The primitive XY/XX sex chromosomes of *Hyla femoralis* (Anura, Hylidae). *Cytogenet. Genome Res.*, 101, 74-79.
- Schmid, M., Steinlein, C., Bogart, J. P., Feichtinger, W., León, P., La Marca, E., Díaz, L. M., Sanz, A., Chen, S. H. & Hedges, S. B. (2010). *The Chromosomes of Terraranan Frogs: insights into vertebrate cytogenetics*. Basel, SW. S. Karger AG, 559 p.

- Schmid, M., Steinlein, C., Bogart, J. P., Feichtinger, W., Haaf, T., Nanda, I., del Pino, E. M., Duellman, W. E. & Hedges, S. B. (2012). The hemiphractid frogs. Phylogeny, embryology, life history, and cytogenetics. *Cytogenet. Genome Res.*, 138, 69-384.
- Schmid, M. & Steinlein, C. (2015). Chromosome banding in Amphibia. XXXII. The genus *Xenopus* (Anura, Pipidae). *Cytogenet. Genome Res.*, 145, 201-217.
- Session, A. M., Yoshinobu, U., Taejoon, K., Chapman, J. K., Toyoda, A., Takahashi, S., Fukui, A., Hikosaka, A., Suzuki, A., Kondo, M., Heeringen, S. J., Quigley, I., Heinz, S., Ogino, H., Ochi, H., Hellsten, U., Lyons, J. B., Simakov, O., Putnam, N., Stites, J., Kuroki, Y., Tanaka, T., Michiue, T., Watanabe, M., Bogdanovic, O., Lister, R., Georgiou, G., Paranjpe, S. S., Kruijsbergen, I., Shu, S., Carlson, J., Kinoshita, T., Ohta, Y., Mawaribuchi, S., Jenkins, J., Grimwood, J., Schmutz, J., MitroS, T., Mozaffari, S., Suzuki, Y., Haramoto, Y., Yamamoto, T. S., Takagi, C., Heald, R., Miller, K., Haudenschild, C., Kitzman, J., Nakayama, T., Izutsu, Y., Robert, J., Fortriede, J., Burns, K., Lotay, V., Karimi, K., Yasuoka, Y., Dichmann, D. S., Flajnik, M. F., Houston, D. W., Shendure, J., DuPasquier, L., Vize, P. D., Zorn, A. M., Ito, I., Marcotte, E., Wallingford, J. B., Ito, Y., Asashima, M., Ueno, N., Matsuda, Y., Veenstra, G. J. V., Fujiyama, A., Harland, R. M., Taira, M. & Rokhsar, D. S. (2016). Genome evolution in the allotetraploid frog *Xenopus laevis*. *Nature*, 538, 336-343.
- Sharbel, T. F., Green, D. M. & Houben, A. (1998). B chromosome origin in the endemic New Zealand frog *Leiopelma hochstetteri* through sex chromosome evolution. *Genome*, 41, 14-22.
- Shibata, F., Hizume, M. & Kuroki, Y. (1999). Chromosome painting of Y chromosomes and isolation of a Y chromosome-specific repetitive sequence in the dioecious plant *Rumex acetosa*. *Chromosoma*, 108, 266-270.
- Silva, D. M. Z., Daniel, S. N., Camacho, J. P. M., Utsunomia, R., Ruiz-Ruano, F. J., Penitente, M., Pansonato-Alves, J. C., Hashimoto, D. T., Oliveira, C., Porto-Foresti, F. & Foresti, F. (2016). Origin of B chromosomes in the genus *Astyanax* (Characiformes, Characidae) and the limits of chromosome painting. *Mol. Genet. Genomics*, 291, 1407-1418.
- Sinclair, A. H., Berta, P., Palmer, M. S., Hawkins, J. R., Griffiths, B. L., Smith, M. J., Foster, J. W., Frischauf, A. M., Lovell-Badge, R. & Goodfellow, P. N. (1990). A gene from the human sex determining region encodes a protein with homology to a conserved DNA-binding motif. *Nature*, 346, 240-244.
- Singh, L., Purdom, F. & Jones, K. W. (1976). Satellite DNA and evolution of sex chromosomes. *Chromosoma*, 59, 43-62.
- Singh, L., Purdom, I. F. & Jones, K. W. (1980). Sex chromosome associated satellite DNA: evolution and conservation. *Chromosoma*, 79, 137-157.
- Slamovits, H. C. & Rossi, M. S. (2002). Satellite DNA: agent of chromosomal evolution in mammals. *Mastozoologia Neotropical*, 9, 297-308.
- Smith, C. A., Roeszler, K. N., Ohnesorg, T., Cummins, D. M., Farlie, P. G., Doran, T. J. & Sinclair, A. H. (2009). The avian Z-linked gene DMRT1 is required for male sex determination in the chicken. *Nature*, 461, 267-271.
- Steinemann, S. & Steinemann, M. (2005). Retroelements: tools for sex chromosome evolution. *Cytogenet. Genome Res.*, 110, 134-143.
- Stöck, M., Horn, A., Grossen, C., Lindtke, D., Sermier, R., Betto-Colliard, C., Dufresnes, C., Bonjour, E., Dumas, Z., Luquet, E., Maddalena, T., Sousa, H. C., Martinez-Solano, I. &

- Perrin, N. (2011). Ever-young sex chromosomes in European tree frogs. *PLoS Biol.*, 9, e1001062.
- Stöck, M., Savary, R., Zaborowska, A., Gorecki, G., Brelsford, A., Rozenblut-Kocisty, B., Ogielska, M. & Perrin, N. (2013a). Maintenance of ancestral sex chromosomes in Palearctic tree frogs: direct evidence from *Hyla orientalis*. *Sex. Dev.*, 7, 261-266.
- Stöck, M., Savary, R., Betto-Colliard, C., Biollay, S., Jourdan-Pineau, H. & Perrin, N. (2013b). Low rates of X-Y recombination, not turnovers, account for homomorphic sex chromosomes in several diploid species of Palearctic green toads (*Bufo viridis* subgroup). *J. Evol. Biol.*, 26, 674-682.
- Suárez, P., Cardozo, D., Baldo, D., Pereyra, M. O., Faivovich, J., Orrico, V. G. D., Catroli, G. F., Grabiele, M., Bernarde, P. S., Nagamachi, C. Y., Haddad, C. B. & Pieczarka, J. C. (2013). Chromosome evolution in Dendropsophini (Amphibia, Anura, Hyliinae). *Cytogenet. Genome Res.*, 141, 295-308.
- Sun, Y. B., Xiong, Z. J., Xiang, X. Y., Liu, S. P., Zhou, W. W., Tu, X. L., Zhong, L., Wang, L., Wu, D. D., Zhang, B. L., Zhu, C. L., Yang, M. M., Chen, H. M., Li, F., Zhou, F., Feng, S. H., Huang, C., Zhang, G. J., Irwin, D., Hillis, D. M., Murphy, R. W., Yang, H. M., Che, J., Wang, J. & Zhang, Y. P. (2015). Whole-genome sequence of the Tibetan frog *Nanorana parkeri* and the comparative evolution of tetrapod genomes. *P. Natl. Acad. Sci. USA*, 112, E1257-E1262.
- Targueta, C. P., Rivera, M., Souza, M. B., Recco-Pimentel, S. M. & Lourenço, L. B. (2010). Cytogenetic contributions for the study of the Amazonian *Engystomops* (Anura; Leiuperidae) assessed in the light of phylogenetic relationships. *Mol. Phylogenet. Evol.*, 54, 709-725.
- Targueta, C. P., Rivera, M. & Lourenço, L. B. (2012). Karyotypic differentiation via 2n reduction and a finding of a case of triploidy in anurans of the genus *Engystomops* (Anura, Leiuperidae). *Genetica*, 139, 1339-1347.
- Teixeira, L., Targueta, C. P., Seger, K. R., Orrico, V. D. G. & Lourenço, L. B. (2016). Comparative cytogenetics of treefrogs of the *Dendropsophus marmoratus* (Laurenti, 1768) group: conserved karyotypes and interstitial telomeric sequences. *Comp. Cytogenet.*, 10, 753-767.
- Terurel, M., Cabrero, J., Montiel, E. E., Acosta, M. J., Sánchez, A. & Camacho, J. P. M. (2009a). Microdissection and chromosome painting of X and B chromosomes in *Locusta migratoria*. *Chrom. Res.*, 17, 11-18.
- Teruel, M., Cabrero, J., Perfectti, F., Acosta, M. J. & Sánchez, A. (2009b). Microdissection and Chromosome Painting of X and B chromosomes in the Grasshoppers *Eyprepocnemis plorans*. *Cytogenet. Genome Res.*, 125, 286-291.
- Toledo, F. L., Siqueira, S., Duarte, T. C., Veiga-Menoncello, A. C. P., Recco-Pimentel, S. M. & Haddad, C. F. B. (2010). Description of a new species of *Pseudopaludicola* Miranda-Ribeiro, 1926 from the state of São Paulo, Southeastern Brazil (Anura, Leiuperidae). *Zootaxa*, 2496, 38-48.
- Tomatis, C. G., Baldo, D., Kolenc, F. & Borteiro, C. (2009). Chromosomal variation in the species of the *Physalaemus henselii* group (Anura, Leiuperidae). *J. Herpetol.*, 43, 555-560.
- Ugarčević, D. & Plohl, M. (2002). Variation in satellite DNA profiles – causes and effects. *EMBO J.*, 21, p. 5955-5959.
- Uno, Y., Nishida, C., Oshima, Y., Yokoyama, S., Miura, I., Matsuda, Y. & Nakamura, M. (2008). Comparative chromosome mapping of sex-linked genes and identification of sex

- chromosomal rearrangements in the Japanese wrinkled frog (*Rana rugosa*, Ranidae) with ZW and XY sex chromosome systems. *Chromosome Res.*, 16, 637–647.
- Uno, Y., Nishida, C., Takagi, C., Ueno, N. & Matsuda, Y. (2013). Homoeologous chromosomes of *Xenopus laevis* are highly conserved after whole-genome duplication. *Heredity*, 111, 430–436.
- Uno, Y., Nishida, C., Takagi, C., Igawa, T., Ueno, N., Sumida, M. & Matsuda, Y. (2015). Extraordinary diversity in the origins of sex chromosomes in anurans inferred from comparative gene mapping. *Cytogenet. Genome Res.*, 145, 218–229.
- Vacher, J. P., Kok, P. J. R., Rodrigues, M. T., Lima, J. D., Lorenzini, A., Martinez Q., Fallet, M., Courtois, E. A., Blanc, M., Gaucher, P., Dewynter, M., Jairam, R., Ouboter, P., Thébaud, C. & Fouquet, A. (2017). Cryptic diversity in Amazonian frogs: Integrative taxonomy of the genus *Anomaloglossus* (Amphibia: Anura: Aromobatidae) reveals a unique case of diversification within the Guiana Shield. *Mol. Phylogenet. Evol.*, 112, 158–173.
- Veiga-Menoncello, A. C. P., Lourenço, L. B., Strussmann, C., Rossa-Feres, D. C., Andrade, G. V., Giarretta, A. & Recco-Pimentel, S. M. (2014). A phylogenetic analysis of *Pseudopaludicola* (Anura) providing evidence of progressive chromosome reduction. *Zool. Scripta*, 43, 261–272.
- Vences, M., Thomas, M., Van Der Meijden, A., Chiari, Y. & Vieites, D. R. (2005a). Comparative performance of the 16S rRNA gene in DNA barcoding of amphibians. *Front. Zool.*, 2, 1–12.
- Vences, M., Thomas, M., Bonett, R. M. & Vieites, D. R. (2005b). Deciphering amphibian diversity through DNA barcoding: chances and challenges. *Phil. Trans. R. Soc. B.*, 360, 1859–1868.
- Vicari, M. R., Pistune, H. F. M., Castro, J. P., Almeida, M. C., Bertollo, L. A., Moreira-Filho, O., Camacho, J. P. & Artoni, R. F. (2011). New insights on the origin of B chromosomes in *Astyanax scabripinnis* obtained by chromosome painting and FISH. *Genetica*, 139, 1073–1081.
- Vittorazzi, S. E., Lourenço, L. B., Del-Grande, M. L. & Recco-Pimentel, S. M. (2011). Satellite DNA derived from 5S rDNA in *Physalaemus cuvieri* (Anura, Leiuperidae). *Cytogenet. Genome Res.*, 134, 101–107.
- Vittorazzi, S. E., Lourenço, L. B. & Recco-Pimentel, S. M. (2014a). Long-time evolution and highly dynamic satellite DNA in leptodactylid and hylodid frogs. *BMC Genet.*, 15, 111.
- Vittorazzi, S. E., Quinderé, Y. R. S. D., Recco-Pimentel, S. M., Tomatis, C., Baldo, D., Lima, J. R. F., Ferro, J. M., Lima, J. D. & Lourenço, L. B. (2014b). Comparative cytogenetics of *Physalaemus albifrons* and *Physalaemus cuvieri* species groups (Anura, Leptodactylidae). *Comp. Cytogenet.*, 8, 103–123.
- Vittorazzi, E., Lourenço, L. B. & Recco-Pimentel, S. M. (2016). Chromosomal analysis of *Physalaemus kroyeri* and *Physalaemus cicada* (Anura, Leptodactylidae). *Com. Cytogenet.*, 10, 311–323.
- Volff, J. N., Nanda, I., Schmid, M. & Scharl, M. (2007). Governing sex determination in fish: regulatory patches and ephemeral dictators. *Sex. Dev.*, 1, 85–99.
- Wang, X., Zhang, Q., Ren, J., Jiang, Z., Wang, C., Zhuang, W. & Zhai, T. (2009). The preparation of sex-chromosome-specific painting probes and construction of sex chromosome DNA library in half-smooth tongue sole (*Cynoglossus semilaevis*). *Aquaculture*, 297, 78–84.

- Warburton, P. E., Dolled, M., Mahmood, R., Alonso, A., Li, S., Naritomi, K., Tohma, T., Nagai, T., Hasegawa, T., Ohashi, H., Govaerts, L. C., Eussen, B. H., Van Hemel, J. O., Lozzio, C., Schwartz, S., Dowhanick-Morrisette, J. J., Spinner, N. B., Rivera, H., Crolla, J. A., Yu, C. & Warburton, D. (2000). Molecular cytogenetic analysis of eight inversion duplications of human chromosome 13q that each contain a neocentromere. *Am. J. Hum. Genet.*, 66, 1794-1806.
- Wu, G. F. & Zhao, E. M. (1985). Preliminary studies on karyotypes of the genus *Amolops* of the Hengduan Mountains. *Acta Herpetol. Sinica*, 4, 276-282.
- Yamada, K., Nishida-Umehara, C. & Matsuda, Y. (2004). A new Family of satellite DNA sequences as a major component of centromeric heterochromatin in owls (Strigiformes). *Chromosoma*, 112, 277-287.
- Yoshimura, A., Nakata, A., Mito, T. & Noji, S. (2006). The characteristics of karyotypes and telomeric satellite DNA sequences in the cricket, *Gryllus bimaculatus* (Orthoptera, Gryllidae). *Cytogenet. Genome Res.*, 112, 329-336.
- Yoshimoto, S., Okada, E., Umemoto, H., Tamura, K., Uno, Y., Nishida-Umehara, C., Matsuda, Y., Takamatsu, N., Shiba, T. & Ito, M. (2008). A W-linked DM-domain gene, DM-W, participates in primary ovary development in *Xenopus laevis*. *P. Natl. Acad. Sci. USA*, 105, 2469-2474.

Chapter 55

HUMAN CYTOGENETIC BIOMONITORING

***Armanda Teixeira-Gomes^{1,2}, Solange Costa^{1,2,*}
and João Paulo Teixeira^{1,2}***

¹EPIUnit – Instituto de Saúde Pública, Universidade do Porto,
Porto, Portugal

²Environmental Health Department, National Institute of Health,
Porto, Portugal

ABSTRACT

Humans are daily exposed to a variety of potentially harmful agents in the air they breathe, liquids they drink, food they eat and products they use. Long-standing evidence of the bond between health and environment has led to the recognition for the need of sustainable development. On the other hand, there is an increasing global awareness of the inevitable limits of individual health care and the need to complement such services with effective public health strategies. According to World Health Organization (WHO), cancer is a leading cause of death worldwide. Additionally, at least 200 000 people die every year from occupational or work-related cancers. Exposure assessment aims at prevention. Establishing the health effects of various activities and exposures requires information about the levels of exposure and the biological effects resulting from the interaction between the organism and the chemical agent. The resulting data provides a basis for designing effective prevention and mitigation strategies. Cytogenetic endpoints have long been applied in surveillance of human genotoxic exposure and early effects of genotoxic carcinogens. Assays measuring chromosomal aberrations (CAs), micronucleus (MN) and sister chromatid exchange (SCE) in lymphocytes are well-established techniques extensively used in human biomonitoring studies to assess DNA damage at the chromosomal level. The relevance of cytogenetic alterations as a cancer risk biomarker is further supported by epidemiologic data linking CAs and MN with cancer risk in human populations. Thus, the use of cytogenetic markers in human biomonitoring is of paramount importance due to its predictability regarding deleterious effects resulting from the exposure to environmental stressors.

* Corresponding Author's Email: solange.costa2@gmail.com.

Keywords: human biomonitoring, cytogenetic, chromosomal aberrations, micronucleus and sister chromatid exchange

AN INTRODUCTION TO HUMAN BIOMONITORING

Humans are exposed to several potential harmful agents in their daily life activities, in the air, water, soil, food and products they use [1]. Identification, evaluation and control of environmental hazards is of extreme importance to prevent the health effects of exposure to substances. Human biomonitoring is frequently used to provide early warning signals of excessive exposure, allowing the measure of the substances themselves, their metabolites or markers of subsequent health effects in body fluids or tissues [2]. Studying adverse health outcomes related to the environmental exposures (in the living and working environment) is a major societal challenge today.

Exposure can be assessed by biological monitoring, with measures of biological endpoints (biomarkers or biological indicators). Biological monitoring assesses the interaction of potential harmful agents with the biological systems by analyzing chemicals, metabolites, enzymes and other biochemical substances in tissues, body fluids or other accessible samples [1]. Other method used to protect human health in case of exposure to potential harmful substances is called environmental monitoring, that encompasses the measure of those substances in environmental matrices like air, water, soil, food, etc. Environmental monitoring is particularly important to the determination of the sources of exposure. The two methods, biological and environmental monitoring, complete each other and are commonly used simultaneously, since they provide different but complementary information [3, 4]. Nevertheless, human biomonitoring is essential because it allows a better estimation of the internal dose of a compound (dose really taken up) and consequently, its potential health risks, despite the route of exposure - inhalation, absorption through the skin, and ingestion.

Human biomonitoring depends on the use of biomarkers that are defined as a xenobiotically induced alteration in cellular or biochemical components or processes, structures, or functions measurable in a biological system or sample [5]. A good biomarker must be sensitive, specific, relevant, reproducible and measurable in the population. Biomarkers are traditionally divided into three main categories - biomarkers of exposure, susceptibility and effect. Biomarkers of exposure are defined as the compound, parent compound or its metabolites, or the product of an interaction between the xenobiotic and its target molecule or cell that is measured in a compartment within an organism (e.g., lead levels in blood; arsenic levels in hair). Biomarkers of susceptibility are indicators of an inherent or acquired capacity of an organism to respond to the challenge of exposure to a specific chemical (e.g., polymorphic genes of metabolizing enzymes). Biomarkers of effect are measurable biochemical, physiological, behavioral, functional or other alterations within an organism induced by the exposure and depending on the magnitude of the alterations can be recognized as associated with an established or possible health impairment or disease [1, 6].

Biomarkers of exposure are specific for the identification of the agent of exposure, while the biomarkers of effect are usually unspecific for the causative agent [6]. Although they are less specific in the identification of the agent of exposure, they can be more predictive of the induced toxicity. They are biological indicators of the response of the body to exposures indicating early sub-clinical changes, which can develop into pathological consequences [7].

Biomarkers of effect either indicate early processes that precede a disease or predict the development and presence of disease [8]. The cytogenetic biomarkers that are going to be addressed further in this chapter belong to the category of biomarkers of effect, and are extensively used tools in human biomonitoring studies, particularly in occupational studies, to evaluate the genotoxic potential of the exposure to hazardous chemicals [9].

CYTOGENETIC BIOMARKERS IN HUMAN BIOMONITORING

Cytogenetic biomarkers are extensively used tools in human population studies to assess the impact of environmental, occupational and medical factors on genomic stability [9]. Nowadays, the most used cytogenetic endpoints in human biomonitoring studies are chromosomal aberrations (CAs) and micronucleus assay (MN). The increased frequency of these indicators in exposed workers is a consistent finding in several studies [10-17]. Also, the interest in assays measuring these cytogenetic alterations has increased since they have been linked with cancer risk in human epidemiologic studies [18, 19].

One critical issue in the use of cytogenetic endpoints in human biomonitoring is the selection of the biological specimen to investigate [20]. These endpoints are usually assessed in readily available surrogate cells to estimate events occurring in target tissues and to provide early warning signals for adverse health effects. Despite the documented usefulness of non-blood cells (e.g., buccal, nasal, urothelial) in biomonitoring [21] the peripheral blood lymphocytes (PBLs) are the most frequently used cells in human studies. The major reason to use lymphocytes is because they can be damaged in any tissue/organ specific toxic environment, since these cells circulate throughout the body and have reasonably long life-span [22]. The preference in the use of lymphocytes is also given to the fact that they exist in large number in the human body and are relatively simple to harvest.

In this chapter we are going to review the mechanisms and methodology of the major cytogenetic endpoints used in human biomonitoring: CAs, MN and sister chromatid exchanges (SCEs). An overview of these cytogenetic assays can be observed in Figure 1.

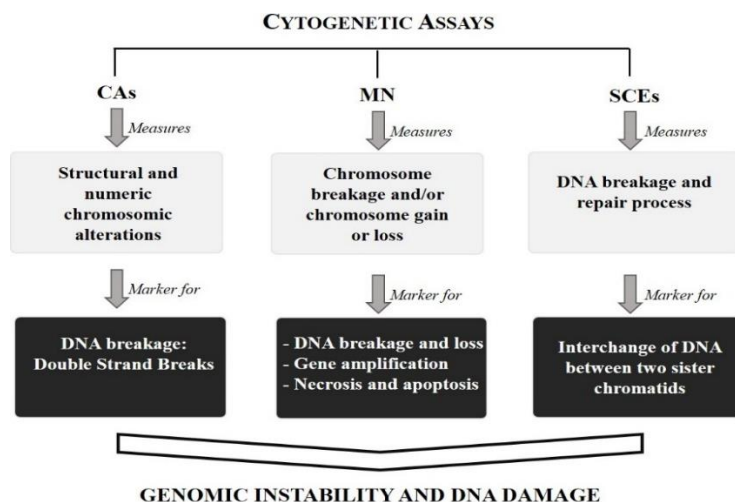


Figure 1. Overview of the cytogenetic assays.

Chromosomal Aberrations (CAs): Mechanisms and Methodology

Chromosomal aberrations (CAs) are changes in normal chromosome structure or number that can occur as a consequence of exposure to chemicals or radiation [23]. Chromosome structural changes have been assessed for decades by CA assay in PBLs to evaluate the exposure to contaminants in occupational and environmental settings.

Structural CAs may be induced by direct DNA breakage, replication on a damaged DNA template, inhibition of DNA synthesis or other mechanisms (e.g., topoisomerase II inhibitors) [23]. The formation of structural CAs requires one or several double-strand breaks (DSBs). Mechanistic and morphological differences categorize these CAs in two types: chromosome-type aberrations and chromatid-type aberrations. Concerning the morphology, chromosome-type aberrations involve both sister chromatids of a chromosome or more chromosomes, and chromatid-type aberrations only one chromatid of one or several chromosomes [24]. Regarding the mechanism, chromosome-type aberrations and chromatid-type aberrations differ in the type of initial lesion (induced by different agents) and the DNA repair mechanisms that follow. Those lesions are commonly repaired by three main mechanisms: homologous recombination (HR), non-homologous end joining (NHEJ) repair and single-strand annealing (SSA). HR is an error-free process, highly accurate that precisely restores the original sequence at the break. NHEJ is an error-prone repair process that joins directly the two broken ends which usually leads to small-scale alterations at the break site. SSA mainly leads to the formation of interstitial deletions [25].

Chromosome-type aberrations are usually generated *in vivo* by S-phase independent clastogens in resting G₀/G₁ lymphocytes. After DNA synthesis and chromosome duplication, the aberrations formed are doubled and chromosome-type breaks and exchanges (e.g., dicentric and ring chromosomes) are seen in metaphase. The duplication originates symmetric lesions on both sister chromatids. They reflect DSBs that are incompletely repaired or unrepaired by NHEJ repair and SSA.

Chromatid-type aberrations arise predominantly *in vitro* during the S-phase of the cultured lymphocytes. These type of breaks and exchanges are the response to base modifications and single-strand breaks (SSBs) induced *in vivo* by S-phase dependent clastogens. DSBs are a consequence of these lesions during DNA replication. Incomplete or failed repair of these lesions by conservative HR will generate chromatid-type aberrations formation in a subsequent metaphase [26, 27]. Chromatid-type aberrations formation is the mechanism most likely to be applicable to human biomonitoring studies since it represents the common process by which clastogenic chemicals induce CA.

As previously mentioned, the detection of structural CAs in human biomonitoring studies is performed in lymphocytes using the CA assay. This assay is the most widely used and best validated because the mechanisms for CAs induction are better understood and most environmental toxic substances have been shown to induce structural CAs [22, 28]. In fact, the CA assay is recognized as one of key tests for genotoxic compounds, being its protocol defined by OECD guidelines for the testing of chemicals (OECD TG 473) [29].

Lymphocytes are cultured in the presence of a mitogen, for instance phytohemagglutinin (PHA), in order to stimulate cell division. A spindle inhibitor (e.g., colcemid, colchicine) is added to the culture after a determined period to stop the replication when cells are in the metaphase. A schematic of the assay is represented in Figure 2.

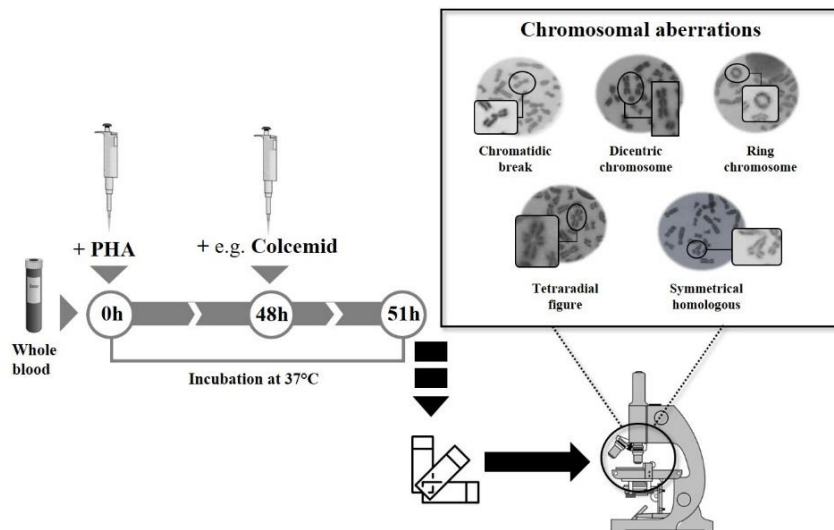


Figure 2. Schematic representation of CAs assay described by Roma-Torres et al. 2006 [31] and examples of microscopical observations.

Afterwards, the metaphase-arrested cells are ready for scoring. This method is not appropriate for the estimation of numerical CAs, since chromosome loss can occur. Numerical CAs are changes in normal chromosome number, a consequence of abnormal cell division. Cells are considered aneuploid when they contain extra (hyperploid) or missing (hypoploid) chromosome. These changes may occur due to damage to the mitotic spindle and associated elements, damage to chromosome sub-structures, alterations in cell physiology and mechanical disruption [26, 30]. The additional use of fluorescence in situ hybridization (FISH) methods that have higher sensitivity and efficiency, enables the detection of subtle aberrations, as chromosome-specific rearrangements and/or loss, that are not apparent when using the conventional Giemsa staining [24].

Micronucleus (MN): Mechanisms and Methodology

Micronucleus (MN) are extra-nuclear bodies resultant from damaged chromosome fragments and/or whole chromosomes that are not included in the daughter nuclei at completion of telophase during mitosis. The chromosome fragments or whole chromosomes lag behind during the segregation process in anaphase because they did not attach properly to the spindle. The DNA material is eventually enclosed by a nuclear membrane being morphologically similar to nuclei (with the exception of its smaller size) after conventional nuclear staining [32].

Chromosomal fragments can result from direct DNA breakage, conversion of SSBs into DSBs during replication, or inhibition of DNA synthesis [26]. Whole chromosomes are generally formed from defects in the chromosome segregation machinery, such as deficiencies in the cell cycle checkpoint genes, dysfunction of the mitotic spindle, defects in kinetochore or other parts of the mitotic activity [32]. MN formation can also be related to damage to chromosomal sub-structures, hypomethylation of centromeric and pericentromeric DNA or mechanical disruption [26, 33]. Hence, MN may represent a measure of both chromosome

breakage and chromosome loss, so an increased frequency of micronucleated cells can reflect exposures to clastogenic (e.g., structural CAs) or aneugenic (e.g., numerical CAs) agents.

The MN assay can also provide parameters of genetic damage [e.g., nucleoplasmic bridge (NPB) and nuclear buds (NBUDs)] apart from the chromosome breakage and loss. DSBs misrepair may result in the development of a dicentric chromosome and an acentric fragment. Typically, at anaphase the dicentric chromosome forms a NPB between the two daughter nuclei, when its centromeres are pulled to the opposite poles of the cell, and the acentric fragment results in a micronuclei [32, 33]. NBUDs are characterized by having the same morphology as a micronuclei, but they are attached to the nucleus by a narrow or wide nucleoplasmic connection [33]. The nuclear budding takes place during S-phase of the cell cycle and cells are sometimes referred as “broken egg” cells. Nuclear budding seems to be the mechanism by which cells remove amplified and/or excess DNA, being NBUDs considered a biomarker of gene amplification and/or altered gene dosage [32].

Once formed, a micronucleated cell can be eliminated by apoptosis, the MN can be expelled from the cell, retained in the cytoplasm or incorporated into the main nucleus. However, the post-mitotic fate of the MN and the micronucleated cell is still poorly understood [20].

Currently, MN are extensively used as genotoxicity biomarker and has found a place in human biomonitoring studies, being one of the best validated techniques to evaluate chromosomal damage in humans [34]. In fact, MN assay offers more advantages than SCE and CA assays - MN scoring is simpler, requires shorter training and is less time consuming.

MN frequency can be assessed in exfoliated epithelial cells (buccal, nasal mucosa, urine), erythrocytes [OECD TG 474 [35]] or in lymphocytes. The *in vitro* mammalian cell MN test protocol for testing chemicals, encompassing cultured primary human or other mammalian PBLs, was adopted in 2010 by the OECD (OECD TG 487) [36]. In human population studies, the PBLs are the most commonly used tissue and the most frequently used test the cytokinesis-block micronucleus (CBMN) assay. This assay has been extensively validated worldwide for assessing environmental effects on chromosome damage in blood and epithelial tissues in human populations. The MN frequencies can be influenced by age, sex, nutritional status or smoking habits [37, 38]. A schematic representation of the CBMN assay is presented in Figure 3.

Like the CA assay, CBMN assay is performed in cultured PBLs that have divided once and were stimulated to proliferate in the presence of a mitogen (PHA). MN can only be expressed in cells that have completed one nuclear division, and for that reason is necessary to differentiate these cells from resting cells to assess the cytogenetic damage [23]. Cytochalasin-B (actin-polymerization inhibitor) is added before the first mitosis to block the cytokinesis allowing the identification of the once-divided cells in culture (binucleated cells) and the cells that did not divide or escaped the blocking of cytokinesis (mononucleated cells). Of note that only the cells that completed one nuclear division during *in vitro* culture can express the damage inflicted *in vivo*. The presence of MN in binucleated cells corresponds to pre-existing *in vivo* chromosome damage that have accumulated in the DNA since the last replication and lesions expressed as MN during the first *in vitro* mitosis. In mononucleated cells, MN are a result of genomic instability accumulated over many years in stem cells and circulating lymphocytes. The CBMN test also provides information on the nuclear division index by the registration of mono, bi, tri and tetranucleated cells, offering useful information regarding the cytostatic

potential of exposures by the identification of compounds that stimulate or delay cell division [39].

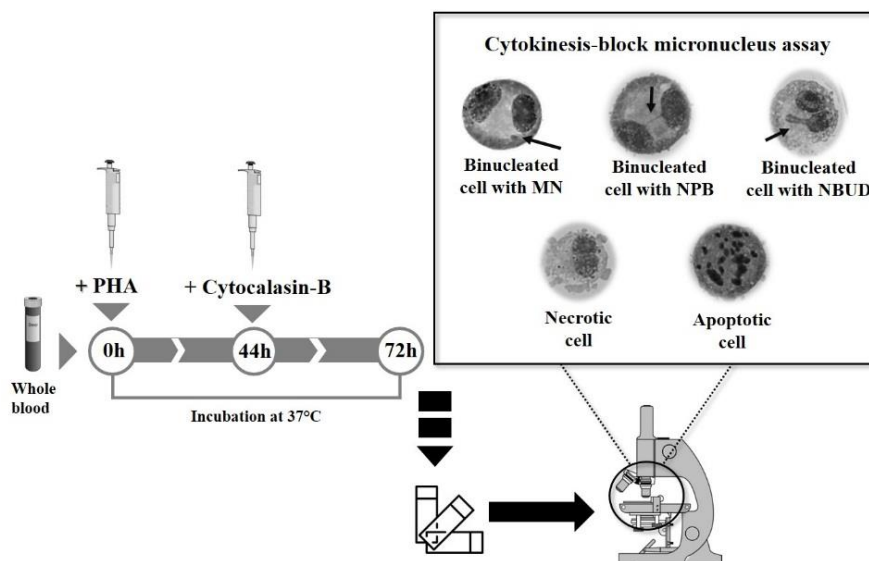


Figure 3. Schematic representation of CBMN assay described by Costa et al. 2006 [12] and examples of microscopical observations.

The combination of the CBMN assay with FISH probes contributes to the higher specificity and sensitivity of the method, since it allows the differentiation of MN containing a whole chromosome (centromere positive MN), resulting from the action of aneugenic agents, or an acentric chromosome fragment (centromere negative MN), reflecting the exposure to clastogenic agents.

In conclusion, besides aneugenic and clastogenic events this assay can provide information on several genotoxic and cytotoxic markers, such as, NPB (marker of chromosome rearrangement), NBUDs (marker of gene amplification) and nuclear division index (estimation of cell division inhibition).

Sister Chromatid Exchanges (SCEs): Mechanisms and Methodology

Sister chromatid exchanges (SCEs) are the result of interchanges of DNA between two sister chromatids in a duplicating chromosome. These reciprocal DNA exchanges take place during the S-phase of the cell cycle, when DNA is duplicated, resulting in two identical daughter chromatids. The effective breakage of DNA in sister chromatids followed by the reunion of the chromatids with one another, apparently at the same location, results in the SCEs [40]. Even though the mechanism of formation of SCE is not completely understood, they seem to be a consequence of the replication of a damaged DNA template, possibly at the replication fork [26]. The likely pathway for SCEs formation involves the initial collapse of a replication fork when facing a previously existent gap or SSB in one parental strand [41]. Subsequently, the formation of a replication-associated DSB with one free end takes place, initiating the repair process by conservative HR with the invasion of the intact sister strand to serve as a template

for DNA synthesis. Resolution of the resulting Holliday junction by non-crossing over will follow, triggering the induction of SCEs [42] (Figure 4). Another possible mechanism for SCEs formation is the initial stall of the replication fork due to the existence of obstacles, namely adducts, on the DNA template. The cleavage of these intermediary structures by endonucleases will induce DSBs with one free end that can be further repaired by conservative HR as previously described, resulting in SCEs [42].

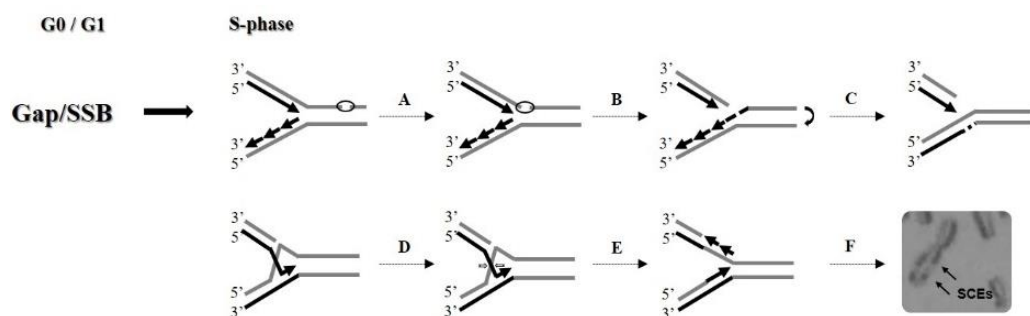


Figure 4. Mechanism of SCEs formation. A) Replication fork approaches a SSB. B) Collapsed replication fork; fork breaks; repair synthesis. C) 5' to 3' double-strand break resection. D) HR pathway, strand invasion. E) Resolution of the Holliday junction in the orientation shown by the arrows results in SCE, as illustrated by the grey/black color junctions in the new "parental" strands. F) The replication fork is restored (taken from Costa et al. 2013 [43]).

Taylor and coworkers were the first observing SCEs with the use of autoradiography on cells that underwent a cycle of triitated thymidine incorporation followed by a replication cycle in nonradioactive medium [44]. Nowadays, 5-bromo-2'-deoxyuridine (BrdU), an analogue of DNA thymidine, is used for a better visualization of the SCEs. A standard fluorescence assay is used based on differential staining of the sister chromatids after inducing cell division in the presence of BrdU [45]. BrdU is efficiently incorporated into the DNA during replication due to the closer resembles with thymidine. This technique takes advantage of the semiconservative nature of the DNA replication and when cells are cultured through a single replication cycle with the presence of BrdU in the medium, one DNA strand in each daughter chromatid is substituted by this thymidine analogue. After the second replication round the two sister chromatids differ in BrdU presence: one chromatid contains one substituted DNA strand, while both strands of its sister chromatid are substituted. The chromatids can be further differentiated with Hoechst dye, that fluoresces at lower intensity when bound to DNA substituted with BrdU than when bound to unsubstituted DNA [46]. Photosensitization with UV light is conducted, resulting in the selective degradation of the highly substituted chromatid, since it seems to preferentially break the bonds present in non-histone proteins of the chromatid containing more BrdU. Giemsa staining follows and the asymmetric distribution of BrdU-substituted DNA in sister chromatids can be observed by light microscopy, because brdU incorporation leads to a much weaker staining being the chromatid with more BrdU lighter in appearance [41]. SCE frequencies can afterwards be scored using a semi-automated computer-based metaphase finder. In Figure 5 a schematic representation of SCEs assay can be observed.

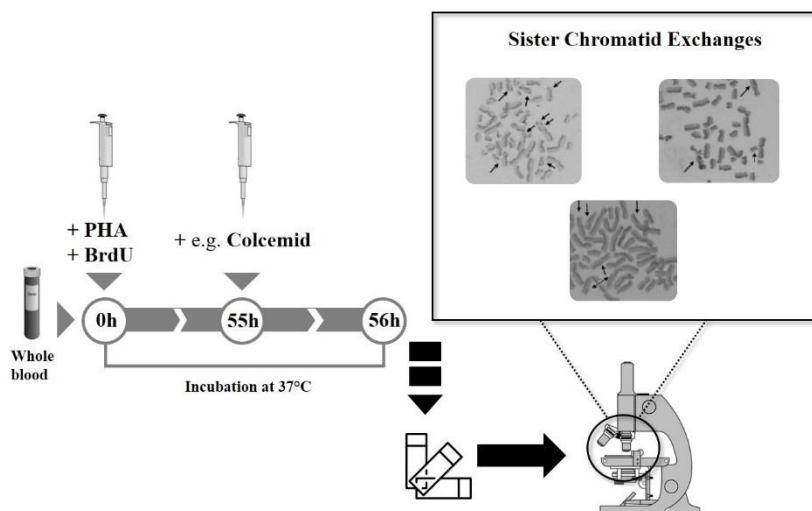


Figure 5. Schematic representation of SCEs assay described by Teixeira et al. 2004 [47] and examples of microscopical observations.

CAs, MN, SCEs AND RISK PREDICTIVITY OF HUMAN DISEASES

Environmental exposures related to our lifestyle, infections and the occupation that we have throughout our life are considered a major cause of cancer in humans. According to the World Health Organization (WHO), cancer is a leading cause of death worldwide, being responsible for 8.8 million deaths in 2015. Globally, one in six deaths is due to cancer. It has been estimated that 3 to 6% of all cancers in the world are resultant from occupational exposures. Additionally, at least 200 000 people die every year from occupational or work-related cancers [48].

Exposure assessment aims at prevention. Looking at these numbers, it is of extreme importance to identify hazardous environmental substances and predict the development of cancer. Thus, a great challenge in biomarker research is the identification of long-term risks of developing cancer and the identification of early markers of occupational and environmental toxicity. The use of cytogenetic markers in human biomonitoring is of paramount importance due to its predictability regarding deleterious effects in health related with exposure to environmental stressors. Regarding carcinogenicity, the most frequently used biomarkers reflect genotoxic effects. Accordingly, these cytogenetic endpoints have been applied either in the surveillance of human exposure or in the investigation of the genotoxic potential of new compounds by the U.S. National Toxicology Program (NTP) and the International Agency for Research on Cancer (IARC). Epidemiologic data linking high frequencies of CAs and MN in PBLs with cancer risk in human populations supports the relevance of cytogenetic alterations as cancer risk biomarkers [18, 19].

SCEs were efficiently induced by several mutagenic or carcinogenic agents, particularly those that result in covalent DNA adducts or interfere directly or indirectly with DNA replication [45]. This assay has also gained popularity when it was observed that the cells that derived from patients with Bloom's syndrome, characterized by a high propensity for cancer, exhibit a hallmark high occurrence of SCE [49] due to dysregulation of HR by a defective

helicase function [41]. Defects in HR process might be the cause of genetic instability, increased SCE and high incidence of cancer in early life associated with this disorder. Actually, increases in SCE frequency have been used as a biomarker in the clinical diagnose of Bloom's syndrome [41, 50].

SCEs is considered a good biomarker of effect to genotoxic/mutagenic agents, although no association between high SCEs frequencies and cancer risk was observed [18, 28, 51-53]. In the early 1990s, the preliminary reports from the Nordic Study Group on the Health Risk of Chromosome Damage mentioned the lack of association between the frequency of SCE and the incidence of cancer in humans [52, 54]. The results obtained in these initial studies were later confirmed in 1998 by an European cohort study [18] and in 2004 in a study from Bonassi and coworkers [53]. An association between SCEs level and cancer risk is difficult to evaluate because the baseline levels of SCEs are different among individuals and across studies [51]. Consequently, the use of SCE as biomarker for human biomonitoring has been declining, and CAs and MN are the preferred methods nowadays.

As previously mentioned, CAs have been associated with cancer risk prediction, being the most extensively used biomarker in human populations exposed to genotoxic agents [26]. The first reports regarding possible associations between the frequency of CAs and the incidence of cancer in humans emerged in the early 1990s, as for SCEs. Several studies with Italian and Nordic cohorts reported that the CAs levels in lymphocytes revealed to be predictive of overall cancer risk in humans [18, 52, 55, 56]. In fact, one of the latest studies mentioned, a report from 1998 of the European Study Group on Cytogenetic Biomarkers and Health, had the following results using both cohorts that comprised 3541 subjects examined between 1970 and 1988: subjects with higher levels of CAs presented an elevated standardized incidence ratio for cancer of 1.53 for the Nordic cohort and a mortality ratio of 2.01 for the Italian cohort demonstrating an increased cancer incidence in healthy individuals [18]. Another report for the referred European Group using both cohorts confirmed that the increased risk of cancer was related to high levels of CAs in lymphocytes, but independent of the exposure to carcinogenic agents [57]. Although, a subsequent study from Smerhovsky and coworkers that analyzed data from 3973 subjects reported a strong association between the frequency of CAs and the incidence of cancer in workers exposed to radon – an increase of 1% in the frequency of CAs was associated with an increase in cancer risk of 64% [58]. Nevertheless, the data collected in this study could not assess this association with exposure to other chemicals. Later, other study using the Nordic and Italian cohorts concluded that both chromosome-type aberrations and chromatid-type aberrations have the similar cancer risk predictive values [27], however some studies showed stronger association of chromosome-type aberrations with cancer risk predictivity [51]. A later study from 2007 of Boffetta and coworkers in a cohort with total of 6430 subjects from 9 laboratories in central Europe confirmed these previous results, observing relative risks of cancer of 1.78 and 1.81 for the medium and high tertiles levels of CAs [59]. Also, this study assessed the relation between the frequency of CAs and the risk of specific cancers and showed a relation between CAs and stomach cancer and possibly the colon and rectum cancers, regarding the lung and breast cancers the results were only suggestive of an association. A study from Bonassi and coworkers helped to clarify some of the referred issues, using a larger pooled analysis with 22 358 cancer-free subjects that underwent genetic screening and with follow-up time for cancer incidence of 10.1 years [60]. The following results were observed: an increased relative risk of cancer was found for high levels of CAs, this increase was mostly driven by chromosome-type aberrations, the strongest association was found for the stomach cancer and

the effect of CA levels on overall cancer risk was not altered by exposure. These results were in general consistent with previous studies. A recent study in patients diagnosed with colorectal, lung and breast cancer also reported that CAs serve as predictive markers of cancer but in this particular study for the colorectal cancer association, only chromatid-type aberrations were significantly elevated [61]. The results obtained in all studies highlight the importance of CAs as biomarkers of cancer risk prediction in humans.

MN can be a result of chromatid-type aberrations occurring during the replication of DNA on a damaged template or chromosome-type aberrations initiated before mitosis and duplicated at replication [26]. Hence, taking into consideration the cancer risk predictivity of CAs and the mechanistic similarities between CAs and MN formation, it was expected that MN would also be possible biomarkers of cancer predictivity [20]. In the 1990s the first studies performed in this regard found no relation between MN frequency in PBLs and the cancer risk in humans [18, 52, 56]. Nevertheless, later in 2007 Bonassi and coworkers reported that MN formation is associated with early events in carcinogenesis [19]. The results emerged from a cohort of 6718 individuals that were screened for MN frequency between 1980 and 2002 and a significant increase of all cancers incidence was found in otherwise healthy subjects with an elevated frequency of MN in PBLs. The link between MN induction in PBLs and cancer development was further confirmed by case control and meta-analysis studies [62, 63]. A significantly higher MN frequency was observed in lymphocytes of individuals who developed cancer within 14 years after blood sampling (cases; 4.7 ± 3.4 MN/1000 BN cells) as compared to those who were still cancer free at the end of the follow-up period (controls; 1.5 ± 1.7 ; $p < 0.0001$) [62]. Additionally, subjects with a high MN frequency (>2.5 MN/1000 BN cells) had an increased risk of cancer death when compared with those with low frequency of MN (≤ 2.5 MN/1000 BN cells). A meta-analysis with findings from 37 publications about the association between MN frequency in PBLs and the presence of cancer diagnosis revealed a 28 to 64% increase in the baseline MN frequency of untreated cancer patients when compared to cancer-free referents [63]. Other recent studies using the CBMN assay reported an association between the increased MN frequencies in lymphocytes and the risk of colorectal and bladder cancers [64, 65]. A study from Maffei and coworkers found an increase of MN frequency in PBLs of patients with colorectal cancer (16.82 ± 6.56) when compared to controls (8.00 ± 1.77) [64]. This report also mentioned that the MN frequency found for the control group was in accordance with the MN frequency of healthy controls, adjusted for age and gender, of their laboratory database (7.54 ± 1.74). Recently, an increased frequency of MN (9.51 ± 4.73 versus 7.73 ± 3.91) and NBUDs (5.06 ± 3.36 versus 3.99 ± 2.44) was observed in bladder cancer patients compared to controls [65]. In addition to the cancer predictivity of MN, other biomarkers of the CBMN assay (i.e., NPB and NBUDs) have been reported as stronger predictors of cancer risk [65-67]. All the recently accumulated data on the cancer predictive value of elevated MN frequencies makes the CBMN assay a good candidate for wide application in human biomonitoring studies.

In Figure 6 a chronological representation of the major findings regarding the aforementioned cytogenetic biomarkers and the cancer risk predictivity in humans can be observed.

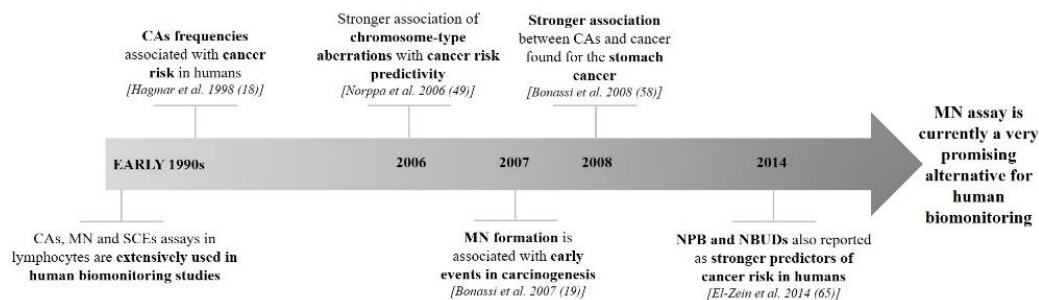


Figure 6. Chronological representation of the major findings regarding cytogenetic biomarkers and cancer risk predictivity.

APPLICATION IN OTHER HUMAN CELLS

Other cell types widely used to assess the formation of MN in human biomonitoring studies are the exfoliated cells, such as buccal, nasal or urothelial cells. These non-blood cells have some advantages when compared with lymphocytes for the use in large human biomonitoring studies: are easier to collect by non-invasive procedures; and, in some cases, these tissues are the actual target of the carcinogenic agents of exposure. For instance, they are in immediate contact with genotoxic agents by inhalation and ingestion, also kidney and bladder cells are in contact with the metabolites of the agents of exposure [68].

An advantage for the use of these types of cells is the elimination of the culture step, since these are usually rapidly dividing tissues. Cells are collected (swabbed in the case of buccal and nasal cells and isolated from urine in the case of urothelial cells) and dropped or smeared into the slides and stained with the appropriate dye [69].

Increases in MN frequencies in exfoliated cells in response to specific exposures have been widely reported. Some examples are the increases in MN frequencies in buccal, nasal and urothelial cells of populations exposed to formaldehyde [70-74], arsenic [75-77] and smoking [78-82].

Apart from the use of these cells in occupational and environmental biomonitoring studies, they have also been related with risk predictivity of human diseases [83]. In 2000, Bloching and coworkers suggested that the use of MN assay in buccal mucosa could be important in the prediction of cancer risk in the upper aerodigestive tract [84]. Exfoliated buccal cells collected from breast and uterus of untreated cancer patients presented a significant increase in MN number when compared to control subjects [85]. A study performed in 45 cancer patients with lung, stomach or colorectal cancer found significant increases in micronucleate cells number in buccal mucosa of cancer patients when compared with healthy control subjects [86]. Patients with lung cancer presented the highest number of buccal cells with MN per 1000 cells [86]. A study from Bonassi and coworkers under the frame of the HUMAN MicroNucleus project on exfoliated buccal cells (HUMN_{XL}) was conducted combining data of buccal MN values of 5424 subjects from 30 laboratories with the aim of studying several conditions that could impact MN frequency [87]. Significant increases in MN frequency were observed for the majority of occupational exposures and also cancer diagnosis. Associations between elevated MN frequency was observed for oropharyngeal cancer, respiratory cancer and all the other cancers pooled together. Also, in this study no associations were found for elevated MN

frequency and neurodegenerative diseases [87]. A subsequent systematic review and meta-analysis from Bolognesi and coworkers focusing in the clinical application of the micronucleus test in exfoliated buccal cells suggests that the MN assay can be useful in the prescreening and follow-up of precancerous oral lesions [83]. Increased MN frequency was observed in oral and neck cancer patients and leukoplakia, a premalignant stage of cancer, when compared with controls. Besides, in this study, an increased MN frequency was observed for patients with Alzheimer's disease and Down syndrome [83].

The MN test in urothelial cells has also been used to address the possibility of the assay to be conducted as a diagnostic tool. An increased MN frequency in urothelial cells has been observed in cancer patients [88-90]. Arora and coworkers performed a study with 30 cases of urine samples, where 15 were reported as normal urothelial cells and the other 15 as atypical urothelial cells that later turn out as malignant. MN were only observed in atypical cells ($2.53 \pm 0.99\%$), whereas no MN was found in the control cases [88]. A study using MN test in urothelial cells was conducted in women diagnosed with cervix cancer. This study reported an elevated frequency of micronucleated cells in 72% of the cancer patients when compared to 16.7% of controls [89]. Another study observed an increased MN frequency in cervix cancer patients, and found a linear association between the mean MN count and the stage of cervix cancer [90].

These studies suggest that the MN test in exfoliated cells might be an important indicator for malignancy and other diseases. The use of this type of cells may be advantageous in screening programs for occupational and environmental exposures and also for other human diseases, especially given the non-invasive method for collecting the samples.

CONCLUSION

We are all exposed to several harmful agents in our daily life activities, being those exposures possibly responsible for health impairments in the future. Research focused on the development of integrated approaches regarding the health risk assessment of populations exposed to hazard compounds is of paramount importance. Thus, genotoxicity biomarkers used in environmental and occupational health have been increasingly applied in human population studies.

Cytogenetic biomarkers are the most frequently used endpoints given their sensitivity to measure exposures of genotoxic agents and their role as early cancer risk predictors. These biomarkers are usually measured in lymphocytes, as the cytogenetic damage present in PBLs reflects cumulative exposure events and possible health consequences, as carcinogenesis, which can be related to chronic genotoxic exposure. CAs are the most validated biomarker in this regard although time-consuming. On the other hand, MN assay is a very promising alternative that is more appropriate for human biomonitoring, since it offers an easier technical procedure than SCE and CA assays, and has also been widely associated with cancer risk predictivity in several studies.

In order to efficiently use these cytogenetic endpoints in human biomonitoring is important to be aware of the need for the biomarkers to be validated and to have good protocols and sufficient technical expertise in the laboratories that conduct the assays.

One of main issues concerning cytogenetic biomarkers is the interpretation of altered levels of these indicators at individual level. Traditionally, risk predictions are only valid at a group level and therefore effect of inter-individual variability is removed. Hence, it will be important to conduct further research regarding the alterations on these endpoints at individual level.

Cytogenetic biomarkers such as SCEs, CAs and MN are extensively used as early predictors of clinical disease that can contribute to the implementation of new effective disease prevention policies in occupational and environmental settings. Further human biomonitoring studies encompassing cohorts with larger populations should continue to be conducted with the aim of obtaining a possible cut-off/control level that might be indicative of cancer/exposure risk prediction.

ACKNOWLEDGMENTS

The microscopic photos were kindly provided by Solange Costa (for Figures 2, 4 and 5) and Fátima Brandão (for Figure 3). Figure 4 was kindly provided by Solange Costa. The work developed by Armanda Teixeira-Gomes and Solange Costa is supported by the Portuguese Foundation for Science and Technology (FCT) under the grants SFRH/BD/121802/2016 and SFRH/BPD/100948/2014, respectively.

REFERENCES

- [1] IPCS. *Environmental Health Criteria 155: Biomarkers and risk assessment: Concepts and Principles*. World Health Organization, Geneva, Switzerland., 1993.
- [2] WHO. *Human biomonitoring: facts and figures*. World Health Organization Regional Office for Europe, Copenhagen, 2015.
- [3] Angerer, J; Ewers, U; Wilhelm, M. Human biomonitoring: state of the art. *Int J Hyg Environ Health*, 2007, 210(3-4), 201-28.
- [4] IPCS. *Concise international chemical assessment document (CICAD)*; no. 40 - Formaldehyde. World Health Organization, Geneva, Switzerland., 2002.
- [5] Klaassen, CD; Amdur, MO. *Casarett and Doull's toxicology: the basic science of poisons*. McGraw-Hill New York, 1996.
- [6] Silins, I; Hogberg, J. Combined toxic exposures and human health: biomarkers of exposure and effect. *Int J Environ Res Public Health*, 2011, 8(3), 629-47.
- [7] Links, JM; Kensler, TW; Groopman, JD. Biomarkers and mechanistic approaches in environmental epidemiology. *Annu Rev Public Health*, 1995, 16, 83-103.
- [8] Kyrtopoulos, SA. Biomarkers in environmental carcinogenesis research: striving for a new momentum. *Toxicol Lett.*, 2006, 162(1), 3-15.
- [9] Ladeira, C; Viegas, S. Human Biomonitoring—An overview on biomarkers and their application in Occupational and Environmental Health. *Biomonitoring*, 2016, 3(1).

- [10] Costa, S; Carvalho, S; Costa, C; Coelho, P; Silva, S; Santos, LS; et al. Increased levels of chromosomal aberrations and DNA damage in a group of workers exposed to formaldehyde. *Mutagenesis.*, 2015, 30(4), 463-73.
- [11] Costa, S; Ceppi, M; Costa, C; Silva, S; Pereira, C; Laffon, B; et al. The cytokinesis-block micronucleus (CBMN) assay in human populations exposed to styrene: A systematic review and meta-analysis. *Mutat Res.*, 2016, 770(Pt A), 92-105.
- [12] Costa, S; Coelho, P; Costa, C; Silva, S; Mayan, O; Santos, LS; et al. Genotoxic damage in pathology anatomy laboratory workers exposed to formaldehyde. *Toxicology*, 2008, 252(1-3), 40-8.
- [13] Costa, S; Pina, C; Coelho, P; Costa, C; Silva, S; Porto, B; et al. Occupational exposure to formaldehyde: genotoxic risk evaluation by comet assay and micronucleus test using human peripheral lymphocytes. *J Toxicol Environ Health A.*, 2011, 74(15-16), 1040-51.
- [14] Costa, S; Garcia-Leston, J; Coelho, M; Coelho, P; Costa, C; Silva, S; et al. Cytogenetic and immunological effects associated with occupational formaldehyde exposure. *J Toxicol Environ Health A.*, 2013, 76(4-5), 217-29.
- [15] Teixeira, JP; Gaspar, J; Coelho, P; Costa, C; Pinho-Silva, S; Costa, S; et al. Cytogenetic and DNA damage on workers exposed to styrene. *Mutagenesis.*, 2010, 25(6), 617-21.
- [16] Costa, C; Silva, S; Neves, J; Coelho, P; Costa, S; Laffon, B; et al. Micronucleus frequencies in lymphocytes and reticulocytes in a pesticide-exposed population in Portugal. *J Toxicol Environ Health A.*, 2011, 74(15-16), 960-70.
- [17] Costa C; Costa S; Silva S; Coelho P; Botelho M; Gaspar J; et al. DNA damage and susceptibility assessment in industrial workers exposed to styrene. *J Toxicol Environ Health A.* 2012, 75(13-15), 735-46.
- [18] Hagmar, L; Bonassi, S; Stromberg, U; Brogger, A; Knudsen, LE; Norppa, H; et al. Chromosomal aberrations in lymphocytes predict human cancer: a report from the European Study Group on Cytogenetic Biomarkers and Health (ESCH). *Cancer Res.*, 1998, 58(18), 4117-21.
- [19] Bonassi, S; Znaor, A; Ceppi, M; Lando, C; Chang, WP; Holland, N; et al. An increased micronucleus frequency in peripheral blood lymphocytes predicts the risk of cancer in humans. *Carcinogenesis.*, 2007, 28(3), 625-31.
- [20] Mateuca, RA; Decordier, I; Kirsch-Volders, M. Cytogenetic methods in human biomonitoring: principles and uses. *Methods Mol Biol.*, 2012, 817, 305-34.
- [21] Salama, SA; Serrana, M; Au WW. Biomonitoring using accessible human cells for exposure and health risk assessment. *Mutat Res.*, 1999, 436(1), 99-112.
- [22] Au, WW. Usefulness of biomarkers in population studies: from exposure to susceptibility and to prediction of cancer. *Int J Hyg Environ Health.*, 2007, 210(3-4), 239-46.
- [23] Mateuca, R; Lombaert, N; Aka, PV; Decordier, I; Kirsch-Volders, M. Chromosomal changes: induction, detection methods and applicability in human biomonitoring. *Biochimie.*, 2006, 88(11), 1515-31.
- [24] Suspiro, A; Prista, J. Biomarkers of occupational exposure do anticancer agents: a minireview. *Toxicol Lett.*, 2011, 207(1), 42-52.

- [25] Obe, G; Pfeiffer, P; Savage, JR; Johannes, C; Goedecke, W; Jeppesen, P; et al. Chromosomal aberrations: formation, identification and distribution. *Mutat Res.*, 2002, 504(1-2), 17-36.
- [26] Albertini, RJ; Anderson, D; Douglas, GR; Hagmar, L; Hemminki, K; Merlo, F; et al. IPCS guidelines for the monitoring of genotoxic effects of carcinogens in humans. International Programme on Chemical Safety. *Mutat Res.*, 2000, 463(2), 111-72.
- [27] Hagmar, L; Stromberg, U; Bonassi, S; Hansteen, IL; Knudsen, LE; Lindholm, C; et al. Impact of types of lymphocyte chromosomal aberrations on human cancer risk: results from Nordic and Italian cohorts. *Cancer Res.*, 2004, 64(6), 2258-63.
- [28] Bonassi, S; Au, WW. Biomarkers in molecular epidemiology studies for health risk prediction. *Mutat Res.*, 2002, 511(1), 73-86.
- [29] OECD. Test No. 473: *In Vitro Mammalian Chromosomal Aberration Test*: OECD Publishing.
- [30] Bourner, RD; Parry, EM; Parry, JM. Chemically induced aneuploidy: investigations into chromosome specific effects in mitosis. *Mutat Res.*, 1998, 404(1-2), 191-7.
- [31] Roma-Torres, J; Teixeira, JP; Silva, S; Laffon, B; Cunha, LM; Mendez, J; et al. Evaluation of genotoxicity in a group of workers from a petroleum refinery aromatics plant. *Mutat Res.*, 2006, 604(1-2), 19-27.
- [32] Fenech, M; Kirsch-Volders, M; Natarajan, AT; Surralles, J; Crott, JW; Parry, J; et al. Molecular mechanisms of micronucleus, nucleoplasmic bridge and nuclear bud formation in mammalian and human cells. *Mutagenesis.*, 2011, 26(1), 125-32.
- [33] Fenech, M. Cytokinesis-block micronucleus assay evolves into a “cytome” assay of chromosomal instability, mitotic dysfunction and cell death. *Mutat Res.*, 2006, 600(1-2), 58-66.
- [34] Bolognesi, C; Fenech, M. Micronucleus assay in human cells: lymphocytes and buccal cells. *Methods Mol Biol.*, 2013, 1044, 191-207.
- [35] OECD. Test No. 474: *Mammalian Erythrocyte Micronucleus Test*: OECD Publishing.
- [36] OECD. Test No. 487: *In Vitro Mammalian Cell Micronucleus Test*: OECD Publishing.
- [37] Fenech, M. Important variables that influence base-line micronucleus frequency in cytokinesis-blocked lymphocytes-a biomarker for DNA damage in human populations. *Mutat Res.*, 1998, 404(1-2), 155-65.
- [38] Bonassi, S; Fenech, M; Lando, C; Lin, YP; Ceppi, M; Chang, WP; et al. HUMAN MicroNucleus project: international database comparison for results with the cytokinesis-block micronucleus assay in human lymphocytes: I. Effect of laboratory protocol, scoring criteria, and host factors on the frequency of micronuclei. *Environ Mol Mutagen.*, 2001, 37(1), 31-45.
- [39] Fenech, M. The advantages and disadvantages of the cytokinesis-block micronucleus method. *Mutat Res.*, 1997, 392(1-2), 11-8.
- [40] Latt, SA; Schreck, RR. Sister chromatid exchange analysis. *Am J Hum Genet.*, 1980, 32(3), 297-313.
- [41] Wilson, DM; 3rd. Thompson, LH. Molecular mechanisms of sister-chromatid exchange. *Mutat Res.*, 2007, 616(1-2), 11-23.

- [42] Helleday, T. Pathways for mitotic homologous recombination in mammalian cells. *Mutat Res.*, 2003, 532(1-2), 103-15.
- [43] da Costa, SCB. *Occupational exposure to formaldehyde-Genotoxic damage and susceptibility evaluation in anatomical pathology laboratory workers*: Universidade do Porto (Portugal), 2013.
- [44] Taylor, JH; Woods, PS; Hughes, WL. The organization and duplication of chromosomes as revealed by autoradiographoc studies using tritium-labeled thymidine. *Proc Natl Acad Sci U S A.*, 1957, 43(1), 122-8.
- [45] Natarajan, A; Tucker, J; Sasaki, M. Monitoring cytogenetic damage *in vivo*. In: *SCOPE 52 Methods to Assess DNA Damage and Repair: Interspecies Comparisons Scientific committee on problems of the environment international council of scientific unions*, 1994, 52, 95-.
- [46] Perry, P; Wolff, S. New Giemsa method for the differential staining of sister chromatids. *Nature.*, 1974, 251, 156.
- [47] Teixeira, JP; Gaspar, J; Silva, S; Torres, J; Silva, SN; Azevedo, MC; et al. Occupational exposure to styrene: modulation of cytogenetic damage and levels of urinary metabolites of styrene by polymorphisms in genes CYP2E1, EPHX1, GSTM1, GSTT1 and GSTP1. *Toxicology.*, 2004, 195(2-3), 231-42.
- [48] WHO. World Health Organization: *Cancer*. <http://www.who.int/cancer/en/>, 2017.
- [49] Chaganti, RS; Schonberg, S; German, J. A manyfold increase in sister chromatid exchanges in Bloom's syndrome lymphocytes. *Proc Natl Acad Sci U S A.*, 1974, 71(11), 4508-12.
- [50] Ellis, NA; Proytcheva, M; Sanz, MM; Ye, TZ; German, J. Transfection of BLM into cultured bloom syndrome cells reduces the sister-chromatid exchange rate toward normal. *Am J Hum Genet.*, 1999, 65(5), 1368-74.
- [51] Norppa, H; Bonassi, S; Hansteen, IL; Hagmar, L; Stromberg, U; Rossner, P; et al. Chromosomal aberrations and SCEs as biomarkers of cancer risk. *Mutat Res.*, 2006, 600(1-2), 37-45.
- [52] Hagmar, L; Brogger, A; Hansteen, IL; Heim, S; Hogstedt, B; Knudsen, L; et al. Cancer risk in humans predicted by increased levels of chromosomal aberrations in lymphocytes: Nordic study group on the health risk of chromosome damage. *Cancer Res.*, 1994, 54(11), 2919-22.
- [53] Bonassi, S; Lando, C; Ceppi, M; Landi, S; Rossi, AM; Barale, R. No association between increased levels of high-frequency sister chromatid exchange cells (HFCs) and the risk of cancer in healthy individuals. *Environ Mol Mutagen.*, 2004, 43(2), 134-6.
- [54] Brogger, A; Hagmar, L; Hansteen, IL; Heim, S; Hogstedt, B; Knudsen, L; et al. An inter-Nordic prospective study on cytogenetic endpoints and cancer risk. Nordic Study Group on the Health Risk of Chromosome Damage. *Cancer Genet Cytogenet.*, 1990, 45(1), 85-92.
- [55] Bonassi, S; Abbondandolo, A; Camurri, L; Dal Pra, L; De Ferrari, M; Degrassi, F; et al. Are chromosome aberrations in circulating lymphocytes predictive of future cancer onset

- in humans? Preliminary results of an Italian cohort study. *Cancer Genet Cytogenet.*, 1995, 79(2), 133-5.
- [56] Hagmar, L; Bonassi, S; Stromberg, U; Mikoczy, Z; Lando, C; Hansteen, IL; et al. Cancer predictive value of cytogenetic markers used in occupational health surveillance programs: a report from an ongoing study by the European Study Group on Cytogenetic Biomarkers and Health. *Mutat Res.*, 1998, 405(2), 171-8.
- [57] Bonassi, S; Hagmar, L; Stromberg, U; Montagud, AH; Tinnerberg, H; Forni, A; et al. Chromosomal aberrations in lymphocytes predict human cancer independently of exposure to carcinogens. European Study Group on Cytogenetic Biomarkers and Health. *Cancer Res.*, 2000, 60(6), 1619-25.
- [58] Smerhovsky, Z; Landa, K; Rossner, P; Brabec, M; Zudova, Z; Hola, N; et al. Risk of cancer in an occupationally exposed cohort with increased level of chromosomal aberrations. *Environ Health Perspect.*, 2001, 109(1), 41-5.
- [59] Boffetta, P; van der Hel, O; Norppa, H; Fabianova, E; Fucic, A; Gundy, S; et al. Chromosomal aberrations and cancer risk: results of a cohort study from Central Europe. *Am J Epidemiol.*, 2007, 165(1), 36-43.
- [60] Bonassi, S; Norppa, H; Ceppi, M; Stromberg, U; Vermeulen, R; Znaor, A; et al. Chromosomal aberration frequency in lymphocytes predicts the risk of cancer: results from a pooled cohort study of 22 358 subjects in 11 countries. *Carcinogenesis.*, 2008, 29(6), 1178-83.
- [61] Vodenkova, S; Polivkova, Z; Musak, L; Smerhovsky, Z; Zoubkova, H; Sytarova, S; et al. Structural chromosomal aberrations as potential risk markers in incident cancer patients. *Mutagenesis.*, 2015, 30(4), 557-63.
- [62] Murgia, E; Ballardini, M; Bonassi, S; Rossi, AM; Barale, R. Validation of micronuclei frequency in peripheral blood lymphocytes as early cancer risk biomarker in a nested case-control study. *Mutat Res.*, 2008, 639(1-2), 27-34.
- [63] Iarmarcovai, G; Ceppi, M; Botta, A; Orsiere, T; Bonassi, S. Micronuclei frequency in peripheral blood lymphocytes of cancer patients: a meta-analysis. *Mutat Res.*, 2008, 659(3), 274-83.
- [64] Maffei, F; Zolezzi Moraga, JM; Angelini, S; Zenesini, C; Musti, M; Festi, D; et al. Micronucleus frequency in human peripheral blood lymphocytes as a biomarker for the early detection of colorectal cancer risk. *Mutagenesis.*, 2014, 29(3), 221-5.
- [65] Pardini, B; Viberti, C; Naccarati, A; Allione, A; Oderda, M; Critelli, R; et al. Increased micronucleus frequency in peripheral blood lymphocytes predicts the risk of bladder cancer. *Br J Cancer.*, 2017, 116(2), 202-10.
- [66] El-Zein, RA; Schabath, MB; Etzel, CJ; Lopez, MS; Franklin, JD; Spitz, MR. Cytokinesis-blocked micronucleus assay as a novel biomarker for lung cancer risk. *Cancer Res.*, 2006, 66(12), 6449-56.
- [67] El-Zein, RA; Lopez, MS; D'Amelio, AM; Jr. Liu, M; Munden, RF; Christiani, D; et al. The cytokinesis-blocked micronucleus assay as a strong predictor of lung cancer: extension of a lung cancer risk prediction model. *Cancer epidemiology, biomarkers &*

- prevention: a publication of the American Association for Cancer Research, cosponsored by the American Society of Preventive Oncology.*, 2014, 23(11), 2462-70.
- [68] Fenech, M; Holland, N; Chang, WP; Zeiger, E; Bonassi, S. The HUman MicroNucleus Project--An international collaborative study on the use of the micronucleus technique for measuring DNA damage in humans. *Mutat Res.*, 1999, 428(1-2), 271-83.
- [69] Stich, HF; San, RH; Rosin, MP. Adaptation of the DNA-repair and micronucleus tests to human cell suspensions and exfoliated cells. *Ann N Y Acad Sci.*, 1983, 407, 93-105.
- [70] Suruda, A; Schulte, P; Boeniger, M; Hayes, RB; Livingston, GK; Steenland, K; et al. Cytogenetic effects of formaldehyde exposure in students of mortuary science. *Cancer Epidemiol Biomarkers Prev.*, 1993, 2(5), 453-60.
- [71] Viegas, S; Ladeira, C; Nunes, C; Malta-Vacas, J; Gomes, M; Brito, M; et al. Genotoxic effects in occupational exposure to formaldehyde: A study in anatomy and pathology laboratories and formaldehyde-resins production. *J Occup Med Toxicol.*, 2010, 5(1), 25.
- [72] Costa, S; Brandão, F; Coelho, M; Costa, C; Coelho, P; Silva, S; et al. Micronucleus frequencies in lymphocytes and buccal cells in formaldehyde exposed workers. *WIT Transactions on Biomedicine and Health.*, 2013, 16, 83-94.
- [73] Ying, CJ; Yan, WS; Zhao, MY; Ye, XL; Xie, H; Yin, SY; et al. Micronuclei in nasal mucosa, oral mucosa and lymphocytes in students exposed to formaldehyde vapor in anatomy class. *Biomed Environ Sci.*, 1997, 10(4), 451-5.
- [74] Ballarin, C; Sarto, F; Giacomelli, L; Bartolucci, GB; Clonfero, E. Micronucleated cells in nasal mucosa of formaldehyde-exposed workers. *Mutat Res.*, 1992, 280(1), 1-7.
- [75] Gonsebatt, ME; Vega, L; Salazar, AM; Montero, R; Guzman, P; Blas, J; et al. Cytogenetic effects in human exposure to arsenic. *Mutat Res.*, 1997, 386(3), 219-28.
- [76] Warner, ML; Moore, LE; Smith, MT; Kalman, DA; Fanning, E; Smith, AH. Increased micronuclei in exfoliated bladder cells of individuals who chronically ingest arsenic-contaminated water in Nevada. *Cancer Epidemiol Biomarkers Prev.*, 1994, 3(7), 583-90.
- [77] Moore, LE; Smith, AH; Hopenhayn-Rich, C; Biggs, ML; Kalman, DA; Smith, MT. Micronuclei in exfoliated bladder cells among individuals chronically exposed to arsenic in drinking water. *Cancer Epidemiol Biomarkers Prev.*, 1997, 6(1), 31-6.
- [78] Nersesyan, A; Muradyan, R; Kundi, M; Knasmueller, S. Impact of smoking on the frequencies of micronuclei and other nuclear abnormalities in exfoliated oral cells: a comparative study with different cigarette types. *Mutagenesis.*, 2011, 26(2), 295-301.
- [79] Piyathilake, CJ; Macaluso, M; Hine, RJ; Vinter, DW; Richards, EW; Krumdieck, CL. Cigarette smoking, intracellular vitamin deficiency, and occurrence of micronuclei in epithelial cells of the buccal mucosa. *Cancer Epidemiol Biomarkers Prev.*, 1995, 4(7), 751-8.
- [80] Burgaz, S; Iscan, A; Buyukbingol, ZK; Bozkurt, A; Karakaya, AE. Evaluation of micronuclei in exfoliated urothelial cells and urinary thioether excretion of smokers. *Mutat Res.*, 1995, 335(2), 163-9.
- [81] Lehucher-Michel, MP; Di Giorgio, C; Amara, YA; Laget, M; Botta, A. The micronucleus assay in human exfoliated urothelial cells, effect of smoking. *Mutagenesis.*, 1995, 10(4), 329-32.

- [82] Reali, D; Di Marino, F; Bahramandpour, S; Carducci, A; Barale, R; Loprieno, N. Micronuclei in exfoliated urothelial cells and urine mutagenicity in smokers. *Mutat Res.*, 1987, 192(2), 145-9.
- [83] Bolognesi, C; Bonassi, S; Knasmueller, S; Fenech, M; Bruzzone, M; Lando, C; et al. Clinical application of micronucleus test in exfoliated buccal cells: A systematic review and metanalysis. *Mutat Res Rev Mutat Res.*, 2015, 766, 20-31.
- [84] Bloching, M; Hofmann, A; Lautenschlager, C; Berghaus, A; Grummt, T. Exfoliative cytology of normal buccal mucosa to predict the relative risk of cancer in the upper aerodigestive tract using the MN-assay. *Oral Oncol.*, 2000, 36(6), 550-5.
- [85] Nersesyan, AK; Adamyan, RT. Micronuclei level in exfoliated buccal mucosa cells of patients with benign and malignant tumors of female reproductive organs and breast. *Tsitol Genet.*, 2004, 38(3), 72-5.
- [86] Yildirim, IH; Yesilada, E; Yologlu, S. Micronucleus frequency in peripheral blood lymphocytes and exfoliated buccal cells of untreated cancer patients. *Genetika.*, 2006, 42(5), 705-10.
- [87] Bonassi, S; Coskun, E; Ceppi, M; Lando, C; Bolognesi, C; Burgaz, S; et al. The HUman MicroNucleus project on eXfoLiatted buccal cells (HUMN(XL)), the role of life-style, host factors, occupational exposures, health status, and assay protocol. *Mutat Res.*, 2011, 728(3), 88-97.
- [88] Arora, SK; Dey, P; Saikia, UN. Micronucleus in atypical urothelial cells. *Diagn Cytopathol.*, 2010, 38(11), 811-3.
- [89] Gandhi, G; Sharma, P. The micronucleus test in urothelial cells of cervix cancer patients. *Indian J Hum Genet.*, 2002.
- [90] Sundararajan, SK; Natarajan, PS. Kanchana. Micronucleus Assay in Urothelial Cells in Cancer Cervix. *J Clin Diagn Res.*, 2017, 11(3), Xc01-xc3.

Chapter 56

COMPLEX KARYOTYPES AND THEIR DETECTION IN HEMATOLOGIC MALIGNANCIES: THE ONGOING ROLE FOR CLASSICAL CYTOGENETICS

Helena Urbankova*, PhD

Department of Hemato-Oncology, Faculty of Medicine and Dentistry, Palacky
University Olomouc and University Hospital Olomouc, Olomouc, Czech Republic

ABSTRACT

The great era for classical cytogenetics started sixty years ago with the description of the twenty-three pairs of human chromosomes (Tjio and Levan, 1956) and the discovery of the Philadelphia chromosome, the first known chromosomal defect associated with a specific type of cancer (Nowell and Hungerford, 1960). Since then many recurrent chromosomal aberrations linked to specific hematological malignancies have been detected. The vast majority of these abnormalities can be detected by modern molecular genetic methods so it might seem there is no longer a need for microscopy. However, there is still a significant portion of hemato-oncologic patients for which the classical cytogenetic investigation is appropriate, i.e., cases with complex karyotypes. Complex karyotypes are characterized by the presence of three or more unrelated chromosomal aberrations co-existing in a single clone. Their occurrence is associated with adverse outcomes across the entire spectrum of hematologic malignancies. These aberrations are also a powerful diagnostic indicator for molecular targeted therapies, allogeneic stem cell transplantation or other generally more aggressive treatment strategies. Since the presence of complex karyotypes would be missed when using only molecular genetic methods, this highlights the irreplaceable role of classical cytogenetics as a first-tier analysis for the evaluation of complex structural chromosomal abnormalities in hemato-oncologic patients. Ideally, classical cytogenetics is then followed by more precise molecular genetic methods to identify specific chromosomal aberrations more deeply. Here we focus on the complex

* Corresponding Author's Email: Helena.Urbankova@fnol.cz.

karyotype issues in the myelodysplastic diseases, leukemias, lymphomas and multiple myelomas as seen daily in our center.

Keywords: complex karyotype, cytogenetics, leukemia, lymphoma

INTRODUCTION

Genomic Instability and Chromosomal Aberrations

Genomic instability and the creation of chromosomal aberrations is a hallmark of many types of cancer, including hematologic malignancies. Such genetic instability results in the deregulation of basic cell functions such as proliferation, growth, cell death, replication, angiogenesis, etc. (Hanahan & Weinberg, 2011). The molecular cause of genomic instability in hematologic neoplasms remains unclear, but according to recent sequencing studies, the so called mutator hypothesis seems to be flawed. Instead the sequencing results suggest that the DNA repair genes are frequently free of mutations before therapy. More likely, the mutation status of several other genes appears to be important: the tumor suppressor gene *TP53*, the ataxia telangiectasia mutated (*ATM*) gene and the cyclin-dependent kinase inhibitor 2A gene (*CDKN2A*). Mutations in these genes play a key role in the oncogene-induced DNA replication stress model, the currently most trusted theory of oncogene-induced DNA instability (Gaillard, Garcia-Muse, & Aguilera, 2015; Negrini, Gorgoulis, & Halazonetis, 2010). It is believed that the activation of oncogenes causes DNA replication stress and formation of DNA double strand breaks (DSBs) in precancerous lesions. The continuing generation of DSBs leads to genomic instability and together with defects in the DNA repair and checkpoint genes (*ATM* and *TP53*) results in cancer (Bartkova et al., 2005; Halazonetis, Gorgoulis, & Bartek, 2008). Nevertheless, the precise nature of oncogene-induced replicative stress is still not clear.

In the last 5 years, the deep sequencing of the cancer genomes revealed novel catastrophic mechanisms as the cause of complex chromosomal changes, i.e., chromothripsis, kataegis and chromoplexy. One or more chromosomes carry multiple and adjacent intrachromosomal rearrangements. Such aberrant chromosomes underwent deep fragmentation and further resealing and are no longer regarded as an independent item. More probably this deep fragmentation occurred during a unique catastrophic event (Palumbo & Russo, 2016; Stephens et al., 2011). Interestingly, this hypothesis contrasts starkly with previously described models of carcinogenesis as the consecutive accumulation of aberrations. Evidence for chromothripsis is found in different types of cancer, e.g., multiple myeloma, medulloblastoma, neuroblastoma and colorectal cancers. Moreover, in some studies, chromothripsis was associated with more aggressive types of cancer. This discrepancy needs to be further examined.

Generally, genomic instability generates point mutations leading to large chromosomal rearrangements and possibly results in complex changes affecting many chromosomes. As tumorigenesis is still believed to be a multistep process, we can distinguish between primary

and secondary abnormalities, which are changing and accumulating during cancer development (Heim & Mitelman, 1989).

Primary aberrations are connected with the earliest stages of cancer and are mostly found as sole karyotypic changes associated with the already described molecular consequence, e.g., activation of a particular oncogene or inactivation of a specific tumor-suppressor gene. However sometimes primary aberrations can present simply as submicroscopic mutations and may not be visible in the karyotype.

Secondary aberrations develop in cells with existing primary changes during later stages of the disease, probably as the consequence of the altered cancer cell metabolism (e.g., oxidative stress, defect checkpoints, mutations in DNA repair genes, etc.). Secondary aberrations are generally more numerous and less specific, but we still see non-randomness in many of them. It remains to be determined if all of them are pathologically relevant and have functional impacts on the course of the disease.

There are specific changes with known impacts on the course of the disease, but there is also a wide spectrum of aberrations with unclear functions. Elucidating this issue would require both an overview of all chromosomal aberrations that occur in cancer genomes as well as their analyses in large cohorts of patients. Important information can be lost when using only molecular genetic analyses, which either focus on particular targets and can therefore miss aberrations in other locations, or do not distinguish between different clones in a population of tumor cells. It is still uncertain whether the proposed scenario of primary and secondary changes is correct. It is well known that a primary change detected in one type of neoplasm can also occur as a secondary change in a different neoplasm and have a different clinical impact. Moreover, there is one more criterion: the aberrations must be clonal, meaning that they must be detected in at least two identical cells and, in the case of a loss of genetic material, even three cells. The significance of multiple chromosomal changes remains unexplained. Despite of this it is now clear that patients diagnosed with a hematologic malignancy that have three or more aberrations in the same clone (the so called complex karyotypes) are associated with adverse outcomes.

There are two types of chromosomal aberrations - structural and numerical. They lead to balanced/unbalanced relocations of chromosomal material or result in a loss or gain of entire chromosomes or chromosomal segments. Based on the type of aberration present, different detection methods are available.

Methods of Detection of Chromosomal Aberrations

Conventional cytogenetics is based on the chromosomes present in the metaphase stage of the cell cycle, when chromatin is highly condensed and a typical chromosome morphology is visible. Therefore cells (mostly from the peripheral blood, bone marrow or other biologic material) are grown short term in a culturing medium and arrested in a metaphase using a mitotic inhibitor. The cell suspension is further harvested and fixed using Carnoy's solution. The most frequently used technique to stain chromosome bands in hemato-oncology is G-

banding. The cell suspension is pretreated with trypsin and stained with a Giemsa staining solution. Twenty-three chromosome pairs can then be identified in humans based on their size, position of a centromere and a specific bright and dark banding pattern. The assembled karyotype is described using the International System for Human Cytogenetic Nomenclature (ISCN) 2016 (McGowan-Jordan J., Simons A., & Schmid M., 2016) and the detected abnormalities are defined. The advantage of this technique is that it allows the identification of balanced/unbalanced aberrations in a single cell at a low cost. Unfortunately, it has a relatively low resolution depending on the number and type of dividing cells and the quality of metaphase chromosomes, which can allow smaller aberrations to escape observation. Therefore, molecular cytogenetic methods based on complementary hybridization of a fluorescent labeled probe to patient's DNA were introduced.

These techniques brought a rapid improvement to the detection of chromosomal rearrangements. Fluorescence *in situ* hybridization (FISH) is the most frequently used technique. It was developed in 1982 by Langer-Safer et al., (Langer-Safer, Levine, & Ward, 1982) and its use for chromosome classification was described later in 1986 by Pinkel et al., (Pinkel, Straume, & Gray, 1986). FISH is able to detect and localize the presence or absence of a specific DNA sequence either on a particular chromosome or in an interphase nuclei without the need of metaphase chromosomes.

A large number of fluorescent probes targeting various sequences of the genome are now commercially available. FISH probes can be divided into three categories: locus specific, centromeric and whole chromosome painting probes. The selection of the appropriate probe for the detection of a specific chromosomal rearrangement associated with a particular type of hemato-oncologic disease depends on previous knowledge of the possible aberration. We can only routinely use a combination of two or three fluorochromes in a single hybridization and are therefore able to detect only a limited number of changes. Performing multiple hybridizations is time consuming and expensive. For these reasons using the FISH technique alone is not ideal for detecting multiple chromosomal aberrations and complex karyotypes. Similarly, the use of a singular molecular method such as allele-specific polymerase chain reaction (PCR), multiplex ligation-dependent probe amplification (MLPA), Sanger sequencing, etc., is also inadequate.

The multicolor FISH (M-FISH) approach is a much better tool. It was developed in the 1990s and published by Schröck et al., (Schröck et al., 1996) and Speicher et al., (Speicher, Gwyn Ballard, & Ward, 1996). This method is based on the complementary hybridization of probes to chromosomes in metaphase. The probes are labelled with a different fluorochrome combination specific to each chromosome. This technique allows the rapid and unequivocal detection of complex chromosomal rearrangements, whether they are balanced or unbalanced. Unfortunately the resolution of the M-FISH approach is also limited because of its correlation to chromosomes.

The introduction of a comparative genomic hybridization (CGH) was a key innovation. CGH is based on a competitive hybridization between the reference and tumor genomic DNA to the normal metaphase chromosomes. This approach could reveal unbalanced chromosomal aberrations by a single genome wide experiment (Kallioniemi et al., 1992). However, this

technique also has disadvantages: it's inability to detect balanced changes of genomes together with the relatively low resolution caused by the hybridization to chromosomes.

A few years later a variant of this technique was introduced: an arrayCGH (Pinkel et al., 1998; Solinas-Toldo et al., 1997). This method is also based on a competitive hybridization of the reference and tumor genomic DNA, but instead of binding to chromosomes it is applied to a chip with a spectrum of immobilized fragments of DNA. This enables the arrayCGH technique to detect genome wide unbalanced abnormalities with a high resolution defined by size and genome localization of the used fragments. ArrayCGH became a fast and reliable diagnostic tool for many cancer types, as well as hemato-oncologic diseases (Carter, Fiegler, & Piper, 2002; Fiegler et al., 2003; Vermeesch et al., 2007; Vermeesch et al., 2005). When combined with conventional cytogenetics and FISH, it provides a powerful methodology for the detection of complex chromosomal aberrations in hematologic malignancies (Knijnenburg et al., 2005; Kolialexi, Tsangaris, Kitsiou, Kanavakis, & Mavrou, 2005).

Next generation sequencing (NGS) strategies were developed recently (Sikkema-Raddatz et al., 2013). They are able to detect either balanced or unbalanced aberrations through the whole genome, but they have not added any major findings to the cytogenetic landscape of hematologic neoplasms. Moreover, they are quite expensive and require extensive biostatistical analysis. This is most likely the reason why these genome wide sequencing strategies are not currently widely used in diagnostic practice.

As it is well known that complex chromosomal changes impact prognosis in many hemato-oncologic diseases including the myelodysplastic syndrome (MDS), acute myeloid leukemia (AML), B-cell lymphoblastic leukemia/lymphoma (B-ALL), chronic myeloid leukemia (CML), chronic lymphocytic leukemia (CLL) and multiple myeloma (MM).

Unfortunately there is still a certain bias in defining a complex karyotype and its impact in various hematologic malignancies and there is a need for standardization (Peterson, 2017).

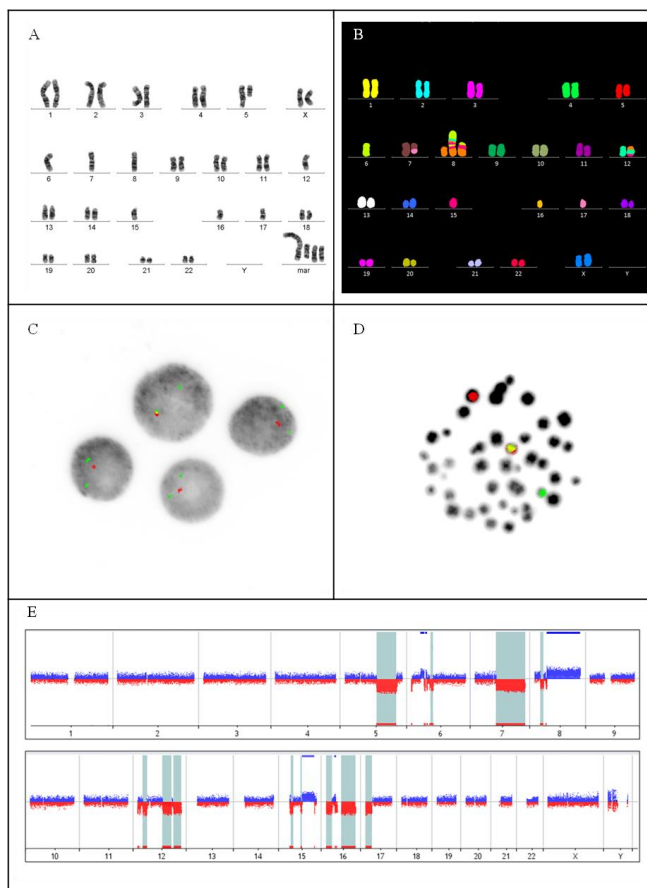
COMPLEX KARYOTYPE IN SELECTED HEMATOLOGIC MALIGNANCIES

Myelodysplastic Syndrome (MDS)

Myelodysplastic syndrome is a heterogeneous group of clonal bone marrow disorders characterized by impaired maturation of the hematopoietic cells associated with one or more peripheral blood cytopenias, which can also progress to acute myeloid leukemia (Vardiman, 2003). The current diagnostics follow the established WHO (World Health Organization) 2016 criteria based on histology, blast counts and cytogenetic findings (Arber et al., 2016). At the time of diagnosis, recurrent chromosomal aberrations are found in 40-70% of patients with primary MDS and these aberrations predict the rate of survival and the risk of leukemic transformation (Haase et al., 2007). The majority of these chromosomal changes are unbalanced (exhibiting loss of a part or an entire chromosome), but balanced translocations and complex

derivative chromosomes can also be detected. The most frequent cytogenetic aberrations include: del(5)(q) found in 10-20% of primary MDS, -7/del(7)(q) (10%), +8 (10%), del(20)(q) (5%) and others (Olney & Le Beau, 2001). All of these changes could be seen as solo aberrations or as a part of a complex karyotype, which has a big impact on the prognosis and the course of the disease. For example patients with isolated del(5)(q) have a good prognosis with a median survival of 4.8 years, whereas patients carrying the del(5)(q) as a part of a complex karyotype have a poor prognosis with early progression to AML. Complex karyotypes are observed in 10% of primary MDS and up to 90% of therapy-related MDS (Schanz et al., 2012). According to a Revised International Prognostic Scoring System for myelodysplastic syndromes (IPSS-R) (Greenberg et al., 2012), a complex karyotype in MDS is defined by the presence of three chromosomal changes in the poor prognostic group (with median survival 1.5 years), whereas the very poor prognostic group carry more abnormalities (median survival 0.7 years). Moreover according to the study of Göhring G. et al., (Gohring et al., 2010) aimed at the identification of prognostic factors in childhood myelodysplastic syndrome, it is important to define even the types of aberrations included in a complex karyotype. They defined the so called structurally complex karyotype, which is characterized by three or more chromosomal aberrations, at least one of which is structural. Patients with complex karyotypes but without structural rearrangement had a five-year overall survival similar to patients with normal karyotype, while patients with a structurally complex karyotype had a very poor outcome. Other authors studied the so called monosomal karyotype, which is defined by the presence of two or more autosomal monosomies or one monosomy with at least one structural abnormality. Such monosomal karyotypes seem to be associated with poor risk or very poor risk IPSS-R categories, but its prognostic value remains controversial. Other studies show that the impact of a complex karyotype as detected by conventional cytogenetics is increased by the presence of cytogenetically invisible mutations in the *TP53* gene, which can be detected by sequencing analyses. Patients with the *TP53* mutations and complex karyotype were associated with an extremely poor survival rate and a frequent early relapse, whereas outcomes were unambiguously better in the *TP53*-mutated patients without complex karyotype (Yoshizato et al., 2017).

In summary, not only is the presence of complex karyotype significant, but the number and the type of changes are also important considerations for establishing the diagnosis, prognosis and therapeutic strategies in patients with MDS. For these reasons the role of conventional cytogenetics remains pivotal in MDS. Cytogenetic results are very often complicated and biased, and therefore must be evaluated very carefully by a skilled cytogeneticist or a clinician experienced in cytogenetics in order to correctly interpret the karyotypic data. The combination of conventional cytogenetics with the results of molecular cytogenetic and genomic methods can produce more refined results and improve the risk stratification of MDS patients.



A. The complex karyotype detected by conventional cytogenetics (G-banding) of the bone marrow sample, marker chromosomes with unknown origin are present. The karyotype was described according to ISCN 2016: 43,XX,del(5)(q?),-6,-7,-8,-12,-15,-16,-17,?der(18),+4mar[cp24].

B. By M-FISH we proved the presence of the complex karyotype including both structural and numerical aberrations and we were able to specify more precisely particular abnormalities and marker chromosomes, e.g., monosomy 6, derivative chromosome der(7)t(7;17), trisomy 8 with two derivative chromosomes 8, etc. The karyotype was corrected: 43,XX,del(5)(q31q?35),-6,?der(7)t(7;17)(p?12;q?12),der(8)t(6;8)(?p;p?12)ins(8;15)(p?12;q?),der(8)(6qter>6q?13::12?>12?::8?>8?::15q?>15q?::6?>6?::15q?>15q?::8p11>8qter),+8,der(12)t(8;12)(?q;p?12)t(12;15)(q?24;q?),-15,-16,-17[10]

C. Interphase FISH with the locus specific probe XL Del(5)(q31) (MetaSystems, Germany) confirmed the deletion of 5q31 in 81% of cells. The orange labeled probe is designed to hybridize to the EGR1 locus at 5q31, whereas the green probe covers specific locus at 5p15 and functions as a control probe.

D. The result of metaphase FISH with centromeric probes specific for the centromeric regions of chromosome 7 (labeled in orange) and 17 (green) (Abbott Molecular, USA). The yellow signal is proving the presence of the dicentric chromosome dic(7;17). Based on this finding, which was not visible previously, the original description of der(7)t(7;17) was corrected for dic(7;17)(q11;p11) with consequently demonstrated deletion of the *TP53* gene.

E. ArrayCGH allowed us to detect already known changes as well as other previously hidden abnormalities and to precisely define the breakpoint regions and the aberrations of small sizes. In total we found 26 unbalanced genomic changes through the entire genome (SurePrint G3 ISCA CGH+SNP Microarray, 4x180K, Agilent), 21 losses (5q21.1q34, 6p25.3, 6p21.1, 6p12.3, 6p11.2q12, 6q12, 7q11.22q36.3, 8p21.3p12, 8p12p11.21, 12p13.33, 12p13.31, 12p12.3p11.22, 12q21.2q23.3, 12q24.11q24.33, 15q11.2q13.2, 15q21.3, 15q26.2, 16p13.3p12.3, 16p12.1, 16p12.1q24.3 and 17p13.3p11.2) and 5 gains (6p22.1p21.1, 6p21.1p12.3, 8p11.21q24.3, 15q21.3q26.2 and 16p12.1). Chromosomes are listed from the left to the right, from 1 to 9 and 10 to X and Y, respectively. The regions of gain are represented by the blue bars along the chromosomes, the regions of loss by the red ones.

All images were created in the Laboratory of Cytogenetics and Molecular Cytogenetics, Dpt. of Hemato-oncology Olomouc, Czech Republic, namely by: Kropackova J., Holzerova M. and Urbankova H.

Figure 1. Results of cytogenetics and molecular cytogenetic/genetic methods used to determine complex genomic aberrations in a patient with MDS diagnosed in our center.

Adult Acute Myeloid Leukemia (AML)

Acute myeloid leukemia is a neoplasm of the myeloid line of blood cells characterized by an accumulation of immature myeloid blasts in the bone marrow. This leads to an abnormal hematopoiesis and to a lack of differentiated granulocytes, monocytes, thrombocytes or erythrocytes. Based on the morphology, cytochemistry, immunophenotyping, genetic features and clinical characteristics, subgroups of AML can be defined with differing clinical relevance. There are currently seven subgroups of AML described (Arber et al., 2016; Swerdlow S. H., Campo E., Harris N. L., al., & 2008) and the majority of them has detectable clonal chromosomal aberrations. These aberrations, either numerical or structural, are summarized at the Mitelman Database of Chromosome Aberrations and Gene Fusions in Cancer (Mitelman, 2017). Cytogenetics provides strong prognostic information for prediction of the time of remission and post-remission therapy. Chromosomal abnormalities with good prognoses include t(8;21)(q22;q22.1), inv(16)(p13.1;q22) or t(16;16)(p13.1;q22) and t(15;17)(q22;q21). Patients carrying a normal karyotype are placed in the AML category with an average risk and require further molecular genetic analyses for accurate risk stratification. Patients with detected myelodysplasia-related changes del(5)(q), monosomies of chromosomes 5 or 7, translocations or inversions of chromosome 3, t(6;9), t(9;22), or abnormalities of chromosome 11q23 (*MLL* gene) have a particularly poor prognosis. Despite the fact that each specific category of AML has its own prognostic and treatment implications, for practical purposes antileukemic therapy at the diagnosis of AML is similar for all subtypes. The term complex karyotype in AML was introduced in the 1980s and at that time it was already observed to be associated with a poor prognosis (Berger et al., 1987; Levin, Le Coniat, Bernheim, & Berger, 1986). It was found that complex karyotypes are frequently characterized by the presence of certain chromosomal aberrations, for example monosomies of chromosome 5 and 7, deletions of 5q, 7q, 12p and 17p, moreover they are often associated with mutations of the *TP53* gene (Bowen et al., 2009).

The currently reported frequency of complex karyotypes with three and more unrelated chromosomal aberrations is 10-40% in all AML patients and increases with age (Stolzel et al., 2016). According to ELN (European Leukemia Net) recommendations, complex karyotype with three and more changes is classified as the adverse risk category. In comparison, the UK national classification system of Medical Research Council (MRC) for AML established the cutoff for adverse prognosis as the presence of at least four or more aberrations, and they consider the karyotype complex if it has at least five aberrations. Moreover, the definition of complex karyotype in AML is further complicated by the nature of its components. Breems et al., (Breems & Lowenberg, 2011) recently investigated the prognostic value of a monosomal karyotype, which is defined as either one autosomal monosomy and the presence of at least one structural aberration, or at least two distinct autosomal monosomies. They proved that the monosomal karyotype has a highly unfavorable impact on prognosis. On the contrary, it is well known that patients with complex karyotypes containing prognostically favorable aberrations, such as t(8;21), inv(16)/t(16;16) or t(15;17), remain in the good prognostic group despite the presence of the complex karyotype (Peterson, 2017).

Nevertheless, in general an increasing number of abnormalities included in complex karyotypes definitely worsens the overall survival of AML patients. Indeed, there is a need for a uniformly defined complex karyotype as an adverse-risk factor between different institutions. This issue should be figured out in the future by a survival analysis of a large multicenter cohort of AML patients with complex aberrant karyotypes.

Chronic Lymphocytic Leukemia (CLL)

Chronic lymphocytic leukemia is the most common type of leukemia in adults. It affects B-lymphocytes, which grow in an uncontrolled manner and accumulate in the bone marrow, peripheral blood and lymph nodes. In the past the use of conventional cytogenetics in the detection of chromosomal aberrations in CLL was not very successful due to neoplastic CLL cells having a very low mitotic activity, resulting in a lack of metaphase chromosomes in the cultures of peripheral blood cells.

Mitogenic stimulation of the CLL cultured cells with CpG-oligonucleotides and interleukin 2 was identified more than ten years ago as a reliable and reproducible tool that could obtain metaphase chromosomes and enable the assembly of a karyotype, even in CLL samples (Dicker, Schnittger, Haferlach, Kern, & Schoch, 2006). After an appropriate stimulation, chromosomal aberrations can now be detected in more than 80% of patients with CLL. The most frequent changes are deletions of the 13q14, 11q22, 17p13 and trisomy of chromosome 12 (Dohner et al., 2000). Deletions of the 11q22 (*ATM* gene) and 17p13 (*TP53* gene) are considered high risk factors with a poor outcome (Rigolin et al., 2017). The presence of complex aberrant karyotype was found in 10-20% of CLL patients (Haferlach, Dicker, Schnittger, Kern, & Haferlach, 2007; Mayr et al., 2006). Interestingly, many breakpoints are located in regions of recurrent losses in CLL, e.g., 13q14 and 17p13, which means conversely that complex karyotypes in CLL are often associated with 17p abnormalities (either deletions or mutations of the *TP53* gene) (Fink et al., 2006). In addition, the CLL subgroup with a complex karyotype seems to be strongly associated with an unmutated *IGVH* status and CD38 expression (Haferlach et al., 2007). According to recent studies, the complex karyotype became a strong independent predictive marker of rapid progression, resistance to therapy, early relapse and a short overall survival (Kujawski et al., 2008; Mayr et al., 2006; Puiggros et al., 2017; Van Den Neste et al., 2007).

Conventional cytogenetics represents the best tool for the detection of complex karyotypic abnormalities. Nevertheless, we also routinely use an interphase FISH technique which can estimate an aberrations in six precisely selected loci. The technique can provide information even in the absence of metaphase chromosomes or when the targeted abnormality is below the resolution of classical cytogenetics. We use commercially available probes for the numerical aberrations of 6q21, 8q24 (*MYC* gene), 11q22 (*ATM* gene), chromosome 12, 13q14, 13q34 and 17p13 (*TP53* gene) and rearrangement of the *IGH* gene (14q32). To complete the whole picture of a genome we perform an arrayCGH. The final karyotype is based on a combination of the results of all these methods.

Our data on complex karyotypes in CLL patients are in concordance with other authors (Kruzova L., 2017). We have detected complex karyotype in approximately 12% of untreated CLL cases diagnosed in our center and we have proven their adverse impact on the overall survival of patients. We did not observe any worsening effects due to the presence of *ATM* or *TP53* deletions as a part of a complex karyotype, which is in agreement with the finding of Thompson et al., (Thompson et al., 2015) and others. Moreover, we investigated particular chromosomal aberrations included in the complex karyotypes and their impact on the course of the disease in detail, but found no specific influence on the outcome in any way. Also we have suggested that patients with complex karyotypes are recommended for new treatments which are applied already in the first line of therapy.

Mantle Cell Lymphoma (MCL)

MCL comprises 3-10% of B-cell non-Hodgkin's lymphoma (B-NHL). Unlike other NHL types, nearly all cases are characterized by some level of the leukemic course of the disease with neoplastic cells detectable in the peripheral blood, as well as in bone marrow and other locations. This is advantageous for conventional cytogenetics, as the MCL infiltrated samples are available by non-invasive approach. The cytogenetic hallmark of the disease is the presence of t(11;14)(q13;q32), first identified by Banks et al., in 1992 (Banks et al., 1992), which leads to a juxtaposition of the *CCND1* gene to the *IGH* locus and to an overexpression of the cyclin D1. In approximately 20% of patients the translocation is a part of a complex karyotype (Gazzo et al., 2005) with the recurrent loss of der(11) and other changes. Complex karyotype is associated with a more aggressive disease and a shorter overall survival in MCL, but it is not clear if it can be regarded as an independent prognostic factor (Cohen et al., 2015; Sarkozy et al., 2014). This disease is quite rare and the data were mostly retrieved retrospectively. That is why our patients had been treated with different therapies and it is impossible to reliably compare their outcomes. We have analyzed cases diagnosed with MCL in our center from 2006 until now and have found complex karyotypes in 19% of patients (Obr A., unpublished data). Complex karyotypes were frequently accompanied by mutations in the *TP53* gene and a poor prognosis. We would need to analyze a larger cohort of patients treated with similar modern approaches to prove the independent prognostic impact of the presence of a complex karyotype in MCL.

Multiple Myeloma (MM)

Multiple myeloma is a cancer of the plasma cells - terminally differentiated B-lymphocytes which are normally responsible for the production of antibodies. Neoplastic plasma cells can form a mass in the bone marrow or soft tissue. Multiple myeloma is diagnosed based on the detection of abnormal levels of antibodies in the peripheral blood or urine, the finding of neoplastic plasma cells in bone marrow biopsy, and bone lesions revealed by medical imaging.

Cytogenetic analysis is an important part of MM diagnostics as chromosomal aberrations are powerful indicators of prognosis. Patients can be stratified into two prognostic categories according to the chromosome ploidy status and the rearrangements of the *IGH* gene. Generally, the hypodiploid group with $t(4;14)(p16;q32)$ or $t(14;16)(q32;q23)$ is associated with a high-risk disease, whereas the hyperdiploid patients with $t(11;14)(q13;q32)$ are considered a better prognostic group. During the disease progression the secondary chromosomal aberrations are developing. The most frequent are *MYC* rearrangements, deletions of the 13q, 17p, deletions of the 1p and/or amplification of the 1q. These secondary changes are characteristic of a highly proliferating and rapidly developing disease (Sawyer, 2011). Nevertheless, conventional cytogenetics is difficult because of the low level of bone marrow infiltration and the low mitotic index of plasma cells. Therefore cytogenetics is either unsuccessful or the results show normal karyotype in a majority of cases. In cases where abnormal karyotype is identified in a highly proliferative disease frequently multiple structural and numerical changes are found. Moreover patients often have various abnormal clones. Interphase FISH detects numerical and structural abnormalities in up to 90% of patients (Drach et al., 1995). To deal with a low infiltration of plasma cells in bone marrow samples, conventional interphase FISH was adapted to detect changes specifically in the plasma cells using a cytoplasmic light chain immunofluorescent labeling or purified CD138+ cells (Ahmann, 1998). The disadvantages of this approach are the restricted number of analyzed loci and lack of metaphase chromosomes, which subsequently also limits the detection of the complex karyotypes in MM. According to Nemec et al., (Nemec et al., 2012), metaphase cytogenetic analysis revealed the presence of complex karyotype in 19% of cases. Patients with a complex karyotype, translocation $t(4;14)$ and/or gain of the 1q21 locus had a shorter time to progression and a lower overall survival.

Recently, the arrayCGH technique has also found its place in the routine analysis of the genomic profile of the plasma cell, as it uses DNA from separated CD138+ cells. This method is able to identify multiple novel genetic aberrations that previously escaped detection. Nevertheless, the precise definition of the complex karyotype could not be matched in the majority of MM cases as we have no cytogenetic data proving a co-existence of various genomic changes in one clone. Perhaps this is the reason the complex karyotype is not discussed in MM as often as in other hematologic malignancies, even though it is probably present in a vast majority of them.

CONCLUSION

There is accumulative evidence that classical cytogenetics remains a valuable basic method for the analysis of genetic aberrations in the hemato-oncologic patients over decades. Previously it has been partially replaced by more precise molecular cytogenetics and genomic methods. However, with the recently recognized significance of the complex karyotypic changes in a wide spectrum of hematologic neoplasms, its role is again becoming crucial in the diagnostics and prognosis prediction. The correlation of cytogenetic results with clinical data and modern molecular genetic methods may provide clinicians with a broad and complex

overview of biological events taking place in tumor cells and may help them in therapy related decision making. Whereas in some hemato-oncologic diseases the definition of the complex karyotype is clear, in others it remains a topic of discussion. For example, in AML not only the number of unrelated clonal chromosomal aberrations but also its specific composition appear to be important. Complex karyotype is well defined and reported in CLL where it became an independent prognostic marker. On the contrary in other hemato-oncologic entities with problematic neoplastic infiltration of the bone marrow or peripheral blood (e.g., MM, B-/T-cell lymphomas other than MCL etc.), complex karyotypes are difficult to check with only conventional cytogenetics. Regardless of the scarcity of cytogenetic data, the results of other molecular cytogenetic methods show evidence of multiple genomic aberrations. In the future we can probably expect some adjustments to the definitions of certain questionable complex karyotypes in hematologic neoplasms.

In summary, conventional cytogenetics ideally complemented by other molecular cytogenetic and genomic methods remains an essential tool for the detection of the complex karyotype as an indicator of a poor prognosis across many hemato-oncologic diseases.

REFERENCES

- Arber, D. A., Orazi, A., Hasserjian, R., Thiele, J., Borowitz, M. J., Le Beau, M. M., ... Vardiman, J. W. (2016). The 2016 revision to the World Health Organization classification of myeloid neoplasms and acute leukemia. *Blood*, 127(20), 2391-2405. doi: 10.1182/blood-2016-03-643544.
- Banks, P. M., Chan, J., Cleary, M. L., Delsol, G., De Wolf-Peeters, C., Gatter, K., ... et al., (1992). Mantle cell lymphoma. A proposal for unification of morphologic, immunologic, and molecular data. *Am J Surg Pathol*, 16(7), 637-640.
- Bartkova, J., Horejsi, Z., Koed, K., Kramer, A., Tort, F., Zieger, K., ... Bartek, J. (2005). DNA damage response as a candidate anti-cancer barrier in early human tumorigenesis. *Nature*, 434(7035), 864-870. doi: 10.1038/nature03482.
- Berger, R., Bernheim, A., Ochoa-Noguera, M. E., Daniel, M. T., Valensi, F., Sigaux, F., ... Boiron, M. (1987). Prognostic significance of chromosomal abnormalities in acute nonlymphocytic leukemia: a study of 343 patients. *Cancer Genet Cytogenet*, 28(2), 293-299.
- Bowen, D., Groves, M. J., Burnett, A. K., Patel, Y., Allen, C., Green, C., ... Linch, D. C. (2009). TP53 gene mutation is frequent in patients with acute myeloid leukemia and complex karyotype, and is associated with very poor prognosis. *Leukemia*, 23(1), 203-206. doi: 10.1038/leu.2008.173.
- Breems, D. A., & Lowenberg, B. (2011). Acute myeloid leukemia with monosomal karyotype at the far end of the unfavorable prognostic spectrum. *Haematologica*, 96(4), 491-493. doi: 10.3324/haematol.2011.043208.
- Carter, N. P., Fiegler, H., & Piper, J. (2002). Comparative analysis of comparative genomic hybridization microarray technologies: report of a workshop sponsored by the Wellcome Trust. *Cytometry*, 49(2), 43-48. doi: 10.1002/cyto.10153.

- Cohen, J. B., Ruppert, A. S., Heerema, N. A., Andritsos, L. A., Jones, J. A., Porcu, P., ... Blum, K. A. (2015). Complex karyotype is associated with aggressive disease and shortened progression-free survival in patients with newly diagnosed mantle cell lymphoma. *Clin Lymphoma Myeloma Leuk*, 15(5), 278-285 e271. doi: 10.1016/j.clml.2014.12.012.
- Dicker, F., Schnittger, S., Haferlach, T., Kern, W., & Schoch, C. (2006). Immunostimulatory oligonucleotide-induced metaphase cytogenetics detect chromosomal aberrations in 80% of CLL patients: A study of 132 CLL cases with correlation to FISH, IgVH status, and CD38 expression. *Blood*, 108(9), 3152-3160. doi: 10.1182/blood-2006-02-005322.
- Dohner, H., Stilgenbauer, S., Benner, A., Leupolt, E., Krober, A., Bullinger, L., ... Lichter, P. (2000). Genomic aberrations and survival in chronic lymphocytic leukemia. *N Engl J Med*, 343(26), 1910-1916. doi: 10.1056/NEJM200012283432602.
- Drach, J., Angerler, J., Schuster, J., Rothermundt, C., Thalhammer, R., Haas, O. A., ... Huber, H. (1995). Interphase fluorescence in situ hybridization identifies chromosomal abnormalities in plasma cells from patients with monoclonal gammopathy of undetermined significance. *Blood*, 86(10), 3915-3921.
- Fiegler, H., Carr, P., Douglas, E. J., Burford, D. C., Hunt, S., Scott, C. E., ... Carter, N. P. (2003). DNA microarrays for comparative genomic hybridization based on DOP-PCR amplification of BAC and PAC clones. *Genes Chromosomes Cancer*, 36(4), 361-374. doi: 10.1002/gcc.10155.
- Fink, S. R., Smoley, S. A., Stockero, K. J., Paternoster, S. F., Thorland, E. C., Van Dyke, D. L., ... Dewald, G. W. (2006). Loss of TP53 is due to rearrangements involving chromosome region 17p10 approximately p12 in chronic lymphocytic leukemia. *Cancer Genet Cytogenet*, 167(2), 177-181. doi: 10.1016/j.cancergencyto.2006.01.005.
- Gaillard, H., Garcia-Muse, T., & Aguilera, A. (2015). Replication stress and cancer. *Nat Rev Cancer*, 15(5), 276-289. doi: 10.1038/nrc3916.
- Gazzo, S., Felman, P., Berger, F., Salles, G., Magaud, J. P., & Callet-Bauchu, E. (2005). Atypical cytogenetic presentation of t(11;14) in mantle cell lymphoma. *Haematologica*, 90(12), 1708-1709.
- Gohring, G., Michalova, K., Beverloo, H. B., Betts, D., Harbott, J., Haas, O. A., ... Schlegelberger, B. (2010). Complex karyotype newly defined: the strongest prognostic factor in advanced childhood myelodysplastic syndrome. *Blood*, 116(19), 3766-3769. doi: 10.1182/blood-2010-04-280313.
- Greenberg, P. L., Tuechler, H., Schanz, J., Sanz, G., Garcia-Manero, G., Sole, F., ... Haase, D. (2012). Revised international prognostic scoring system for myelodysplastic syndromes. *Blood*, 120(12), 2454-2465. doi: 10.1182/blood-2012-03-420489.
- Haase, D., Germing, U., Schanz, J., Pfeilstocker, M., Nosslinger, T., Hildebrandt, B., ... Steidl, C. (2007). New insights into the prognostic impact of the karyotype in MDS and correlation with subtypes: evidence from a core dataset of 2124 patients. *Blood*, 110(13), 4385-4395. doi: 10.1182/blood-2007-03-082404.
- Haferlach, C., Dicker, F., Schnittger, S., Kern, W., & Haferlach, T. (2007). Comprehensive genetic characterization of CLL: a study on 506 cases analyzed with chromosome banding analysis, interphase FISH, IgV(H) status and immunophenotyping. *Leukemia*, 21(12), 2442-2451. doi: 10.1038/sj.leu.2404935.
- Halazonetis, T. D., Gorgoulis, V. G., & Bartek, J. (2008). An oncogene-induced DNA damage model for cancer development. *Science*, 319(5868), 1352-1355. doi: 10.1126/science.1140735.

- Hanahan, D., & Weinberg, R. A. (2011). Hallmarks of cancer: the next generation. *Cell*, 144(5), 646-674. doi: 10.1016/j.cell.2011.02.013.
- Heim, S., & Mitelman, F. (1989). Primary chromosome abnormalities in human neoplasia. *Adv Cancer Res*, 52, 1-43.
- Kallioniemi, A., Kallioniemi, O. P., Sudar, D., Rutovitz, D., Gray, J. W., Waldman, F., & Pinkel, D. (1992). Comparative genomic hybridization for molecular cytogenetic analysis of solid tumors. *Science*, 258(5083), 818-821.
- Knijnenburg, J., Szuhai, K., Giltay, J., Molenaar, L., Sloos, W., Poot, M., ... Rosenberg, C. (2005). Insights from genomic microarrays into structural chromosome rearrangements. *Am J Med Genet A*, 132A(1), 36-40. doi: 10.1002/ajmg.a.30378.
- Kolialexi, A., Tsangaris, G. T., Kitsiou, S., Kanavakis, E., & Mavrou, A. (2005). Impact of cytogenetic and molecular cytogenetic studies on hematologic malignancies. *Anticancer Res*, 25(4), 2979-2983.
- Kruzova L., S. P., Holzerova M., Vatolikova M., Divoka M., Turcsanyi P., Urbanova R., Kudelka M., Radvansky M., Kriegova E., Papajik T., Urbankova H. . (2017). Gender, Age and TP53 Aberrations in Context of Complex Karyotypes in Chronic Lymphocytic Leukemia Patients in the Era of Chemo +/- Immunotherapy *Blood*, 130(Suppl 1), 2991.
- Kujawski, L., Ouillet, P., Erba, H., Saddler, C., Jakubowiak, A., Kaminski, M., ... Malek, S. N. (2008). Genomic complexity identifies patients with aggressive chronic lymphocytic leukemia. *Blood*, 112(5), 1993-2003. doi: 10.1182/blood-2007-07-099432.
- Langer-Safer, P. R., Levine, M., & Ward, D. C. (1982). Immunological method for mapping genes on Drosophila polytene chromosomes. *Proc Natl Acad Sci US*, 79(14), 4381-4385.
- Levin, M., Le Coniat, M., Bernheim, A., & Berger, R. (1986). Complex chromosomal abnormalities in acute nonlymphocytic leukemia. *Cancer Genet Cytogenet*, 22(2), 113-119.
- Mayr, C., Speicher, M. R., Kofler, D. M., Buhmann, R., Strehl, J., Busch, R., ... Wendtner, C. M. (2006). Chromosomal translocations are associated with poor prognosis in chronic lymphocytic leukemia. *Blood*, 107(2), 742-751. doi: 10.1182/blood-2005-05-2093.
- McGowan-Jordan J., Simons A., & Schmid M., e. (Eds.). (2016). *An International System for Human Cytogenomic Nomenclature (2016)*.
- Mitelman, F. (2017). *Mitelman Database of Chromosome Aberrations and Gene Fusions in Cancer* / Mitelman Database <https://omictools.com/mitelman-database-of-chromosome-aberrations-and-gene-fusions-in-cancer-tool>.
- Negrini, S., Gorgoulis, V. G., & Halazonetis, T. D. (2010). Genomic instability--an evolving hallmark of cancer. *Nat Rev Mol Cell Biol*, 11(3), 220-228. doi: 10.1038/nrm2858.
- Nemec, P., Zemanova, Z., Kuglik, P., Michalova, K., Tajtlova, J., Kaisarova, P., ... Czech Myeloma, G. (2012). Complex karyotype and translocation t(4;14) define patients with high-risk newly diagnosed multiple myeloma: results of CMG2002 trial. *Leuk Lymphoma*, 53(5), 920-927. doi: 10.3109/10428194.2011.634042.
- Olney, H. J., & Le Beau, M. M. (2001). The cytogenetics of myelodysplastic syndromes. *Best Pract Res Clin Haematol*, 14(3), 479-495. doi: 10.1053/beh.2001.0151.
- Palumbo, E., & Russo, A. (2016). Chromosome Imbalances in Cancer: Molecular Cytogenetics Meets Genomics. *Cytogenet Genome Res*, 150(3-4), 176-184. doi: 10.1159/000455804.
- Peterson, J. F. (2017). The Complexities of Defining a Complex Karyotype in Hematological Malignancies: A Need for Standardization? *Acta Haematol*, 138(1), 65-66. doi: 10.1159/000477894.

- Pinkel, D., Seagraves, R., Sudar, D., Clark, S., Poole, I., Kowbel, D., ... Albertson, D. G. (1998). High resolution analysis of DNA copy number variation using comparative genomic hybridization to microarrays. *Nat Genet*, 20(2), 207-211. doi: 10.1038/2524.
- Pinkel, D., Straume, T., & Gray, J. W. (1986). Cytogenetic analysis using quantitative, high-sensitivity, fluorescence hybridization. *Proc Natl Acad Sci US*, 83(9), 2934-2938.
- Puiggros, A., Collado, R., Calasanz, M. J., Ortega, M., Ruiz-Xiville, N., Rivas-Delgado, A., ... Espinet, B. (2017). Patients with chronic lymphocytic leukemia and complex karyotype show an adverse outcome even in absence of TP53/ATM FISH deletions. *Oncotarget*, 8(33), 54297-54303. doi: 10.18632/oncotarget.17350.
- Rigolin, G. M., Formigaro, L., Cavallari, M., Quaglia, F. M., Lista, E., Urso, A., ... Cuneo, A. (2017). An extensive molecular cytogenetic characterization in high-risk chronic lymphocytic leukemia identifies karyotype aberrations and TP53 disruption as predictors of outcome and chemorefractoriness. *Oncotarget*, 8(17), 28008-28020. doi: 10.18632/oncotarget.15883.
- Sarkozy, C., Terre, C., Jardin, F., Radford, I., Roche-Lestienne, C., Penther, D., ... Castaigne, S. (2014). Complex karyotype in mantle cell lymphoma is a strong prognostic factor for the time to treatment and overall survival, independent of the MCL international prognostic index. *Genes Chromosomes Cancer*, 53(1), 106-116. doi: 10.1002/gcc.22123.
- Sawyer, J. R. (2011). The prognostic significance of cytogenetics and molecular profiling in multiple myeloma. *Cancer Genet*, 204(1), 3-12. doi: 10.1016/j.cancergencyto.2010.11.002.
- Schanz, J., Tuchler, H., Sole, F., Mallo, M., Luno, E., Cervera, J., ... Haase, D. (2012). New comprehensive cytogenetic scoring system for primary myelodysplastic syndromes (MDS) and oligoblastic acute myeloid leukemia after MDS derived from an international database merge. *J Clin Oncol*, 30(8), 820-829. doi: 10.1200/JCO.2011.35.6394.
- Schrock, E., du Manoir, S., Veldman, T., Schoell, B., Wienberg, J., Ferguson-Smith, M. A., ... Ried, T. (1996). Multicolor spectral karyotyping of human chromosomes. *Science*, 273(5274), 494-497.
- Sikkema-Raddatz, B., Johansson, L. F., de Boer, E. N., Almomani, R., Boven, L. G., van den Berg, M. P., ... Sinke, R. J. (2013). Targeted next-generation sequencing can replace Sanger sequencing in clinical diagnostics. *Hum Mutat*, 34(7), 1035-1042. doi: 10.1002/humu.22332.
- Solinas-Toldo, S., Lampel, S., Stilgenbauer, S., Nickolenko, J., Benner, A., Dohner, H., ... Lichter, P. (1997). Matrix-based comparative genomic hybridization: biochips to screen for genomic imbalances. *Genes Chromosomes Cancer*, 20(4), 399-407.
- Speicher, M. R., Gwyn Ballard, S., & Ward, D. C. (1996). Karyotyping human chromosomes by combinatorial multi-fluor FISH. *Nat Genet*, 12(4), 368-375. doi: 10.1038/ng0496-368.
- Stephens, P. J., Greenman, C. D., Fu, B., Yang, F., Bignell, G. R., Mudie, L. J., ... Campbell, P. J. (2011). Massive genomic rearrangement acquired in a single catastrophic event during cancer development. *Cell*, 144(1), 27-40. doi: 10.1016/j.cell.2010.11.055
- Stolzel, F., Mohr, B., Kramer, M., Oelschlagel, U., Bochtler, T., Berdel, W. E., ... Ehninger, G. (2016). Karyotype complexity and prognosis in acute myeloid leukemia. *Blood Cancer J*, 6, e386. doi: 10.1038/bcj.2015.114.
- Swerdlow S. H., Campo E., Harris N. L., al., e., & e. (Eds.). (2008). *WHO Classification of Tumours of Haematopoietic and Lymphoid Tissues*.

- Thompson, P. A., O'Brien, S. M., Wierda, W. G., Ferrajoli, A., Stingo, F., Smith, S. C., ... Keating, M. J. (2015). Complex karyotype is a stronger predictor than del(17p) for an inferior outcome in relapsed or refractory chronic lymphocytic leukemia patients treated with ibrutinib-based regimens. *Cancer*, 121(20), 3612-3621. doi: 10.1002/cncr.29566.
- Van Den Neste, E., Robin, V., Francart, J., Hagemeijer, A., Stul, M., Vandenberghe, P., ... Michaux, L. (2007). Chromosomal translocations independently predict treatment failure, treatment-free survival and overall survival in B-cell chronic lymphocytic leukemia patients treated with cladribine. *Leukemia*, 21(8), 1715-1722. doi: 10.1038/sj.leu.2404764.
- Vardiman, J. W. (2003). Myelodysplastic syndromes, chronic myeloproliferative diseases, and myelodysplastic/myeloproliferative diseases. *Semin Diagn Pathol*, 20(3), 154-179.
- Vermeesch, J. R., Fiegler, H., de Leeuw, N., Szuhai, K., Schoumans, J., Ciccone, R., ... Zuffardi, O. (2007). Guidelines for molecular karyotyping in constitutional genetic diagnosis. *Eur J Hum Genet*, 15(11), 1105-1114. doi: 10.1038/sj.ejhg.5201896.
- Vermeesch, J. R., Melotte, C., Froyen, G., Van Vooren, S., Dutta, B., Maas, N., ... Devriendt, K. (2005). Molecular karyotyping: array CGH quality criteria for constitutional genetic diagnosis. *J Histochem Cytochem*, 53(3), 413-422. doi: 10.1369/jhc.4A6436.2005.
- Yoshizato, T., Nannya, Y., Atsuta, Y., Shiozawa, Y., Iijima-Yamashita, Y., Yoshida, K., ... Ogawa, S. (2017). Genetic abnormalities in myelodysplasia and secondary acute myeloid leukemia: impact on outcome of stem cell transplantation. *Blood*, 129(17), 2347-2358. doi: 10.1182/blood-2016-12-754796.

Chapter 57

**NATURAL HYBRIDIZATION BETWEEN
CHROMOSOMAL DISCREPANT SPECIES
AND THE ROLE OF HYBRID SPECIATION IN THE
GENUS *ASTYANAX***

***Francisco de Menezes Cavalcante Sassi¹,
Snaydia Viegas Resende^{1,2}, Rubens Pazza¹
and Karine Frehner Kavalco^{1,2}***

¹Laboratório de Genética Ecológica e Evolutiva, Universidade Federal de Viçosa,
Campus Rio Paranaíba. Rio Paranaíba – MG, Brasil

²Programa de Pós-Graduação em Manejo e Conservação de Ecossistemas Naturais e
Agrários (MCENA), Universidade Federal de Viçosa, Campus Florestal– MG, Brasil

ABSTRACT

Initially identified as *Astyanax schubarti* on the basis of morphological characteristics, its chromosomal analysis revealed a unique diploid number in the genus. With 42 chromosomes and the impossibility of homologous pairing, the karyotype of the individual was compared to a set of haploid complements of *Astyanax schubarti* ($2n = 36$ chromosomes) and *Astyanax fasciatus* ($2n = 48$ chromosomes). Natural hybrids are rare. The viability of a hybrid between species with such chromosomal discrepancy may offer important hypotheses to explain the morphological, molecular, and cytogenetic diversity of the genus.

Keywords: hybridization, speciation, ecological genetics, evolutionary genetics

INTRODUCTION

Natural hybrids emerge following mating between individuals of two populations or distinct species, and they are identified on the basis of one or more hereditary characteristics (Downling and Secor 1997). The role of hybridization in the evolution of species remains controversial, but hybrid speciation, mainly in polyploid plants, is well documented (Mallet 2007). This phenomenon is common among species of teleost fish (Scribner et al., 2001; Salzburger et al., 2002; Meyer et al., 2005), but few studies have focused on the species of the Neotropical region, particularly in South America.

Piabas or *lambaris* in Brazil are representatives of fish belonging to the genus *Astyanax*, which is most abundant in South America (Géry 1977). Among the about 150 species already described, *A. mexicanus* stands out, and because it has cave populations with troglomorphic characteristics, it is used as a model species in developmental and evolutionary biology studies (Jeffery 2001). In South America, where most of the biodiversity of this genus occurs, chromosomal studies have identified several cryptic species in the last 25 years, including the groups *A. scabripinnis* ($2n = 46, 48$, and 50), *A. fasciatus* ($2n = 46$ and 48), and *A. bimaculatus* ($2n = 50$). In addition to the differences in diploid numbers, these groups also present variation in karyotype forms and in the presence of supernumerary chromosomes (Pazza and Kavalco 2007).

Most species of *Astyanax* with a characterized chromosome number have 46 to 50 chromosomes, except for *A. schubarti*, which has $2n = 36$ chromosomes, possibly resulting from Robertsonian translocations from an ancestral karyotype (Daniel-Silva and Almeida-Toledo 2005). *A. fasciatus* can present $2n = 46$ or 48 chromosomes in sympatric or allopatric populations (Artoni et al., 2006; Pazza et al., 2006; Pazza et al., 2008; Pansonato-Alves et al., 2013; Penteado et al., 2013; Pazza et al., 2016). In the Mogi-Guaçu River (upper Rio Paraná Basin), the largest variation in sympatric chromosomes is observed in the species, where in addition to the standard cytotypes (with adequate homology between the chromosomes) $2n = 46$ and 48 , specimens with $2n = 45, 46$ variants, and 47 chromosomes have also been observed, without supernumerary chromosomes (Pazza et al., 2006). These authors suggest that the chromosome number $2n = 48$ is common in the Mogi-Guaçu River basin, whereas the chromosome number $2n = 46$ represents an invasion from the Tietê River basin, and that secondary contact after the recent diversification enabled gene flow between the cytotypes, despite the non-identification of obvious F1 hybrids in the study.

Thus, the present study, which presents a hybrid between species with divergent karyotypes, *A. schubarti* and *A. fasciatus*, provides important data for future studies on the natural history of the group.

MATERIAL AND METHODS

Samples of the genus *Astyanax* were collected in the Mogi-Guaçu River (143 individuals), in the region of Cachoeira de Emas (Pirassununga, São Paulo) between 2001 and 2003 (-21.92695 , -47.3673). Animals were fixed in formalin 10%, put in in ethanol 70% and deposited in the vertebrate collection of the Laboratory of Ecological and Evolutionary Genetics of the Federal University of Viçosa, Rio Paranaíba Campus (LaGEEvo UFV/CRP) - MG, Brazil. The tissue samples used for molecular analyses were stored at -20°C in pure ethanol. Species

identification follow the diagnostic characters for this species, like teeth morphology, fin color after fixation and meristic.

Mitotic chromosomes were obtained by the air-drying technique (Bertollo et al., 1978) from the anterior kidney of the animal. Chromosomes were analyzed after conventional staining with 1% Giemsa, and at least 20 metaphases were counted. The chromosomes were classified as metacentric (m), submetacentric (sm), subtelocentric (st), and acrocentric (a), according to the method described by Levan et al., (1964), using Photoshop CC v3.5 for measurement, and were later organized into karyotypes, according to the convention for fish.

Total DNA was extracted from each individual using a commercially available kit and fragments of organs, such as liver and heart, deposited in the Tissue Bank of the Laboratory of the Ecological and Evolutionary Genetics (LaGEEvo), Federal University of Viçosa, Rio Paranaíba *Campus*.

The mitochondrial gene (mtDNA) cytochrome oxidase I (COI) was amplified with the primers Fish R1 - 5' TAGACTTCTG GGTGGCCAAAGAATCA3' and Fish F1 - 5' TCAACCAACC ACAAAGACATTGGCAC3' (Ward et al., 2005). PCR was performed in a final volume of 25 μ L, with 2.5 μ L of 10X Taq buffer, 1 μ L $MgCl_2$, 1 μ L of each primer, 0.2 μ L of Taq DNA polymerase, 12.8 μ L ultra-pure water, 1.5 μ L of dNTPs, and 5 μ L of the extracted DNA solution. Amplifications were performed in a thermal cycler as follows: initial denaturation at 95 °C for 2 min; and repeated cycles of denaturation at 94 °C for 30 s, primer annealing at 58 °C for 30 s, and a final extension at 72 °C for 1 min. PCR products were visualized on a 1% agarose gel and samples were sent to a third-party company (Macrogen Korea) for purification and sequencing.

Once the sequences were obtained, they were identified using the BLASTn algorithm in the GenBank database (<http://blast.ncbi.nlm.nih.gov>), which is used to compare nucleotide sequences.

RESULTS

One individual (Figure 1) identified as *A. schubarti* Britski (1964), presented 42 chromosomes (10 m + 19 sm + 8 st + 5 a) (Figure 2). It was not possible to completely pair the chromosomes, since measurements showed that the size of each chromosome is different, and morphology too, indicating that there is no homology.

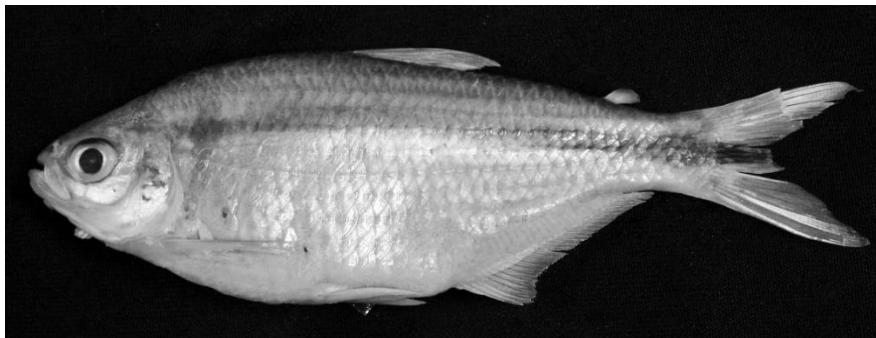


Figure 1. Photograph of the hybrid between *Astyanax fasciatus* and *Astyanax schubarti* fixed in 10% formalin and preserved in 70% alcohol. Standard length = 80.89mm.

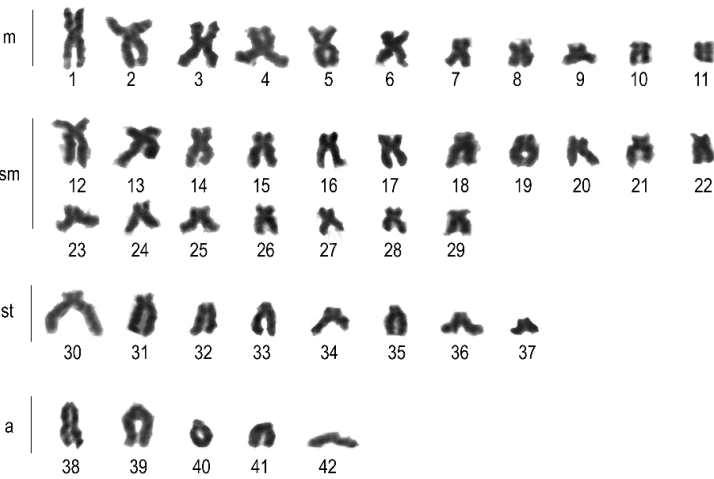


Figure 2. Karyotype of the hybrid stained with Giemsa.

Analysis of genetic identity using BLAST indicated 99% similarity in the gene encoding *cytochrome oxidase I* from the study specimen with most of the sequences of the same gene from *A. fasciatus* present in the database.

DISCUSSION

In this study, we report the first natural hybrid of the genus *Astyanax*. Unlike other natural hybrids found in fish, it is necessary to emphasize discrepancies in the diploid number of parent species. The contrary hypothesis would be a completely new cytotype for the genus *Astyanax* in more than 40 years of cytogenetics studies in the genus, but it is not plausible since we have found only one specimen among more than 240 individuals analyzed.

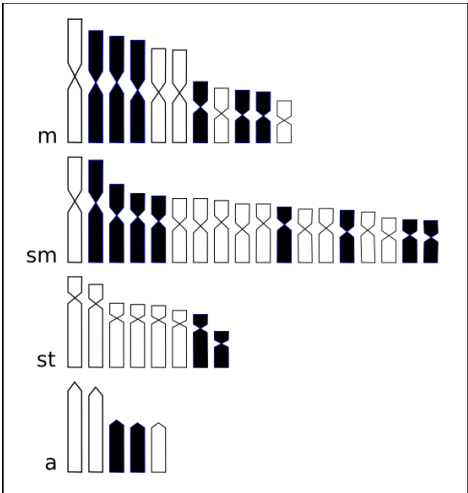


Figure 3. Ideogram representing the probable parental origin of the hybrid chromosomes between *Astyanax schubarti* (in black) and *A. fasciatus* (in white).

As the chromosome number was not consistent with that of any other *Astyanax* species, we tested the hypothesis that it represented a hybrid of *A. schubarti* ($2n = 36$, 12 m + 16 sm + 4 st + 4 a) and *A. fasciatus* ($2n = 48$, 8 m + 22 sm + 12 st + 6 a), which are both sympatric and syntopic species in Cachoeira de Emas. Hence, an ideogram with the haploid chromosome complement of each species was assembled from karyotype data in the literature (Daniel-Silva & Almeida-Toledo 2005; Pazza et al., 2006), and the result (Figure 3) was congruent with the karyotype obtained.

Other natural, or even artificial hybrids, of Neotropical species have parents with at least the same chromosome number. Parents of the natural hybrid between species of the genus *Cichla* (*Cichla monoculus* and *C. temensis*) feature $2n = 48$ chromosomes (Brinn et al., 2004), with little difference in the location of the nucleolus organizer region. Natural hybrids of *Pseudoplatystoma* species have already been observed in some populations using molecular markers (Prado et al., 2012a; Carvalho et al., 2013). However, no cytogenetic differences have been observed between *Pseudoplatystoma corruscans* and *P. reticulatum*, which both have $2n = 56$ chromosomes, between parents and the artificial hybrid (Prado et al., 2012b). The same can be observed in the artificial hybrid between *Piaractus mesopotamicus* and *Colossoma macropomum*, both with $2n = 54$ chromosomes, and one of the oldest produced in Brazil (Almeida-Toledo et al., 1987). At that time, both species were classified under the same genus. In turn, the hybrid between *Colossoma macropomum* and *Piaractus brachipomus* (also with $2n = 54$ chromosomes in both species) can be recognized by the location of the 18S ribosomal gene, because it has four loci distributed in its chromosomal complement while the parents have two and six, respectively (Nirchio et al., 2003). In the *Astyanax* genus, artificial hybrids between surface and cave populations of *A. mexicanus* have been developed for studies investigating the development biology of the eye (Wilkins 1971). Furthermore, mitochondrial DNA sequences suggest introgression events between these forms, which are currently geographically isolated (Dowling et al., 2002). Although no chromosomal data are available, it is likely that both populations have a similar chromosomal complement (Kavalco and Almeida-Toledo 2007), with $2n = 50$ chromosomes. Thus, an intermediate diploid hybrid between the parent species *A. fasciatus* ($2n = 48$) and *A. schubarti* ($2n = 36$) with significantly different diploid numbers is remarkable.

Although being characterized as common in fish (Scribner et al., 2001), natural hybrids in Neotropical fish are scarcely reported. Among the 168 species reported by those authors, none is from South America, although the literature indicates a hybrid between *P. mesopotamicus* and *C. Macropomum* as described previously (Almeida-Toledo et al., 1987). The difficulty in identification of hybrids in nature can be explained by the sterility of interspecific hybrids (Anderson 1948). However, difficulties in identification can be related to morphological characteristics, since these are used initially to identify species. Although some variables are intermediate in hybrids, their distribution is not always uniform, according to multivariate analysis (Neff and Smith 1979), leading the hybrid to present more characteristics of one of the parents. In fact, the hybrid found in the present study was initially identified as *A. schubarti*, due to the presentation of classic morphological characteristics, differing from *A. fasciatus* by the height of the body and the absence of red pigmentation of the caudal and dorsal fins (Britski 1964), and from *A. altiparanae* by the absence of an oval humeral spot (Garutti and Britski 2000). This suggests that the low frequency of hybrids documented in the Neotropical region can be related to the subtlety of morphological characteristics presented by hybrids, impeding their observation when only morphological data are considered.

One of the bets of current science against problems such as this is DNA barcoding, which involves the analysis of a specific mitochondrial sequence (COI) for the taxonomic identification of unknown individuals (Hebert et al., 2003). Analysis of this sequence in hybrid individuals enables the identification of *A. fasciatus*, which suggests that this is the female parental species by mitochondrial inheritance. However, it is important to highlight that DNA barcoding may not always be effective, especially for cryptic species, often with hybridism or with small sample sizes (Moritz and Cicero 2004; Meyer and Paulay 2005). The latter seems to be resolved in the genus *Astyanax*, which has about 150 described species and for which DNA barcoding has identified up to five key groups and several potential new species (Rossini et al., 2016). According to these authors, *A. fasciatus* and *A. schubarti* belong to clade 1 and display the lowest interspecific distance, with clusters including more than one nominal species. The low divergence observed in this clade may be explained by phenotypic plasticity, which hinders the identification of species; different rates of evolution in the *COI* gene between groups, with different groups having more recent adaptive radiation than others; and the description of new species only by their restricted geographical distribution, which would require future studies for synonymization (Rossini et al., 2016). Thus, despite the possibility of reducing gaps with larger sampling, resolution of the molecular taxonomy of *Astyanax* by DNA barcoding can be compromised by species complexes and hybridism.

The viability of a hybrid between species with such disparate chromosome numbers can help in our understanding of hybridity as an important factor for diversification in *Astyanax*. The fact that the individual morphologically of the hybrid resembles one of the parents may aid in the introgression of the genome in the population, since some species of fish may recognize their partners by morphology, as in the case of African cichlids (Kocher 2004). The speed and direction of evolution in new environments can be directly influenced by introgression, as this increases genetic variation (Lu et al., 1995; Abbott et al., 2003) and can produce new genotypes (Svardson 1970; Rieseberg et al., 2003). This may facilitate evolutionary divergence and speciation by reorienting the evolution of populations when there are changes in the environment (Grant et al., 1995). Although the emergence of a natural hybrid is rare, this is not related to evidence of the insignificance of hybridization and introgression in the evolution and diversification of animals (Downling & Secor 1997), since even small rates of hybridization can have a great impact on the evolution of a species (Mallet 2005). A recent adaptive radiation (Ornelas-Garcia et al., 2008) involving morphological and chromosomal variation, can therefore explain the viability in hybrids and their fundamental role in the evolution of the *Astyanax* genus. In addition, this also explains the introgressions and the shared haplotypes between different species observed by molecular data.

Additionally, Scribner et al., (2001) noted four main factors for hybridization in fish: habitat loss, distribution expansion, aquaculture, and introduction. These factors can be easily attributed to anthropogenic action. In view of this, more and more human endeavors aimed at altering aquatic habitats must be monitored using multidisciplinary approaches, including molecular and cytogenetic markers, so that the management of possible anthropological hybridization events is minimized. We conclude that hybridisms between species with discrepant chromosomal numbers can occur in nature and could have great importance in *Astyanax* evolution. More studies are necessary to understand the impact of such events in evolution of this group.

ACKNOWLEDGMENTS

The authors thank FAPEMIG (Fundação de Amparo à Ciência do Estado de Minas Gerais) and to Post-graduation Program MCENA/UFV (Manejo e Conservação de Ecossistemas Naturais e Agrários).

REFERENCES

- Abbot, R. J., James, J. K., Milne, R. I., Gillies, A. C. M. (2003) Plant introductions, hybridization and gene flow. *Philos. Trans. R. Soc. Lond. B., Biol. Sci.* 358: 1123-1132.
- Almeida-Toledo, L. F., Foresti, F., Toledo Filho, S. A., Bernardino, G., Ferrari, W., Alcantara, R. C. G. (1987) Cytogenetic studies of *Colossoma mitrei*, *Colossoma macropomum* and their interspecific hybrid. *Proc. World Symp. Selection Hybrid Genet. Eng. Aquac.* 189-195.
- Anderson, E. (1948) Hybridization of the habitat. *Inter. Journ. Organ. Evol.* 1: 1-9.
- Artoni, R. F., Shibatta, O. A., Gross, M. C., Schneider, C. H., Almeida, M. C., Vicari, M. R., Bertollo, L. A. C. (2006) *Astyanax* aff. *Fasciatus* Cuvier, 1819 (Teleostei; Characidae): evidences of a species complex in the Upper rio Tibagi basin (Paraná, Brazil). *Neotrop. Ichtyol.* 4: 197-202.
- Bertollo, L. A. C., Takahashi, C. S., Moreira-Filho, O. (1978) Cytotaxonomic considerations on *Hoplias lacerdae* (Pisces, Erythrinidae). *Braz. J. Genet.* 1: 103-120.
- Brinn, M. N. A., Porto, J. I. R., Feldberg, E. (2004) Karyological evidence for interspecific hybridization between *Cichla monoculus* and *C. temensis* (Perciformes, Cichlidae) in the Amazon. *Hereditas.* 141: 252-257.
- Britski, H. A. (1964) Sobre uma nova espécie de *Astyanax* do rio Mogi-Guaçu (Pisces, Characidae). [On a new species of *Astyanax* of the river Mogi-Guaçu (Pisces, Characidae)]. *Pap. Avuls. Depart. Zool.* 16: 213-215.
- Carvalho, D. C., Seerig, A. S., Brasil, B. S. A. F., Crepaldi, D. V., Oliveria DAA (2013) Molecular identification of the hybrid between the catfish species *Pseudoplatystoma corruscans* and *Pseudoplatystoma reticulatum* using a set of eight microsatellite markers. *J. Fish. Biol.* 83: 671-676.
- Daniel-Silva, M. F. Z., Almeida-Toledo, L. F. (2005) Chromosome evolution in fish: BrdU replication patterns demonstrate chromosome homeologies in two species of the genus *Astyanax*. *Cytog. Gen. Res.* 109: 497-501.
- Dowling, T. E. and Secor, C. L. (1997) The role of hybridization and introgression in the diversification of animals. *Ann. Rev. Eco. System.* 28: 593-619.
- Dowling, T. E., Martasian, D. P., Jeffery, W. R. (2002) Evidence for multiple genetic forms with similar eyeless phenotypes in the blind cavefish, *Astyanax mexicanus*. *Mol. Biol. Evol.* 19: 446-455.
- Eigenmann, C. H. (1921). The American Characidae. *Mem. Museum Comp. Zool.* 43: 209-310.
- Garutti, V. and Britski, H. A. (2000) Descrição de uma espécie nova de *Astyanax* (Teleostei: Characidae) da bacia do alto Rio Paraná e considerações sobre as demais espécies do gênero na bacia. [Description of a new species of *Astyanax* (Teleostei: Characidae) from

- the upper Paraná River basin and considerations about the other species of the genus in the basin *Comum. Museu Cienc. Tecnol.* 13: 65–88.
- Géry, J. (1977) *Characoids of the world*. 1st edn. Tfh Pubns Inc: Neptune.
- Grant, P. R., Grant, B. R., Petren, K. (1995) Hybridization in the recent past. *Am. Nat.* 166: 56–67.
- Hebert, P. D. N., Cywinska, A., Ball, S. L., deWaard, J. R. (2003) Biological identifications through DNA barcodes. *Proc. Royal. Soc. B.* 270: 313–321.
- Jeffery, W. R. (2001) Cavefish as a model system in evolutionary developmental biology. *Dev. Biol.* 201, 231: 1–12.
- Kavalco, K. F., Almeida-Toledo, L. F. (2007) Molecular cytogenetics of blind Mexican tetra and comments on the karyotypic characteristics of genus *Astyanax* (Teleostei, Characidae). *Zebrafish.* 4: 103–111.
- Kavalco, K. F., Pazza, R., Brandão, K. O., Garcia, C., Bertollo, L. A. C., Almeida-Toledo, L. F. (2016) Chromosomal diversification higher than molecular variation in *Astyanax aff. Fasciatus* (Teleostei, Characidae). *Zebrafish.* 13: 345–353.
- Kocher, T. D. (2004) Adaptive evolution and explosive speciation: the cichlid fish model. *Nature Rev. Genet.* 5: 288–298.
- Levan, A., Fredga, K., Sandberg, A. A. (1964) Nomenclature of centromeric position on chromosomes. *Hereditas.* 52: 201–220.
- Lu, M., Szmids, A. E., Wang, X. (1995) Inheritance of RAPD fragments in haploid and diploid tissues of *Pinus sylvestris* (L.). *Heredity.* 74: 582–589.
- Mallet, J. (2005) Hybridization as an invasion of the genome. *Tren. Ecol. Evol.* 20: 229–237.
- Mallet, J. (2007) Hybrid speciation. *Nature.* 446: 279–283.
- Meyer, C. P., Paulay, G. (2005) DNA Barcoding: error rates based on comprehensive sampling. *Plos. Biol.* 3: 2229–2238.
- Moreira, Filho, O., Bertollo, L. A. C. (1991) *Astyanax scabripinnis* (Pisces, Characidae): a species complex. *Brazilian J. Genet.* 14: 331–357.
- Moritz, C., Cicero, C. (2004) DNA Barcoding: Promise and Pitfalls. *Plos. Biol.* 2: e354.
- Neff, N. A., Smith, G. R. (1979) Multivariate analysis of hybrid fishes. *System Biol.* 28: 176–196.
- Nirchio, M., Fenocchio, A. S., Swarça, A. C., Pérez, J. E., Granado, A., Estrada, A., Ron, E. (2003) Cytogenetic characterization of hybrids offspring between *Colossoma macropomum* (Cuvier, 1818) and *Piaractus brachipomus* (Cuvier, 1817) from Caicara del Orinioco, Venezuela. *Caryologia.* 56: 405–411.
- Ornelas-Garcia, C. P., Domínguez-Domínguez, O., Doadrio, I. (2008) Evolutionary history of the fish genus *Astyanax* Baird & Girard (1854) (Actinopterygii, Characidae) in Mesoamerica reveals multiple morphological homoplasies. *BMC Evol. Biol.* 8: 340.
- Pansonato-Alves, J. C., Hilsdorf, A. W. S., Utsunomia, R., Silva, D. M. Z. A., Oliveira, C., Foresti, F. (2013) Chromosomal mapping of repetitive DNA and Cytochrome c Oxidase I sequence analysis reveal differentiation among sympatric samples of *Astyanax fasciatus* (Characiformes, Characidae). *Cytogenet. Genome Res.* 141: 133–142.
- Pazza, R., Argento, R. F. M., Mariano, I. H. S., Kavalco, K. F. (2016) Phenotypic plasticity and karyotype conservation in allopatric populations of *Astyanax aff. Fasciatus* (Teleostei, Characiformes) at the Upper Paraná river basin. *Acta Zool.* 97: 241–245.
- Pazza, R., Kavalco, K. F. (2007) Chromosomal evolution in the Neotropical characin *Astyanax* (Teleostei, Characidae). *Nucleus.* 50: 519–543.

- Pazza, R., Kavalco, K. F., Bertollo, L. A. C. (2006) Chromosome polymorphism in *Astyanax fasciatus* (Teleostei, Characidae). 1. Karyotype analysis, Ag-NORs and mapping of the 18S and 5S ribosomal genes in sympatric karyotypes and their possible hybrid forms. *Cytogenet. Genome Res.* 112: 313-319.
- Pazza, R., Kavalco, K. F., Prioli, S. M. A. P., Prioli, A. J., Bertollo, L. A. C. (2007) Chromosome polymorphism in *Astyanax fasciatus* (Teleostei, Characidae), Part 3: Analysis of the RAPD and ISSR Molecular markers. *Biochem. Syst. Ecol.* 35: 843-851.
- Pazza, R., Kavalco, S. A. F., Penteado, P. R., Kavalco, K. F., Almeida-Toledo, L. F. (2008) The species complex *Astyanax fasciatus* Cuvier (Teleostei, Characiformes) – a multidisciplinary approach. *J. Fish. Biol.* 72: 2002-2010.
- Penteado, P. R., Kavalco, K. F., Pazza, R. (2013) Um terceiro citótipo de *Astyanax aff. Fasciatus* (Teleostei, Characidae) da bacia do rio São Francisco. *Evol. Cons. Biodivers.* 4: 1-7.
- Prado, F. D., Hashimoto, D. T., Senhorini, J. A., Foresti, F., Porto-Foresti, F. (2012a) Detection of hybrids and genetic introgression in wild stocks of two catfish species (Siluriformes: Pimelodidae): The impact of hatcheries in Brazil. *Fisher Res.* 125-126: 300-305.
- Prado, F. D., Nunes, T. L., Senhorini, J. A., Bortoluzzi, J., Foresti, F., Porto-Foresti, F. (2012b) Cytogenetic characterization of F1, F2 and backcross hybrids of the Neotropical catfish species *Pseudoplatystoma corruscans* and *Pseudoplatystoma reticulatum* (Pimelodidae, Siluriformes). *Genetic Mol. Biol.* 35: 57-64.
- Rieseberg, L. H., Raymond, O., Rosenthal, D. M., Lai, Z., Livingstone, K., Nakazato, T., Durphy, J. L., Schwarzbach, A. E., Donovan, L. A., Lexer, C. (2003) Major ecological transitions in wild sunflowers facilitated by hybridization. *Science.* 29: 1211-1216.
- Rossini, B. C., Oliveira, C. A. M., Melo, F. A. G., Bertaco, V. A., Astarloa, R. N. D., Rosso, J. J. et al., (2016) Highlighting *Astyanax* species diversity through DNA Barcoding. *Plos. One.* 11: e0167203.
- Salzburger, W., Baric, S., Sturmbauer, C (2002) Speciation via introgressive hybridization in East African cichlids. *Mol. Ecol.* 11: 619-625.
- Santos, V. H., Foresti, F., Oliveira, C., Almeida-Toledo, L. F., Toledo-Filho, S. A., Bernardino, G. (2002) Synaptonemal complex analysis in the fish species *Piaractus mesopotamicus* and *Colossoma macropomum*, and in their interspecific hybrid. *Caryologia.* 55: 73-79.
- Scribner, K. T., Page, K. S., Bartron, M. L. (2001) Hybridization in freshwater fishes: a review of case studies and cytonuclear methods of biological inference. *Rev. Fish. Biol. Fisher.* 10: 293-323.
- Svärdson, G. (1970) Significance of introgression in coregonid evolution. *Biology of Coregonid fishes.* 33-59.
- Ward, R. D., Zemlak, T. S., Innes, B. H., Last, P. R., Hebert, P. D. N. (2005) DNA barcoding Australia's fish species. *Philos Trans Royal Soc. B.* 360: 1847–1857.
- Wilkins, H. (1971) Genetic interpretation of regressive evolutionary processes: studies on hybrid eyes of two *Astyanax* cave populations (Characidae, Pisces). *Evolution.* 25: 530-544.

Chapter 58

TRICHORHINOPHALANGEAL SYNDROME (TRPS) REVISITED

N. Selenti*, M. Tzetis, E. Tsoutsou and H. Fryssira

Department of Medical Genetics, National and Kapodistrian University
of Athens, School of Medicine, Aghia Sophia Children's Hospital,
Athens, Greece

ABSTRACT

There are three distinct subtypes of Trichorhinophalangeal syndrome (TRPS); TRPS type I, TRPS type II and TRPS type III. Features common to all three subtypes include sparse, slowly growing scalp hair, laterally sparse eyebrows, a bulbous tip of the nose (pear-shaped), and protruding ears. The diagnosis of TRPS is based in typical clinical and radiographic features, as well as in the identification of causative mutations in the *TRPS1* gene (TRPS type I and III) and in loss of functional copies of the *TRPS1* and *EXT1* genes (TRPS type II). The mode of inheritance in TRPS I and III is autosomal dominant, however *de novo* deletions of *TRPS1* and *EXT1* genes are the main defect of TRPS II. Parental balanced chromosomal rearrangements are an important cause of interstitial aberrations in TRPS. In cases of cytogenetically-invisible alterations, parental FISH analysis as well as aCGH should be considered as part of the clinical baseline testing. Treatment of clinical problems of TRPS types is mainly supportive and includes ectodermal and skeletal issues. The number of distinct syndromes as TRPS is rapidly increasing and their confirmation is necessary especially in cases that typical features are absent. Clinical geneticists should provide information for the families and advise them how to overcome problems.

Keywords: trichorhinophalangeal syndrome (TRPS), genotype/phenotype of TRPS patients, differential diagnosis of TRPS

* Corresponding Author's Email: nikoletta_selenti@hotmail.com.

INTRODUCTION

Trichorhinophalangeal Syndrome (TRPS) is a rare inherited multisystem disorder. This is an autosomal dominant malformation syndrome, which is characterized by craniofacial and skeletal abnormalities, although more than half of the reported cases are sporadic. TRPS exhibits almost complete penetrance, but variable expressivity. Phenotype can vary depending on age and gender, even among patients carrying identical mutations, affected members of the same family, or in monozygotic twins. There are three distinct subtypes TRPS I, II and III and their diagnosis is based on clinical and radiographic features as well as on genetic analysis that is helpful especially in the case of non-classical clinical presentation [1-3].

Phenotype of TRPS Type I, II and III Patients

Clinical features that are common to all subtypes include sparse scalp hair, lateral eyebrows, bulbous tip of the nose, long flat philtrum, thin upper vermilion border and protruding ears [1, 3].

TRPS Type I Patients

TRPS type I is caused by mutations in the *TRPS1* gene located on chromosome 8q24.1 and characterized by distinctive skeletal abnormalities and craniofacial dysmorphism. The name of the condition describes some of the areas of the body that are commonly affected: hair (tricho-), nose (rhino-), and fingers and toes (phalangeal) [3, 4].

TRPS I patients, besides the common clinical features, have abnormalities of the skin, teeth, sweat glands, and nails. Skeletal abnormalities include cone-shaped epiphyses at the phalanges, hip malformations and short stature. Affected individuals often have short feet with toenails and fingernails typically thin and abnormally formed. Hip dysplasia often develops in early adulthood but can occur in infancy or childhood. Children with TRPS I often present hypermobility in many of their joints, however, the joints may degenerate leading to joint pain and a limited range of joint movement [3-5].

TRPS Type II Patients

TRPS II or Langer–Giedion syndrome (LGS) is considered to be an autosomal dominant condition however most cases of TRPS II are not inherited, but occur as random events *-de novo-* during the formation of reproductive cells in a parent of an affected individual [6].

The characteristic phenotype of individuals with TRPS II includes sparse scalp hair, thick eyebrows, bulbous tip of the nose, long flat philtrum, thin upper vermilion border and small teeth either with oligodontia or are supernumerary. Most individuals with TRPS II have mild intellectual disability [5, 6].

Bone and joint malformations are present too. Affected individuals may develop a few to several hundred osteochondromas that can cause pain, limited range of joint movement, or damage to the spinal cord, depending on the location of the osteochondromas. These bone growths typically begin from infancy to early childhood and stop forming around adolescence. TRPS II patients may also exhibit osteopenia as well as short stature, cone-shaped epiphyses at

the phalanges, an unusually large range of joint movement (hypermobility) and hip malformations [3, 6-8].

TRPS Type III Patients

Trichorhinophalangeal syndrome type III (TRPS3), also known as Sugio-Kajii syndrome, is considered as an autosomal dominant inheritance and characterized by bulbous tip of the nose, long flat philtrum, thin upper vermilion border, thin light-colored hair and delayed eruption of teeth. Skeletal abnormalities include cone-shaped epiphyses at the phalanges, brachydactyly, metacarpophalangeal shortening, skeletal dysplasia and short stature. Affected individuals may also exhibit additional skeletal abnormalities including osteochondritis, thoracic scoliosis, abnormal prominence of the breast bone (pectus carinatum), and/or limited movements of certain joints. In addition, some females affected by this disorder may exhibit hip malformations. The range and severity of symptoms may vary from case to case [3, 5, 9].

GENETICS

TRPS type I and III are caused by mutations in the *TRPS1* gene, which is located on 8q24.1. TRPS III represents only the extreme clinical manifestations of TRPSI. Missense mutations in exon 6 of *TRPS1* gene are the most common molecular defect identified in TRPSIII [2].

The *TRPS1* gene encodes a large nuclear transcription factor, protein *TRPS1*, which consists of 1281 amino acids. It binds specifically to GATA sequences and represses expression of GATA-regulated genes at selected sites and stages in vertebrate development. In its structure it combines nine potential zinc-finger motifs of four different types, including DNA-binding GATA motif and IKAROS-like zinc-finger motif, which mediates the transcriptional repressive function at the carboxy terminus. *In vitro* studies demonstrated that the presence of the DNA-binding domain is indispensable for the *TRPS1* repression function. *TRPS1* also regulates chondrocyte proliferation and differentiation and executes multiple functions in proliferating chondrocytes, expanding the region of distal chondrocytes, activating proliferation in columnar cells and supporting the differentiation of columnar into hypertrophic chondrocytes. *TRPS1* transcription factor is a novel transcriptional repressor of the *RUNX2* promoter by identifying single nucleotide variation. *RUNX2* is the master regulator of osteoblast differentiation and is required for chondrocyte hypertrophy. *TRPS1* repression activity is necessary for the timely progression of chondrocyte maturation and synchronization of chondrocyte development with perichondrial mineralization [2, 10, 11].

TRPS type I is mainly caused by entire *TRPS1* deletions and nonsense mutations located before the coding region for the DNA-binding domain. TRPS type III is caused by missense mutations in the DNA-binding domain [10].

The LGS/TRPS type II is caused by contiguous deletion of the *TRPS1* and *EXT1* (exostosin-1) genes. The multiple hereditary exostoses (MHE) type I that is present in TRPS type II is caused by a mutation in *EXT1*. The MHE is a genetically heterogeneous disorder which can be caused by mutations in the *EXT1*, *EXT2* or *EXT3* gene. *EXT1* protein is a glycosyltransferase required for the biosynthesis of heparan-sulfate that is found in Golgi apparatus. It modifies newly produced enzymes and other proteins. *EXT1* protein binds to

EXT2 protein to form a complex, because the EXT1/EXT2 complex possesses substantially higher glycosyltransferase activity than EXT1 or EXT2 alone [12, 13].

Although deletions of the 8q region are rare and considerably variable, the shortest region of deletion overlap (SRO) has been defined at 8q24.1, spanning 2 Mb, including *TRPS1*, *EIF3S3*, *RAD21*, *OPG*, *CXIV* and *EXT1* genes [14].

Cornelia de Lange syndrome-4 (CdLs-4) that is characterized by growth deficiency, mental retardation, microcephaly, bushy eyebrows and synophrys, depressed nasal bridge, micrognathia, micromelia, hearing loss, anteverted nares, prominent symphysis and spurs in the anterior angle of mandible and gastrointestinal problems is caused by heterozygous mutations of the *RAD21* gene [15, 16].

TRPS type I, LGS/TRPS type II and CdLs-4 patients share many facial features. There are TRPSII patients where deletions in 8q24.1 region have been identified involving *RAD21* and not *TRPS1*. The phenotypic consequences of molecular defects in *RAD21*, such as facial dysmorphisms, might be underestimated in these patients [1-3, 15, 16].

DIAGNOSIS

Clinical Diagnosis

Facial, ectodermal features and skeletal findings are characteristics of the diagnosis of TRPS type I, LGS/TRPS type II and TRPS type III patients. For TRPS type II patients multiple osteochondromas and intellectual disability are also present in order to confirm the clinical diagnosis [1-3].

Genetic Diagnosis

Sequence analysis of *TRPS1* is initially performed. Chromosome microarray analysis (CMA), using either aCGH or SNP arrays, or gene-targeted deletion/duplication analysis is recommended in the case of no pathogenic variant detection in *TRPS1* [2, 4, 10].

Since haploinsufficiency of *TRPS1* is responsible for the phenotype of TRPS type I and deletion of copies of the *TRPS1* and *EXT1* is responsible for the phenotype of TRPS type II, karyotype may be recommended in order to detect apparently balanced translocation. Most chromosomal aberrations are unbalanced, resulting in duplication or deletion of genetic material at the chromosomal breakpoints. The frequency of balanced chromosomal insertions is much lower than reciprocal translocations because it requires three chromosomal breaks. Parental balanced insertions increase the possibility of copy number imbalances in the offspring resulting in congenital anomalies of multiple organ systems and psychomotor retardation. aCGH analysis revealed that approximately 2.1% of parents with offspring having developmental anomalies, harbor balanced insertions. In this regard, genomic analysis of the parental DNA of a patient may help us to exclude the possibility of parental balanced translocations as predisposing genetic risk [17, 18].

DIFFERENTIAL DIAGNOSIS

TRPS is often considered in the differential diagnosis of disorders with abnormalities of the hair, nose, and limbs (Table 1) [19-21].

Table 1. Disorders to consider in case of the differential diagnosis of the TRPS

Disorder	Genes	Mode of Inheritance	Clinical Features	
			Overlapping with TRPS	Distinguishing from TRPS
Oculo-dento-digital syndrome	<i>GJA1</i>	Autosomal Dominant	• Slow-growing, dry hair • Underdeveloped alae nasi • Long Philtrum	• Eye deformities • Dental abnormalities
Cartilage-hair syndrome	<i>RMRP</i>	Autosomal Recessive	• Short stature • Fine hair • Cone-shaped epiphyses	• Nasal shape • Immunodeficiency
Ellis-Van Creveld Syndrome	<i>EVC1</i> , <i>EVC2</i>	Autosomal Recessive	• Short stature • Brachydactyly	• Nasal shape • Oral frenula • Polydactyly

TREATMENT

The treatment of TRPS is mainly supportive. Concerning the ectodermal features, practical advice on hair care and the extraction of supernumerary teeth can be considered. Human growth hormone therapy may be recommended for skeletal disorders such as short stature, however reported results vary. Regular simple analgesia is the mainstay treatment for joint pain. Physiotherapy and occupational therapy may ameliorate the mobility. Prosthetic hip implantation should be recommended in those with severe hip dysplasia. Regarding osteopenia, bisphosphonates can be given in individuals with TRPS I and bone fragility. The resection of exostoses should be considered in TRPS patients with restricted range of motion, or nerve compression. Skeletal and orthopedic symptoms continue to be an important issue. Severe scoliosis due to joint laxity, Perthes disease, asymmetry of legs, joint immobility and adjacent exostoses are among the most frequent problems. On the other hand, eye, ear, and cardiac problems are rarely of major relevance [1, 5].

GENETIC COUNSELING

Because of the intrafamilial clinical variability, it is not possible to predict the exact phenotype in family members who have inherited a *TRPS1* pathogenic variant or a deletion that includes *TRPS1*. However, once the genetic alteration that causes the TRPS has been identified in an affected family member, prenatal testing and/or Preimplantation Genetic Diagnosis (PGD) for TRPS is suggested. On the other hand, timely diagnosis of parental balanced

chromosomal rearrangements using the appropriate techniques is important and can reduce the risk of subsequent miscarriages as well as abnormal offspring [1, 5, 18].

With better knowledge on the long-term course, complications and molecular defects that cause TRPS type I, II and III, genetic counseling is necessary by clinical geneticists [5].

REFERENCES

- [1] Gorlin R. J., Cohen M. M. Jr., Wolfson J. Trichorhinophalangeal syndrome. *Am J Dis Child.* 1969, 595-599.
- [2] Lüdecke H.-J., Schaper J., Meinecke P., Momeni P., Groß S., von Holtum D., Hirche H., Abramowicz M. J., Albrecht B., Apacik C., Christen H.-J., Claussen U., Devriendt K., Fastnacht E., Forderer A., Friedrich U., Goodship T. H. J., Greiwe M., Hamm H., Hennekam R. C. M., Hinkel G. K., Hoeltzenbein M., Kayserili H., Majewski F., Mathieu M., McLeod R., Midro A. T., Moog U., Nagai T., Niikawa N., Ørstavik K. H., Plöchl E., Seitz C., Schmidtke J., Tranebjærg L., Tsukahara M., Wittwer B., Zabel B., Gillesen-Kaesbach G. and Horsthemke¹ B. Genotypic and Phenotypic Spectrum in Trichorhinophalangeal Syndrome Types I and III. *Am J Hum Genet.* 2001 January, 81-91.
- [3] Maas S. M., Shaw A. C., Bikker H., Lüdecke H. J., van der Tuin K., Badura-Stronka M., Belligni E., Biamino E., Bonati M. T., Carvalho D. R., Cobben J., de Man S. A., Den Hollander N. S., Di Donato N., Garavelli L., Grønberg S., Herkert J. C., Hoogeboom A. J., Jamsheer A., Latos-Bielenska A., Maat-Kievit A., Magnani C., Marcelis C., Mathijssen I. B., Nielsen M., Otten E., Ousager L. B., Pilch J., Plomp A., Poke G., Poluha A., Posmyk R., Rieubland C., Silengo M., Simon M., Steichen E., Stumpel C., Szakszon K., Polonkai E., van den Ende J., van der Steen A., van Essen T., van Haeringen A., van Hagen J. M., Verheij J. B., Mannens M. M., Hennekam R. C. Phenotype and genotype in 103 patients with trichorhinophalangeal syndrome. *Eur J Med Genet.* 2015 May, 279-292.
- [4] Candamourty R., Venkatachalam S., Karthikeyan B., Babu M. R. Trichorhinophalangeal syndrome type 1: A case report with literature review. *J Nat Sci Biol Med.* 2012 July, 209-211.
- [5] Maas S., Shaw A/, Bikker H/, Hennekam R. C. M. Trichorhino-phalangeal Syndrome. *Gene Reviews*[®], 2017 Apr, 1993-2017.
- [6] Shin H. T., Chang M. W. Trichorhinophalangeal syndrome, type II (Langer-Giedion syndrome). *Dermatol Online J.* (2001).
- [7] Tsang W. K., Yang K. W., Fong C. M. Langer-Giedion syndrome: the evolving imaging features in hands and beyond. *Skeletal Radiol.* 43 (2014) 251-255.
- [8] Riedl S., Giedion A., Schweitzer K., Müllner-Eidenböck A., Grill F., Frisch H., Lüdecke H. J. Pronounced short stature in a girl with trichorhinophalangeal syndrome II (TRPS II, Langer-Giedion syndrome) and growth hormone deficiency. *Am J Med Genet A.* 2004 Dec, 200-203.

- [9] Niikawa N., Kamei T., The Sugio-Kajii syndrome, proposed trichorhinophalangeal syndrome type III, *Am. J. Med. Genet.* 278-280.
- [10] Solc Roman, Klugerova Michaela, Vcelak Josef, Baxova Alice, MiloslavKuklik, JanVseticka, Rastislav Beharka, Katerina Hirschfeldova, Mutation analysis of *TRPS1* gene including core promoter, 5'UTR, and 3'UTR regulatory sequences with insight into their organization., *Clinica Chimica Acta* 2017, 30-36
- [11] Piccione M., Niceta M., Antona V. et al. Identification of two new mutations in *TRPS 1* gene leading to the trichorhinophalangeal syndrome type I and III. *Am J Med Genet A.* 149A (2009) 1837-1841.
- [12] Lüdecke H. J., Wagner M. J., Nardmann J., La Pillo B., Parrish J. E., Willems P. J., Haan E. A., Frydman M., Hamers G. J., Wells D. E., Horsthemke B. Molecular dissection of a contiguous gene syndrome: Localization of the genes involved in the Langer-Giedion syndrome. *Hum Mol Genet* (1995) 31-36.
- [13] Lüdecke H. J., Schmidt O., Nardmann J., von Holtum D., Meinecke P., Muenke M., Horsthemke B. Genes and chromosomal breakpoints in the Langer-Giedion syndrome region on human chromosome 8. *Hum Genet* 105 (1999) 619-628.
- [14] Lüdecke H. J., Johnson C., Wagner M. J., Wells D. E., Turleau C., Tommerup N., Latos-Bielenska A., Sandig K. R., Meinecke P., Zabel B., Horsthemke B. Molecular definition of the shortest region of deletion overlap in the Langer-Giedion syndrome. *Am J Hum Genet* 49 (1991) 1197-1206.
- [15] Chen C. P., Lin S. P., Liu Y. P., Chern S. R., Wu P. S., Su J. W., Chen Y. T., Lee C. C., Wang W. An interstitial deletion of 8q23.3-q24.22 associated with Langer-Giedion syndrome, Cornelia de Lange syndrome and epilepsy. *Gene.* 15 (2013) 176-180.
- [16] Dearnorff M. A., Wilde J. J., Albrecht M., Dickinson E., Tennstedt S., Braunholz D., Mönnich M., Yan Y., Xu W., Gil-Rodríguez M. C., Clark D., Hakonarson H., Halbach S., Michelis L. D., Rampuria A., Rossier E., Spranger S., Van Maldergem L., Lynch S. A., Gillesen-Kaesbach G., Lüdecke H. J., Ramsay R. G., McKay M. J., Krantz I. D., Xu H., Horsfield J. A., Kaiser F. J. *RAD21* mutations cause a human cohesinopathy. *Am J Hum Genet.* 90 (2012) 1014-1027.
- [17] Kang S. H., Shaw C., Ou Z., Eng P. A., Cooper M. L., Pursley A. N., Sahoo T., Bacino C. A., Chinault A. C., Stankiewicz P., Patel A., Lupski J. R., Cheung S. W., Insertional translocation detected using FISH confirmation of arraycomparative genomic hybridization (aCGH) results, *American Journal of Medical Genetics.* Part A 152A (2010) 1111-1126.
- [18] Nowakowska B. A., de Leeuw N., Ruivenkamp C. A., Sikkema-Raddatz B., Crolla J. A., Thoelen R., Koopmans M., den Hollander N., van Haeringen A., van der Kevie-Kersemaekers A. M., Pfundt R., Mieloo H., van Essen T., de Vries B. B., Green A., Reardon W., Fryns J. P., Vermeesch J. R., Parental insertional balanced translocations are an important cause of apparently *de novo* CNVs in patients with developmental anomalies, *European Journal of Human Genetics* 20 (2012) 166-170.
- [19] Frasson M., Calixto N., Cronemberger S., de Aguiar R. A., Leão L. L., de Aguiar M. J., Oculodentodigital dysplasia: study of ophthalmological and clinical manifestations in

three boys with probably autosomal recessive inheritance. *Ophthalmic Genet.* 2004 Sep, 227-236.

[20] Matesic D., Hagan J. B. Cartilage-hair hypoplasia. *Mayo Clin Proc.* 2007 Jun.

[21] Baujat G., Le Merrer M. Ellis-van Creveld syndrome. *Review. Orphanet J Rare Dis.* 2007 Jun.

Chapter 59

EPIGENETIC MODIFICATIONS AND DEVELOPMENTAL ORIGIN OF HEALTH AND DISEASES (DOHAD)

Jia Zheng and Xinhua Xiao*

Department of Endocrinology, Key Laboratory of Endocrinology, Ministry of Health,
Peking Union Medical College Hospital, Diabetes Research Center of Chinese Academy
of Medical Sciences & Peking Union Medical College, Beijing, China

Department of Endocrinology, Peking Union Medical College Hospital, Chinese Academy of
Medical Sciences & Peking Union Medical College, Beijing, China

ABSTRACT

It is generally believed that genotype and adult lifestyle elements are primary risks of some metabolic diseases such as insulin resistance, obesity and diabetes mellitus in later life. However, increasing evidence demonstrates that early life malnutrition during the period of gestation and/or lactation may increase our susceptibility to such metabolic diseases in later life. The underlying mechanism is still not very clear. Recently, epigenetics is hypothesized to be the important molecular basis of the imbalanced early life nutrition and glucose metabolism disorders, which is known as "Developmental Origin of Health and Diseases" (DOHAD). Currently, there are substantial epidemiological studies and experimental animal models that have demonstrated nutritional disturbances during the critical periods of early life development can significantly impact the predisposition to developing some metabolic diseases in later life. The fundamental mechanism is that early developmental nutrition can regulate epigenetic modifications of some genes associated with development and metabolism. DNA methylation is the first discovered and one important epigenetic modification. MicroRNAs are recognized as an important epigenetic modification and they are a major class of small non-coding RNAs (about 20-22 nucleotides) which can mediate posttranscriptional regulation of target genes with cell differentiation and apoptosis. Recent studies suggest that DNA methylation and microRNAs maybe the crucial modulators of fetal epigenetic programming in nutrition and metabolic disorders. This chapter will focus on how early life nutrition can alter the epigenome, produce different phenotypes and alter disease susceptibilities, especially for impaired glucose metabolism.

* Corresponding Author's Email: xiaoxinhua@medmail.com.cn (Tel: +86-10-69155073).

Keywords: epigenetics, DNA methylation, microRNAs, early life nutrition, glucose metabolism

ABBREVIATIONS

AUC:	Area Under the ROC Curve;
CVD:	Cardiovascular diseases;
DMR:	Differentially methylated region;
DOHaD:	Developmental Origin of Health and Diseases;
GR:	Glucocorticoid receptor;
GDM:	Gestational diabetes mellitus;
F1:	First-generation;
FGR:	Fetal growth restriction;
HMGCR:	3-hydroxy-3-methylglutaryl coenzyme A reductase;
HNF4 α :	Hepatocyte nuclear factor 4 receptor α ;
IDF:	International Diabetes Federation;
IGF2:	Insulin like growth factor 2;
IGT:	Impaired glucose tolerance;
IL13ra2:	Interleukin 13 receptor alpha 2;
IUGR:	Intrauterine growth restriction;
LncRNAs:	Long non-coding RNAs;
LP:	Low protein;
MeCP2:	Methyl CpG-binding protein 2;
MOR:	μ -opioid receptor;
PPAR α :	Peroxisome proliferator-activated receptor alpha;
RXRA:	Retinoid X receptor- α ;
SAM:	S-adenosylmethionine;
T2DM:	Type 2 diabetes mellitus.

1. INTRODUCTION

Nowadays, the prevalence of diabetes mellitus is increasing very rapidly worldwide, which is now considered a pandemic non-communicable disease. According to the Diabetes Atlas published by the International Diabetes Federation (IDF) on World Diabetes Day in 2017, 387 million people are suffering from diabetes and the number will rise to 592 million by 2035, which implies one twelfth of the world's population will endure diabetes. In particular, more than 21 million infants acquired diabetes from their mother during pregnancy in 2013. It is estimated that diabetes caused 4.9 million deaths in 2014, that is to say, a person died from the disease every seven seconds [1]. In summary, diabetes is becoming a more and more severe problem for our society.

Unfortunately, the pathogenesis of diabetes has not been clearly understood yet. Although it is generally believed that genes together with adult lifestyle factors are major risks of diabetes. There is substantial evidence that the prenatal and early postnatal nutrition play a key

role in determining our susceptibility to diabetes in later life [2]. Little is known about the molecular mechanisms underlying the interaction between maternal diet and aging. However, the hypothesis that epigenetic mechanisms may link such imbalanced nutrition with altered disease risk has been widely accepted in recent years. Epigenetics is one of the major mechanisms explaining the theory of "Developmental Origin of Health and Diseases" (DOHaD), which focuses on the association between perinatal nutrition and late-onset disease, such as obesity, insulin resistance, impaired glucose tolerance (IGT) and type 2 diabetes mellitus (T2DM) [3]. Therefore, epigenetics is likely to be an important molecular basis of malnutrition during early life and glucose metabolism disorders in later life.

2. WHAT IS EPIGENETICS?

Traditionally, epigenetics has been used to explain the phenotypic events which cannot be explained by genetic mechanisms. The term "epigenetics" was first proposed by Waddington in 1942 and defined as a branch of biology that studied the causal interactions between genes and their products [4]. In 2006, the definition of epigenetics was updated and considered as a mechanism that could affect gene expression without altering the nucleotide sequence. Most importantly, it could be inherited between generations steadily by mitosis and meiosis through cell differentiation [5]. Epigenetic conditions can illustrate the reason why an organism produces many different cell types during its development, despite the fact that most of the cells in a multicellular organism share the same genetic information. Epigenetic modifications can regulate gene expression reversibly. The three major epigenetic processes are DNA methylation, histone modifications and non-coding RNAs. DNA methylation was the first recognized and the most well-characterized epigenetic modification in the 1970s [6]. Histone modification can influence gene expression by altering chromatin structure [7]. Non-coding RNAs, such as microRNA and long non-coding RNA, are also included in epigenetic modifications [8]. Epigenetic processes play a critical role in normal development and differentiation in mammals [9]. Recently, epigenetics was defined as "molecular factors and processes around DNA were mitotically stable and regulated genome activity which were independent of DNA sequence" by Skinner et al. in 2010 [10]. Although the definition of epigenetics has been updated constantly, the three core features are always reserved, namely: (1) without any alterations in DNA sequence (2) heritability (3) plasticity and reversibility. In this chapter, we will focus on DNA methylation and microRNAs.

2.1. DNA Methylation

DNA methylation is the first discovered epigenetic modification and it is also one of the best-studied epigenetic modifications in the context of altered environment. It is a biochemical process involving the covalent addition of a methyl group at the 5' position of cytosine in DNA. This normally occurs on a cytosine followed by a guanine, known as a CpG dinucleotide (the p denotes the intervening phosphate group). It is catalyzed by DNA methyltransferases, and S-adenosylmethionine (SAM) is the methyl donor [11]. DNA methylation typically occurs in CpG dinucleotide context. CpG dinucleotides are not randomly distributed throughout the

genome but they are often grouped in clusters at the 5' regulatory regions of many genes, called CpG islands. Most CpG dinucleotides are methylated, but those located in CpG islands are usually unmodified [12]. However, hypermethylation of these CpG islands has the specific effect on reducing gene expression and repressing transcription, while hypomethylation of CpG islands is related to transcriptional activation [13]. DNA methylation is a common modification in mammalian genomes. It constitutes a stable epigenetic symbol that can stably alter the expression of genes and transmit through DNA replication as cells divide and differentiate from embryonic stem cells into specific tissues [11]. DNA methylation is essential for normal development and is associated with a number of key processes including genomic imprinting, X-chromosome inactivation [14], suppression of repetitive elements and carcinogenesis [15].

2.2. MicroRNAs

Non-coding RNAs are the latest discovered epigenetic modifications including microRNAs and long non-coding RNAs (LncRNAs) [8]. MicroRNAs (miRNAs) are gradually recognized as an important epigenetic modification in recent years. They are a major class of small non-coding RNAs with about 20-22 nucleotides which can mediate posttranscriptional regulation of gene expressions. More precisely, miRNAs can bind to the 3' untranslated regions (3'-UTR) of target genes specifically, resulting in either translation inhibition or mRNA degradation [16]. They are complicated and multi-functional that up to 60% of our genome is regulated by miRNAs and each miRNA can probably regulate several hundred gene expressions. The modulations of miRNAs can facilitate some key developmental processes such as cell proliferation, cell line differentiation and programmed cell apoptosis [8]. Therefore, we will focus on the regulations of microRNAs between fetal epigenetic programming in nutrition and glucose metabolism.

3. EVIDENCE FROM HUMAN COHORTS

3.1. Fetal Programming Hypothesis

In 1977, the association between the quality of early life environment and future risk of disease in later life was first proposed by Forsdahl, who discovered that infant mortality rate was positively associated with an increased risk of cardiovascular diseases (CVD) in middle age [17]. Subsequently, Barker and his colleagues found an inverse relationship between birth weight and increased CVD mortality in 1989 [18]. The effects of early life nutrition on the increased risk of metabolic diseases were the most clearly shown in studies of the Dutch famine during the winter of 1944. These studies showed that the individuals whose mothers were exposed to famine in the late stage of gestation had lower birth weight and an increased risk of obesity, CVD, insulin resistance and hypertension in later life compared with unexposed individuals [19]. What is more, apart from nutrition deficiency in early life, overnutrition is also associated with an significantly increased susceptibility to metabolic disease such as obesity, hypertension, atherosclerosis, insulin resistance and diabetes mellitus [20]. It can involve some organs related to glucose metabolism, including liver, pancreas, adipose tissue

and skeletal muscle through different mechanisms. The nutrition during early life has long term phenotypic effects. This idea was first proposed by professor Barker and Hales in the 1990s, known as “fetal programming hypothesis” [21]. After more than 20-year long term research, it has now been accepted worldwide.

3.2. DNA Methylation and Fetal Programming Hypothesis

Epigenetic events are crucial for early development. However, it can be influenced by some environmental factors, potentially programming the genome for later adverse health outcomes. Gestation is the critical time window for maternal nutrition to affect the offspring. Early life nutrition can induce persistent DNA methylation. More specifically, both under-nutrition and over-nutrition will bring about epigenetic modification during the early life, including the embryonic development and neonatal period. A number of clinical studies have shown that people born during the famine in the Netherlands still exhibited lower methylation level of insulin like growth factor 2 (*IGF2*) after 60 years, compared with those who had no exposure to prenatal famine [22]. It indicates that epigenetic modification of some special genes affected by periconceptional maternal nutrition restriction can be kept for decades. Another study showed that the children whose mother used folic acid had a 4.5% higher methylation in the *IGF2* gene differentially methylated region (DMR) than those whose mother did not take folic acid. It also indicates an inverse independent association between *IGF2* DMR and birth weight, which means that periconceptional folic acid use is associated with epigenetic changes in *IGF2* in the child that may affect intrauterine programming of growth and development with consequences for health and disease throughout life [23]. A recent cohort study demonstrated that retinoid X receptor- α (*RXR α*) methylation had independent association with sex-adjusted childhood fat mass and it could explain more than 25% of the variance in childhood adiposity. Higher methylation of *RXR α* was associated with lower maternal carbohydrate intake in early pregnancy. These researches suggest that prenatal development is a substantial risk of metabolic disease, such as diabetes, obesity and metabolic syndrome [24].

3.3. MicroRNAs and Fetal Programming Hypothesis

It is well known that nutrition during early life development can impact the health of the adult permanently. Fetal growth restriction (FGR), also known as intrauterine growth restriction (IUGR), is a relatively common, pleiotropic complication of pregnancy. It is not only associated with significantly increased perinatal morbidity and mortality, and is also a major determinant of cardiovascular disease and glucose intolerance in adult life. Some clinical studies have indicated that some programmed changes in miRNAs expression may regulate the pathophysiology of intrauterine growth. The placenta is a crucial organ which can regulate the exchange of fetomaternal nutrients for the developing fetus, thereby modulating intrauterine development. More specifically, recent studies have showed that epigenetic modifications of placenta may play an important role on the health of the offspring and maternal nutrition. Some studies showed that several miRNAs in placentas are involved in pre-eclampsia, preterm labor, FGR and newborn neurobehavior, such as miRNA-15b, miRNA-181a, miRNA-210, miR-377, miR-483-5p and miR-493 [25, 26]. Huang et al. indicated that the levels of miRNA-424 were

significantly increased in placentae from women with FGR [27]. Tang et al. also observed that aberrant high expression level of miR-141 might play an important role in the pathogenesis of IUGR which can suppress the expressions of target genes, such as, E2F transcription factor 3 and pleiomorphic adenoma gene 1. Furthermore, they also discovered that miR-141 could serve as a potential biomarker to distinguish between FGR and normal controls with the Area Under the ROC Curve (AUC) of 0.839 and the 88.5% sensitivity and 71.7% specificity [28]. Ferland-McCollough et al. showed that miR-483-3p is upregulated in adipose tissue rather than placentas from low birth weight adult humans [29]. In addition to IUGR, miRNAs programming were also existed in fetal macrosomia. A recent research indicated that aberrant expression of miRNA-21 in placenta is associated with macrosomia which may similarly have long-term effects on health and disease in adult life [30, 31]. Thus, all these evidence support the modifications of miRNAs in the fetal programming.

4. EVIDENCE FROM ANIMAL EXPERIMENTS

In addition to the evidence from studies of human cohorts, the association between dietary changes during specific windows of development and epigenomic modifications in later life has also been reported in several animal models. In this chapter, we will describe some well-known epigenetic changes that have been identified from animal models of obesity, insulin resistance and diabetes.

4.1. DNA Methylation and Animal Experiments

4.1.1. Maternal Nutrition

Protein restriction is frequently used as a model for maternal malnutrition. Low protein (LP) diet is associated with impaired fetal growth, the development of obesity, insulin resistance and diabetes in the offspring [32]. Many epigenetic modifications have been reported in offspring exposed to a maternal LP diet. One of the first studies showing the link between nutritional imbalances during intrauterine development and epigenetic modifications was that feeding a LP diet (9% casein vs. 18% casein) to rats during pregnancy resulted in global DNA hypermethylation in the liver of the fetuses [33]. Recent studies confirmed that maternal LP feeding during pregnancy might also result in locus-specific changes in DNA methylation. For example, feeding pregnant rats a protein-restricted diet induced hypomethylation of the glucocorticoid receptor (*GR*) and Peroxisome proliferator-activated receptor alpha (*PPARα*) promoters in the livers of juvenile and adult offspring [34]. Also in Wistar rats, *PPARα* gene methylation was 20.6% lower with expression 10.5-fold higher and *GR* gene methylation was 22.8% lower with expression 200% higher in pups of dams that were fed protein restriction diet throughout pregnancy [35]. Hepatocyte nuclear factor 4 receptor α (*HNF4α*) gene has been implicated in the etiology of T2DM. Sandovici observed that maternal LP diet during pregnancy and lactation could lead to progressive epigenetic silencing of the entire *HNF4α* locus in rat pancreatic islets, which weakened the promoter-enhancer interaction and resulted in a permanent reduction in *HNF4α* expression [36]. Maternal protein-undernutrition during pregnancy and lactation in Balb/c mice affected the balance between food intake and energy

expenditure in adults. Moreover, this nutritional stress resulted in the removal of methyls at CpGs located in the promoter of leptin, which was associated with DNA hypomethylation of the leptin promoter in adipose tissue [37]. Interestingly, the DNA methylation of LP diet is not restricted to rodents. Male offspring of maternal LP diet of Meishan sows during pregnancy and lactation demonstrated higher *GR* binding to the mitochondrial DNA promoter, which was accompanied by lower cytosine methylation and hydroxymethylation on mitochondrial DNA promoter [38]. Similarly, piglets with maternal LP diet during pregnancy and lactation showed significantly lower body weight and liver weight at weaning. Hepatic activation of 3-hydroxy-3-methylglutaryl coenzyme A reductase (*HMGCR*) gene transcription in LP piglets was associated with promoter hypomethylation [39].

Nowadays, both the incidence of obesity during pregnancy and gestational diabetes mellitus (GDM) are increasing along with the improved living conditions. Therefore, scientists are increasingly concerned about overnutrition on human health. There is increasing evidence that caloric excess in maternal diet is associated with an altered metabolic phenotype. Recently, researches showed that maternal high fat diet might modify DNA methylation and gene expression in the offspring. Maternal high fat diet during gestation was associated with global and gene-specific promoter DNA hypomethylation, like the dopamine reuptake transporter, the μ -opioid receptor (*MOR*) and preproenkephalin in the brain from the first-generation (F1) offspring of C57BL/6J x DBA/2J hybrids. These changes might affect feeding behavior, which can increase obesity and obesity-associated risk for metabolic syndrome [40]. During the postnatal period of the same model, the offspring of maternal high fat diet showed increased binding of Methyl CpG-binding protein 2 (*MeCP2*) to the *MOR* promoter in reward-related regions of the brain, which can repress the transcription of *MOR* gene [41]. A recent study showed that the offspring whose mothers fed on high fat diet during late gestation had increased weight accrual and food intake, exhibiting insulin resistance and hyperlipidemia. Furthermore, the offspring mice emerged with increased methylation of adiponectin and leptin receptor, and decreased methylation of leptin genes [42].

4.1.2. Paternal Nutrition

Apart from the detrimental impacts of maternal malnutrition on abnormal glucose metabolism in offspring, recent studies showed that parental malnourished exposures could also affect the phenotype of the offspring. Paternal lifestyle and particular nutrition factors can affect spermatogenesis at the level of germ and sertoli cells [43] and the composition of seminal fluid [44]. Similarly, alterations in paternal diet are also associated with altered DNA methylation in the offspring. As shown by Ng et al., a paternal high fat diet consumption of SD rats exposure could induce intergenerational transmission of impaired glucose tolerance and insulin homeostasis in their female offspring. The interleukin 13 receptor alpha 2 (*Il13ra2*) promoter was hypomethylated in female offspring after high fat feeding of their fathers [45]. Offspring of fathers fed on a low protein diet in C57BL/6J mice exhibited elevated hepatic expression of many genes involved in glucose metabolism and lipid biosynthesis. Epigenomic bisulfite sequence analysis on DNA from the liver of offspring revealed numerous modest (about 20%) changes in cytosine methylation depending on paternal diet, including a substantial increase in methylation at an intergenic CpG island 50 kb upstream of the *PPAR α* gene [46]. A recent study showed that paternal prediabetes increased the susceptibility of the offspring to diabetes through gametic epigenetic alterations [47]. All these results indicate that parental diet

can also induce impaired glucose metabolism in offspring through the epigenetic modification mechanisms, especially, DNA methylation.

4.2. MicroRNAs and Animal Experiments

4.2.1. Maternal Nutrition

Not only the evidence can be observed in humans, the association between maternal nutrition and microRNAs modulations has also been reported in several animal experiments. One recent study suggested that maternal consumption of a high fat diet can induce glucose intolerant and insulin intolerant, obesity and the abnormal lipid metabolism in the early life of the mice offspring by downregulating the expression of miRNA-122 and upregulating miRNA-370, and then they can modulate the expression of target hepatic β -oxidation-related genes, hence, contributing to metabolic disturbances in adult life [48]. Maternal low protein diet which is an animal model of intrauterine growth restriction also can lead to abnormal glucose metabolism by modulating the expression of some miRNAs, for example, one recent research showed that gestational protein restriction in rats could lead to pancreatic failure in offspring which may result from the transgenerational transmission of miRNA-375 misexpression in β -cells [49]. Lie et al. suggested that maternal undernutrition around the time of conception in sheep could induce changes in the expression of about 22 miRNAs, which may have an influence on altering the abundance of the critical insulin-signaling molecules in skeletal muscle and in the association between maternal undernutrition in the periconceptional period and insulin resistance in adult life [50]. The same animal model also showed that maternal undernutrition around conception can impact the insulin signaling and gluconeogenic factors which can be explained by alterations in the expression of a number of specific candidate microRNAs in the liver of fetal sheep [51].

4.2.2. Paternal Nutrition

Not only maternal malnutrition can have a detrimental effect on glucose metabolism in the offspring, animal studies have also shown the potential importance of paternal diets in future metabolic diseases risk. Paternal lifestyle, particularly, nutrition condition can impact spermatogenesis at the level of germ and sertoli cells [43] and the composition of seminal fluid [44]. Similarly, miRNAs are also an important modulators in paternal diet and glucose metabolism in the offspring. One study indicated that offspring of whose father were fed a low protein diet from weaning until sexual maturity, showed decreased levels of cholesterol esters and elevated hepatic expression of many genes involved in lipid and cholesterol biosynthesis. Furthermore, it was exhibited that a number of miRNAs were also altered in the liver tissues of the offspring persistently by microarray hybridization, namely, miRNA-98, miRNA -21, let-7 and miRNA-199 were upregulated and miRNA-210 was downregulated, many of which are associated with cellular proliferation and target lipid biosynthesis genes [46]. Therefore, all these studies suggested that microRNAs could be the crucial modulators of fetal epigenetic programming in nutrition and glucose metabolism.

To conclude, the studies of human cohorts and animal models demonstrate that developmental programming of adult disease occurs at each nutrition spectrum during the early life, due to maternal caloric deprivation, maternal nutritional excess and paternal malnutrition.

Although the specific mechanisms leading to aberrant glucose metabolism in each situation are different, the current evidence supports that the epigenetic modification, especially, DNA methylation and microRNAs maybe an important molecular link of them, and this epigenome can be inherited from one generation to another.

5. DEVELOPMENTAL PLASTICITY AND REVERSIBILITY

Generally speaking, plasticity is viewed as the ability that one genotype can produce different phenotypes in response to different environments. The time window of maximal plasticity appears to be during development, which is termed as "Developmental plasticity". It has evolved to provide the best chances of survival and reproductive success to the organism. Further studies illustrated that developmental plasticity could be regulated by epigenetic modifications. It was traditionally considered that epigenetic modifications were static in the control of gene expression. However, it is now being altered by the idea that these marks are dynamic induced by some factors. Below are examples supporting the idea, for example, intervention with drugs in early life can bring long-term benefits to health. Offspring born to dams with calorie restriction during gestation and lactation are more liable to develop obesity and insulin resistance [52]. However, when leptin is given to neonatal mice born to dams suffered from calorie restriction, the mice show a phenotype close to healthy controls and do not develop obesity in later life even fed with an energy dense high fat diet [53]. It suggests that developmental metabolic programming is potentially reversible by an intervention late in the phase of developmental plasticity. The long-lasting effects of early leptin intervention has been shown to be related with consistent up-regulation of some critical glucose metabolic genes such as transcription factor, *PPARα*, whereas sustained up-regulation of *PPARα* is regulated by reduced DNA methylation in the promoter of *PPARα* [54]. Stress responses in the adult rat are programmed early in life by maternal care, and is associated with altered transcription factor binding to the hippocampal *GR* promoter [55]. Interestingly, central infusion of these adult offspring with L-methionine, a precursor to S-adenosyl-methionine that serves as the donor of methyl groups for DNA methylation, with altered DNA methylation due to maternal behavior programming during gestation can be reversed, together with altered nerve growth factor-inducible protein-A binding to the *GR* promoter and hypothalamic-pituitary-adrenal and behavioral responses to stress [56, 57]. These results demonstrate that despite the inherent stability of the epigenomic marks established early in life through behavioral programming, they are potentially reversible in the adult brain. Therefore, in view of the reversibility of epigenetic mechanisms, intervention with drugs during early life may ameliorate the glucose metabolism disorders in later life if the early disrupting environment has resulted in modified epigenetics, which can generates long-lasting effects.

CONCLUSION AND FURTHER PERSPECTIVE

Today, apart from genotype and adult lifestyle, early life abnormal nutrition including both undernutrition and overnutrition are closely related to aberrant glucose metabolism in adult such as impaired glucose tolerance, gestational diabetes mellitus and type 2 diabetes. It is

suggested that DNA methylation and microRNAs may be the crucial epigenetic mechanisms accounting for the relationship between the early life nutrition and glucose metabolism in later life. More importantly, DNA methylation and microRNAs can be utilized as novel biomarkers and targets for the diagnosis, prevention and intervention of diabetes in the near future.

REFERENCES

- [1] *IDF Diabetes Atlas, 6th edn.* Available online: <http://www.idf.org/diabetesatlas>.
- [2] Slatkin, M. Epigenetic inheritance and the missing heritability problem. *Genetics*. 2009, 182, 845-850.
- [3] Gluckman, P.D.; Hanson, M.A.; Buklijas, T.; Low, F.M.; Beedle, A.S. Epigenetic mechanisms that underpin metabolic and cardiovascular diseases. *Nature reviews. Endocrinology*. 2009, 5, 401-408.
- [4] Waddington, C.H. The epigenotype. 1942. *International journal of epidemiology*. 2012, 41, 10-13.
- [5] Holliday, R. *Epigenetics A Historical Overview. Epigenetics-U.S.* 2006, 1, 76-80.
- [6] Holliday, R.; Pugh, J.E. DNA Modification Mechanisms and Gene Activity during Development. *Science*. 1975, 187, 226-232.
- [7] Turner, B.M. Histone acetylation as an epigenetic determinant of long-term transcriptional competence. *Cell Mol Life Sci*. 1998, 54, 21-31.
- [8] Aguilera, O.; Fernandez, A.F.; Munoz, A.; Fraga, M.F. Epigenetics and environment: a complex relationship. *J Appl Physiol*. 2010, 109, 243-251.
- [9] Fraga, M.F. Genetic and epigenetic regulation of aging. *Curr Opin Immunol*. 2009, 21, 446-453.
- [10] Skinner, M.K.; Manikkam, M.; Guerrero-Bosagna, C. Epigenetic transgenerational actions of environmental factors in disease etiology. *Trends in endocrinology and metabolism: TEM*. 2010, 21, 214-222.
- [11] Bird, A. DNA methylation patterns and epigenetic memory. *Gene Dev*. 2002, 16, 6-21.
- [12] Gardinergarden, M.; Frommer, M. CpG Islands in Vertebrate Genomes. *J Mol Biol*. 1987, 196, 261-282.
- [13] Reik, W.; Dean, W. DNA methylation and mammalian epigenetics. *Electrophoresis*. 2001, 22, 2838-2843.
- [14] Kacem, S.; Feil, R. Chromatin mechanisms in genomic imprinting. *Mamm Genome*. 2009, 20, 544-556.
- [15] Jaenisch, R.; Bird, A. Epigenetic regulation of gene expression: how the genome integrates intrinsic and environmental signals. *Nat Genet*. 2003, 33, 245-254.
- [16] Reddy, M.A.; Natarajan, R. Epigenetic mechanisms in diabetic vascular complications. *Cardiovascular research*. 2011, 90, 421-429.
- [17] Forsdahl, A. Are poor living conditions in childhood and adolescence an important risk factor for arteriosclerotic heart disease? *British journal of preventive & social medicine*. 1977, 31, 91-95.
- [18] Barker, D.J.; Winter, P.D.; Osmond, C.; Margetts, B.; Simmonds, S.J. Weight in infancy and death from ischaemic heart disease. *Lancet*. 1989, 2, 577-580.

- [19] Painter, R.C.; Roseboom, T.J.; Bleker, O.P. Prenatal exposure to the Dutch famine and disease in later life: an overview. *Reproductive toxicology*. 2005, 20, 345-352.
- [20] Curhan, G.C.; Willett, W.C.; Rimm, E.B.; Spiegelman, D.; Ascherio, A.L.; Stampfer, M.J. Birth weight and adult hypertension, diabetes mellitus, and obesity in US men. *Circulation*. 1996, 94, 3246-3250.
- [21] Hales, C.N.; Barker, D.J.P.; Clark, P.M.S.; Cox, L.J.; Fall, C.; Osmond, C.; Winter, P.D. Fetal and Infant Growth and Impaired Glucose-Tolerance at Age 64. *Brit Med J*. 1991, 303, 1019-1022.
- [22] Tobi, E.W.; Lumey, L.H.; Talens, R.P.; Kremer, D.; Putter, H.; Stein, A.D.; Slagboom, P.E.; Heijmans, B.T. DNA methylation differences after exposure to prenatal famine are common and timing- and sex-specific. *Human molecular genetics*. 2009, 18, 4046-4053.
- [23] Steegers-Theunissen, R.P.; Obermann-Borst, S.A.; Kremer, D.; Lindemans, J.; Siebel, C.; Steegers, E.A.; Slagboom, P.E.; Heijmans, B.T. Periconceptional maternal folic acid use of 400 microg per day is related to increased methylation of the IGF2 gene in the very young child. *PloS one*. 2009, 4, e7845.
- [24] Godfrey, K.M.; Sheppard, A.; Gluckman, P.D.; Lillycrop, K.A.; Burdge, G.C.; McLean, C.; Rodford, J.; Slater-Jefferies, J.L.; Garratt, E.; Crozier, S.R.; Emerald, B.S.; Gale, C.R.; Inskip, H.M.; Cooper, C.; Hanson, M.A. Epigenetic gene promoter methylation at birth is associated with child's later adiposity. *Diabetes*. 2011, 60, 1528-1534.
- [25] Mayor-Lynn, K.; Toloubeydokhti, T.; Cruz, A.C.; Chegini, N. Expression profile of microRNAs and mRNAs in human placentas from pregnancies complicated by preeclampsia and preterm labor. *Reproductive sciences (Thousand Oaks, Calif.)*. 2011, 18, 46-56.
- [26] Maccani, M.A.; Padbury, J.F.; Lester, B.M.; Knopik, V.S.; Marsit, C.J. Placental miRNA expression profiles are associated with measures of infant neurobehavioral outcomes. *Pediatric research*. 2013, 74, 272-278.
- [27] Huang, L.; Shen, Z.; Xu, Q.; Huang, X.; Chen, Q.; Li, D. Increased levels of microRNA-424 are associated with the pathogenesis of fetal growth restriction. *Placenta*. 2013, 34, 624-627.
- [28] Tang, Q.; Wu, W.; Xu, X.; Huang, L.; Gao, Q.; Chen, H.; Sun, H.; Xia, Y.; Sha, J.; Wang, X.; Chen, D.; Xu, Q. miR-141 contributes to fetal growth restriction by regulating PLAG1 expression. *PloS one*. 2013, 8, e58737.
- [29] Ferland-McCollough, D.; Fernandez-Twinn, D.S.; Cannell, I.G.; David, H.; Warner, M.; Vaag, A.A.; Bork-Jensen, J.; Brons, C.; Gant, T.W.; Willis, A.E.; Siddle, K.; Bushell, M.; Ozanne, S.E. Programming of adipose tissue miR-483-3p and GDF-3 expression by maternal diet in type 2 diabetes. *Cell death and differentiation*. 2012, 19, 1003-1012.
- [30] Jiang, H.; Wu, W.; Zhang, M.; Li, J.; Peng, Y.; Miao, T.T.; Zhu, H.; Xu, G. Aberrant upregulation of miR-21 in placental tissues of macrosomia. 2014. 34, 658-663.
- [31] Lucchini, R.; Barba, G.; Giampietro, S.; Trivelli, M.; Dito, L.; De Curtis, M. [Macrosomic infants: clinical problems at birth and afterward]. *Minerva pediatrica*. 2010, 62, 65-66.
- [32] Bocock, P.N.; Aagaard-Tillery, K.M. Animal models of epigenetic inheritance. *Seminars in reproductive medicine*. 2009, 27, 369-379.
- [33] Rees, W.D.; Hay, S.M.; Brown, D.S.; Antipatis, C.; Palmer, R.M. Maternal protein deficiency causes hypermethylation of DNA in the livers of rat fetuses. *The Journal of nutrition*. 2000, 130, 1821-1826.

- [34] Burdge, G.C.; Slater-Jefferies, J.; Torrens, C.; Phillips, E.S.; Hanson, M.A.; Lillycrop, K.A. Dietary protein restriction of pregnant rats in the F0 generation induces altered methylation of hepatic gene promoters in the adult male offspring in the F1 and F2 generations. *The British journal of nutrition*. 2007, 97, 435-439.
- [35] Lillycrop, K.A.; Phillips, E.S.; Jackson, A.A.; Hanson, M.A.; Burdge, G.C. Dietary protein restriction of pregnant rats induces and folic acid supplementation prevents epigenetic modification of hepatic gene expression in the offspring. *The Journal of nutrition*. 2005, 135, 1382-1386.
- [36] Sandovici, I.; Smith, N.H.; Nitert, M.D.; Ackers-Johnson, M.; Uribe-Lewis, S.; Ito, Y.; Jones, R.H.; Marquez, V.E.; Cairns, W.; Tadayyon, M.; O'Neill, L.P.; Murrell, A.; Ling, C.; Constancia, M.; Ozanne, S.E. Maternal diet and aging alter the epigenetic control of a promoter-enhancer interaction at the Hnf4a gene in rat pancreatic islets. *Proceedings of the National Academy of Sciences of the United States of America*. 2011, 108, 5449-5454.
- [37] Jousse, C.; Parry, L.; Lambert-Langlais, S.; Maurin, A.C.; Averous, J.; Bruhat, A.; Carraro, V.; Tost, J.; Letteron, P.; Chen, P.; Jockers, R.; Launay, J.M.; Mallet, J.; Fafournoux, P. Perinatal undernutrition affects the methylation and expression of the leptin gene in adults: implication for the understanding of metabolic syndrome. *FASEB journal : official publication of the Federation of American Societies for Experimental Biology*. 2011, 25, 3271-3278.
- [38] Vo, T.X.; Revesz, A.; Sohi, G.; Ma, N.; Hardy, D.B. Maternal protein restriction leads to enhanced hepatic gluconeogenic gene expression in adult male rat offspring due to impaired expression of the liver X receptor. *The Journal of endocrinology*. 2013, 218, 85-97.
- [39] Cong, R.; Jia, Y.; Li, R.; Ni, Y.; Yang, X.; Sun, Q.; Parvizi, N.; Zhao, R. Maternal low-protein diet causes epigenetic deregulation of HMGCR and CYP7alpha1 in the liver of weaning piglets. *The Journal of nutritional biochemistry*. 2012, 23, 1647-1654.
- [40] Vucetic, Z.; Kimmel, J.; Totoki, K.; Hollenbeck, E.; Reyes, T.M. Maternal high-fat diet alters methylation and gene expression of dopamine and opioid-related genes. *Endocrinology*. 2010, 151, 4756-4764.
- [41] Vucetic, Z.; Kimmel, J.; Reyes, T.M. Chronic high-fat diet drives postnatal epigenetic regulation of mu-opioid receptor in the brain. *Neuropsychopharmacology: official publication of the American College of Neuropsychopharmacology*. 2011, 36, 1199-1206.
- [42] Khalyfa, A.; Carreras, A.; Hakim, F.; Cunningham, J.M.; Wang, Y.; Gozal, D. Effects of late gestational high-fat diet on body weight, metabolic regulation and adipokine expression in offspring. *International journal of obesity (2005)*. 2013, 37, 1481-1489.
- [43] Sharpe, R.M. Environmental/lifestyle effects on spermatogenesis. Philosophical transactions of the Royal Society of London. Series B, *Biological sciences*. 2010, 365, 1697-1712.
- [44] Robertson, S.A. Seminal plasma and male factor signalling in the female reproductive tract. *Cell and tissue research*. 2005, 322, 43-52.
- [45] Ng, S.F.; Lin, R.C.; Laybutt, D.R.; Barres, R.; Owens, J.A.; Morris, M.J. Chronic high-fat diet in fathers programs beta-cell dysfunction in female rat offspring. *Nature*. 2010, 467, 963-966.

- [46] Carone, B.R.; Fauquier, L.; Habib, N.; Shea, J.M.; Hart, C.E.; Li, R.; Bock, C.; Li, C.; Gu, H.; Zamore, P.D.; Meissner, A.; Weng, Z.; Hofmann, H.A.; Friedman, N.; Rando, O.J. Paternally induced transgenerational environmental reprogramming of metabolic gene expression in mammals. *Cell*. 2010, 143, 1084-1096.
- [47] Wei, Y.; Yang, C.R.; Wei, Y.P.; Zhao, Z.A.; Hou, Y.; Schatten, H.; Sun, Q.Y. Paternally induced transgenerational inheritance of susceptibility to diabetes in mammals. *Proceedings of the National Academy of Sciences of the United States of America*. 2014, 111, 1873-1878.
- [48] Benatti, R.O.; Melo, A.M.; Borges, F.O.; Ignacio-Souza, L.M.; Simino, L.A.; Milanski, M.; Velloso, L.A.; Torsoni, M.A.; Torsoni, A.S. Maternal high-fat diet consumption modulates hepatic lipid metabolism and microRNA-122 (miR-122) and microRNA-370 (miR-370) expression in offspring. *The British journal of nutrition*. 2014, 111, 2112-2122.
- [49] Dumortier, O.; Hinault, C.; Gautier, N.; Patouraux, S.; Casamento, V.; Van Obberghen, E. Maternal protein restriction leads to pancreatic failure in offspring: role of misexpressed microRNA-375. *Diabetes*. 2014, 63, 3416-3427.
- [50] Lie, S.; Morrison, J.L.; Williams-Wyss, O.; Suter, C.M.; Humphreys, D.T.; Ozanne, S.E.; Zhang, S.; MacLaughlin, S.M.; Kleemann, D.O.; Walker, S.K.; Roberts, C.T.; McMillen, I.C. Periconceptional undernutrition programs changes in insulin-signaling molecules and microRNAs in skeletal muscle in singleton and twin fetal sheep. *Biology of reproduction*. 2014, 90, 5.
- [51] Lie, S.; Morrison, J.L.; Williams-Wyss, O.; Suter, C.M.; Humphreys, D.T.; Ozanne, S.E.; Zhang, S.; MacLaughlin, S.M.; Kleemann, D.O.; Walker, S.K.; Roberts, C.T.; McMillen, I.C. Impact of embryo number and maternal undernutrition around the time of conception on insulin signaling and gluconeogenic factors and microRNAs in the liver of fetal sheep. *American journal of physiology. Endocrinology and metabolism*. 2014, 306, E1013-1024.
- [52] Vickers, M.H.; Breier, B.H.; Cutfield, W.S.; Hofman, P.L.; Gluckman, P.D. Fetal origins of hyperphagia, obesity, and hypertension and postnatal amplification by hypercaloric nutrition. *American journal of physiology. Endocrinology and metabolism* 2000, 279, E83-87.
- [53] Vickers, M.H.; Gluckman, P.D.; Coveny, A.H.; Hofman, P.L.; Cutfield, W.S.; Gertler, A.; Breier, B.H.; Harris, M. Neonatal leptin treatment reverses developmental programming. *Endocrinology*. 2005, 146, 4211-4216.
- [54] Gluckman, P.D.; Lillycrop, K.A.; Vickers, M.H.; Pleasants, A.B.; Phillips, E.S.; Beedle, A.S.; Burdge, G.C.; Hanson, M.A. Metabolic plasticity during mammalian development is directionally dependent on early nutritional status. *Proceedings of the National Academy of Sciences of the United States of America*. 2007, 104, 12796-12800.
- [55] Weaver, I.C.; Cervoni, N.; Champagne, F.A.; D'Alessio, A.C.; Sharma, S.; Seckl, J.R.; Dymov, S.; Szyf, M.; Meaney, M.J. Epigenetic programming by maternal behavior. *Nature neuroscience*. 2004, 7, 847-854.
- [56] Weaver, I.C.; Champagne, F.A.; Brown, S.E.; Dymov, S.; Sharma, S.; Meaney, M.J.; Szyf, M. Reversal of maternal programming of stress responses in adult offspring through methyl supplementation: altering epigenetic marking later in life. *The Journal of neuroscience : the official journal of the Society for Neuroscience*. 2005, 25, 11045-11054.

- [57] Weaver, I.C. Epigenetic programming by maternal behavior and pharmacological intervention. *Nature versus nurture: let's call the whole thing off. Epigenetics-U.S.* 2007, 2, 22-28.

Chapter 60

EFFECTS OF OXIDATIVE STRESS ON EPIGENETIC MECHANISMS

***Yildiz Dincer**, PhD and Onur Baykara, PhD**

Istanbul University, Cerrahpasa Medical Faculty, Department of Medical Biochemistry,
Istanbul, Turkey

ABSTRACT

Oxidative stress is a state in which production of reactive oxygen species exceeds the capacity of antioxidant systems. Reactive oxygen species have one or more unpaired electrons, making them highly reactive with other cellular molecules such as protein, lipid, and nucleic acid. The peroxidation of polyunsaturated fatty acids in biological membranes results in impaired membrane integrity. Oxidatively modified proteins lose *their* capacity to carry out the *physiological functions* and they may form intracellular aggregates. Attacks of reactive oxygen species to DNA results in strand breakages and base oxidation. Major DNA oxidation product is 8-hydroxydeoxyguanosine which has a pro-mutagenic potential. Due to these damaging effects, oxidative stress plays an important role in various pathologies such as cancer, diabetes, chronic inflammatory diseases and neurodegenerative disorders. Epigenetic changes are regular and natural events which regulate gene expression without changing base sequences on DNA. Dysregulation of regular epigenetic mechanisms is a contributory factor for many of human pathologies. Recently, reactive oxygen species have been shown to cause epigenetic dysregulations that play a pivotal role in human disorders. The basic epigenetic mechanisms and their dysregulation by reactive oxygen species have been reviewed in this chapter.

Keywords: epigenetic modifications, DNA methylation, histone acetylation, miRNAs, oxidative stress, 8-hydroxydeoxyguanosine

* Corresponding Author's Email: yldz.dincer@gmail.com.

ABBREVIATIONS

5-hmC	5-hydroxymethylcytosine
5-mC	5-methylcytosine
8-OHdG	8-hydroxydeoxy guanosine
CoA	Acetyl co-enzyme A
ATG12	Autophagy-12
ATG8	Autophagy-8
CAT	Catalase
CNS	Central Nervous System
CpG sites	Cytosine- phosphate-guanine sites
DNA	Deoxyribonucleic acid
DGCR8	DiGeorge Syndrome Critical Region 8
DNMT	DNA methyltransferase
Exp5	Exportin 5
GSH-Px	Glutathione Peroxidases
GST	Glutathione S-Transferases
GNAT	Gcn5-related N-acetylases
Gnc5,	PCAF and elongator complex protein 3(ELP3)
HAT	Histone Acetyl Transferases
HDACi	Histone Deacetylase Inhibitors
HDAC	Histone Deacetylases
HMTS	Histone Methyl Transferases
FAT10	Human leukocyte antigen F-associated
IUPAC	International Union of Pure and Applied Chemistry
LINE-1	Long-interspersed nuclear element-1
MDBP	Methyl-CpG-binding domain proteins
miRNA	MicroRNA
MDS	Myelodysplastic syndrome
NAD	Nicotinamide adenine dinucleotide (+)
ncRNA	Noncoding RNA
NF- κ b	Nuclear factor- κ b
OFR	Oxygen free radicals
PAZ	Piwi, Argonaut and Zwillie
piRNA	Piwi-interacting RNA
RNA	Ribonucleic acid
rRNA	Ribosomal RNA
RISC	RNA-induced silencing complex
SAM	S-adenosyl-L-methionine
SET	Su(var)3-9, Enhancer of Zeste, Trithorax
siRNA	Small interfering RNA
snoRNA	Small nucleolar RNA
Sumo	Small Ubiquitin-like Modifier
SOD	Superoxide dismutase
tRNA	Transfer RNA

UCRP	Ubiquitin cross-reactive protein
UFM-1	Ubiquitin fold modifier-
UBL5	Ubiquitin like protein-5
UBLs	Ubiquitin like proteins
Ub	Ubiquitin

1. EPIGENETIC MECHANISMS

The term epigenetics was first coined by C.H Waddington in 1942, a word derived from epigenesis and genetics. Epigenesis is the word for describing live organisms differentiating from a single cell into an organ or organism. Epigenetic mechanisms modulate gene expression patterns without affecting the deoxyribonucleic acid (DNA) sequence. To stop confusions, a consensus definition was put on use of the epigenetic trait: "stably heritable phenotype resulting from changes in a chromosome without alterations in the DNA sequence" [1]. These mechanisms are acquired throughout life and depend on environmental clues such as diet, lifestyle and toxin exposure. They are heritable mechanisms so epigenetic changes can be transferred to the new cell when cells are divided with mitosis. If this process takes place in germ cells and meiosis, genomic imprinting may occur. In some occasions, such as differentiation of embryonic stem cells to specific tissues, the expression is stabilized so that the cells cannot go back to their previous state. Although all cells in one individual have the same DNA sequence, epigenetic regulation occurs at the specific gene loci in the specific cells to yield specific cellular phenotypes. A change in the phenotype does not usually effect the genotype. Gene expression can be activated or silenced via epigenetic regulations. Epigenetic changes can alter transcriptional activity of genes, and may mediate the differences in risk for certain diseases. Methylation, phosphorylation, acetylation, ubiquitination, sumoylation are included in epigenetic processes.

Chromatins are complex macromolecules in cells. DNA, RNA, histones and non-histone proteins form chromatin structure. DNA is tightly packed into the nucleus in the chromatin structure, and this complex can be exposed to modification with acetyl groups (acetylation), non-coding RNAs like microRNAs (miRNAs) and small interfering RNAs (siRNAs). These modifications may cause changes in chromatin structure which in turn effect the expression patterns of the genes. So to summarize, epigenetic modifications are typically divided into three categories: (1) DNA methylation (2) histone post-translational modifications and (3) microRNAs.

1.1. DNA Methylation

DNA methylation can be briefly explained as covalent addition of a methyl group (CH₃) to the 5-carbon of cytosine forming 5-methylcytosine (5-mC) mostly located at cytosine-phosphate-guanine sites (CpG sites) that are present in the 5'-untranslated regions of gene promoters. Although DNA methylation usually ends up with suppression of gene transcription due to binding of methyl groups into the major groove of DNA, this process is essential for development and directly associated with processes in the cell such as X-chromosome

inactivation, genomic imprinting and in carcinogenesis. 5-mC can be hydroxylated forming 5-hydroxymethylcytosine (5-hmC) which consists of approximately 1.5% of human DNA [2]. DNA methylation is carried out by DNA methyltransferase (DNMT) activity, and methyl group donor is S-adenosyl-L-methionine (SAM). DNMTs are family of enzymes which catalyze the transfer of methyl groups to DNA. DNMT in mammals contains two main regions, C terminal region and N-terminal region. The C-terminal is the catalytic part containing approximately 500 amino acids, and this is the part which center of enzyme is located. The N terminal includes 621 amino acids, and is not vital for DNMT1 activity [3].

Although DNMTs have been classified in four groups since many years including DNMT1, DNMT2, DNMT3a, and DNMT3b, in mammals, there are actually three types of methyltransferases: DNMT1, DNMT3a, and DNMT3b [4]. DNMT1 is the major methyltransferase for maintenance of methylation, for this reason it is usually called as maintenance methyltransferase [5]. It catalyzes the transfer of a methyl group from SAM on to the 5' position of the cytosine ring, generating 5-methylcytosine. DNMT2 which used to be known as a DNMT, now termed as TRDMT1, is not a methyltransferase, even though its sequence is strongly associated with other DNMTs. Its DNA methylation activity is very low [6]. It only has a C domain and probably has a role in recognition of DNA damage and mutation repair [7]. DNMT2 (TRDMT1-tRNA aspartic acid methyltransferase 1) has both ribonucleic acid (RNA) methyltransferase activity and 5-cytosine DNA methyltransferase activity [8]. Previous studies have shown that DNMT2 functions as tRNA methyl-transferase by methylating cytosine-38 in aspartic acid of tRNA [8, 9].

DNMT3 can methylate both hemimethylated and unmethylated CpG at the same rate. DNMT3 family has three members which are DNMT3a, DNMT3b and DNMT3L. The structure of DNMT3 family is quite similar to DNMT1 and they have a catalytic domain with a regulatory region. DNMT3a and DNMT3b have different functions from those of DNMT1, they serve as *de novo* methyltransferases. For establishment and mitotic inheritance of tissue-specific DNA methylation patterns in normal development, *de novo* methylation is essential. It occurs frequently during early development and gametogenesis in mice to remethylate the repetitive sequences those are demethylated during early stage of embryogenesis [10].

DNA methylation results in gene silencing by the recruitment of methyl CpG-binding transcriptional repressors and by interfering with the DNA binding of transcriptional activators. Hypomethylated DNA is associated with active chromatin. Transcription level of DNA is affected by DNA methylation at the promoter region via two ways: (1) Methylated DNA prevents the binding of transcription factors to the gene promoter (2) methylated DNAs are occupied by methyl-CpG-binding domain proteins (MBDPs). MBDPs bring together the other epigenetic components consequently forming compact and inactive heterochromatin, therefore cause decreased or silenced gene transcription [11].

Aberrant DNA methylation on disease-related gene promoters was determined in various diseases, especially in cancer. DNMT inhibitors are promising therapeutic targets. Both Decitabine (5-aza-2'-deoxycytidine) and Azacytidine are hypomethylating agents. They show their function by inhibiting DNMTs. Decitabine can be incorporated into DNA where Azacytidine incorporates both into DNA and RNA [12], as azacytidine is the chemical analogue of DNA and RNA. Both of the drugs have been approved by the FDA for treatment of myelodysplastic syndrome.

1.2. Histone Post-translational Modifications

In eukaryotic organisms, DNA is wrapped around a protein octamer. This octamer is composed of two units of H2A, H2B, H3 and H4 histone proteins which are linked to each other through a linker protein called H1 histone. This complex unit is called as nucleosome and is the basic unit of the chromosomes. Histones can undergo post-translational covalent modifications which are phosphorylation on serine or threonine residues, methylation on lysine or arginine, acetylation and deacetylation on lysines, ubiquitylation of lysines and sumoylation on lysines [13, 14]. The best defined histone modification is acetylation. Histone modifications are proposed to affect chromosome structure and function, especially during transcription and chromatin remodeling processes. Core modifications of histones can also act on DNA templated processes, like replication and transcriptional processes. They can also effect nucleosomal structure [15, 16]. Simon et al. have shown that histone posttranslational modifications within distinct structured regions of the nucleosome directly regulate the inherent dynamic properties of the nucleosome such as unwrapping and disassembly [17]. North et al. have shown that phosphorylation of H3(T118) alters nucleosome dynamics and remodeling [18]. Histone modifications are catalyzed by histone modification enzymes such as histone acetyltransferases, histone deacetylases, histone methyltransferases and histone demethylases.

Histone acetyltransferases (HAT) acetylate conserved lysine residues on histone tails. HATs transfer an acetyl group from acetyl co-enzyme A (CoA) forming ϵ -N-acetyllysine. Depending on their localization, HATs (EC 2.3.1.48) are classified into two groups, the ones localized in the nucleus and carrying a conserved motif are Type A HATs, and the ones localized in the cytoplasm are Type B HATs [19]. Type A HATs also have three primary subclasses which are 1) GNAT (Gcn5-related N-acetylases) (Gcn5, PCAF and elongator complex protein 3(ELP3)) 2) MYST (HMOF/MYST1, HBO1/MYST2, MOZ/MYST3, MORF/MYST4, Type60 and Type p300/CBP (p300 and CBP) families.

The acetylated lysine is neutralized from positive charges (normally, the histone tails are positively charged which help these tails interact with negatively charged phosphate groups found on DNA backbone), and this acetylation decreases the electrostatic interaction between DNA backbone and histone tails. This mechanism is used to control gene expression by activating and inactivating genes on DNA. Simply, HATs are transcriptional co-activators and can promote transcription factors for activation of transcription. Acetylation of nucleosomes by HATs helps transcription factors to reach the DNA more easily so that transcription can increase.

Histone deacetylases (EC 3.5.1.98, HDAC) remove acetyl groups from an ϵ -N-acetyl lysine amino acid on histone tails, therefore the histones can wrap tighter around the DNA which helps to regulate the expression of a gene. The density of wrapping can alter expression levels. HDACs are transcriptional co-repressors. HDACs usually inactivate gene expression [20]. Depending on their sequence homology and subcellular localization, HDACs can be classified into four classes as I, II (a and b), III and IV. HDAC. Class I, (HDACs 1, 2, 3, and 8) are localized in the nucleus. Class II (HDACs 4, 5, 6, 7, 9, and 10), Class III (SirT 1, 2, 3, 4, 5, 6 and 7) and Class IV HDACs (HDAC 11) can be localized either on nucleus or cytoplasm. Class III HDACs can also be localized in the mitochondria [21, 22]. HDAC Class I, II and IV are zinc-dependent which are all expressed in the brain [23], whereas HDAC III is nicotinamide adenine dinucleotide (NAD⁺) dependent. Although Class I HDACs are known as nuclear enzymes, they were shown to play a role in unfolded protein response by localizing to the

endoplasmic reticulum [24]. In addition to their gene silencing function, sirtuins are the key regulators of multiple biological and metabolic events including energy metabolism, glucose and lipid metabolisms, circadian rhythms, inflammation, stress resistance, apoptosis, and autophagy. Many scientists have performed studies on HDACs and their functions in various diseases to find out the underlying epigenetic mechanism.

HDAC 2 is expressed in the central nervous system (CNS) and has a negative effect on memory and synaptic plasticity. Guan et al. have shown that binding of HDAC2 to the promoters of synaptic-plasticity-related genes and thereby negatively regulates their transcriptions [25]. Graff et al. have knocked down HDAC2 by short-hairpin-RNA and they have shown that structural synaptic plasticity and memory impairments were restored in mice [26]. These findings support the information that HDAC2 is associated with disrupted brain functions in neurodegenerative diseases.

Histone deacetylase inhibitors (HDACi) cause to an increase in histone acetylation and reduce the affinity of the histone to the negatively charged DNA backbone. By this method, chromatin gets less dense so that the transcription factors have the chance of reaching the DNA and increasing gene expression. Certain non-histone proteins (DNA repair proteins, DNA binding proteins, transcription factors, chaperones) are also substrate for HDACs. Therefore HDACs take place in various pathways including inhibition of DNA repair, induction of apoptosis, generation of reactive oxygen species (ROS), inhibition of cell cycle arrest and inhibition of antiangiogenesis [27]. In this context HDACi are considered as an emerging therapeutic target for cancer [28, 29] and HIV [30].

Histone methyl transferases (HMTs) (2.1.1.43 EC) are the enzymes (e.g; histone-lysine N-methyltransferases and histone-arginine N-methyltransferases) which modify the histones. They function in adding methyl groups to lysine (K) and arginine (R) residues. Lysine methyltransferases catalyze mono-, di-, and trimethylation of the lysine ϵ -amino group. Histone methylation can be managed by SAMs as the methyl group donor [31]. HMTs are composed of two classes, lysine specific or arginine specific. Lysine-specific HMTs have two subgroups which carry a special type of protein domain called SET (Su(var)3-9, Enhancer of Zeste, Trithorax) domain containing and non-SET domain containing. Histone methylation controls histone-protein interaction. It mediates the binding of HDACs to histones, thus silences the gene indirectly.

Ubiquitination (Ubiquitylation) of histones is required for histone methylation. Ubiquitin (Ub) is a regulatory protein and is found almost in any tissue of eukaryotes. Ubiquitination may cause the proteins to degrade with proteasomes, change their locations in the cell and increase or decrease protein-protein interactions [32]. Ubiquitination may also play major roles in apoptosis, cell cycle, DNA transcription and repair, response to oxidative stress and many other mechanisms. The protein to be ubiquinated must be activated, conjugated and ligated. These steps can be managed by ubiquitin-activating enzymes (E1s), ubiquitin-conjugating enzymes (E2s), and ubiquitin ligases (E3s), respectively. Ubiquitin first activates E1 and then it is transferred to E2. E3 Ubiquitin ligases are activated by a E2-Ub complex [33]. At the end of this process, ubiquitin is bound to lysine residues on the protein's N-terminus or the protein substrate. The binding can be either with a peptide bond or an isopeptide bond, respectively [34]. In general, defects in ubiquitination mechanism can alter the functions of the genes and proteins. Ubiquitination is associated with many types of disease including cancer, neurodegenerative and inflammatory diseases. [35, 36]. Ubiquitination of histone proteins is required for histone methylation, therefore silences the gene indirectly.

Although ubiquitin is the most studied and commonly known protein, another family of ubiquitin like proteins (UBLs) also exist. These proteins act like ubiquitin and modify proteins but they have distinct functions. One of the major members of this family is Sumos. Small Ubiquitin-like Modifier (Sumo) proteins are a family of small proteins comprising approximately 100 amino acids. Sumoylation is directed by an enzymatic cascade analogous to that involved in ubiquitination but SUMO modification usually functions on the opposite direction of ubiquitination and can help stabilizing protein substrates. As a post-translational modification, sumoylation plays different roles in cellular mechanisms, transportations, apoptosis and response to oxidative stress. Sumoylation regulates protein-protein interaction and localization, and enhances the interaction between histones and HDACs. Therefore it is involved in gene silencing, indirectly. Other members of UBLs include, but not limited to, ubiquitin cross-reactive protein (UCRP), ubiquitin fold modifier- (UFM-1), ubiquitin like protein-5 (UBL5), human leukocyte antigen F-associated (FAT10), autophagy-8 (ATG8), autophagy-12 (ATG12) and many others.

Carbonylation, mostly occurring on lysine and arginine residues, is a consequence of oxidative stress and it is mostly irreversible. Carbonylation causes loss of protein function, the modified proteins are aggregated and they accumulate as unfolded or damaged proteins. [37]. Sharma et al. have shown histone carbonylation to occur on histones H1, H2A, H2B and H3 but not H4 in rat liver. Contrary the wide belief that carbonylation increases by age, they have found the carbonylation was lower in older rats [38].

Besides these most frequently investigated post translational modifications, there are many other modifications such as *acylation* (*alkanoylation*- addition of functional groups on proteins such as addition of fatty acids to amino acids), *alkylation* (transfer of an alkyl group to another molecule, most common form is methylation with a distinction, in methylation only one carbon is transferred while in alkylation long chain of carbon groups are transferred), *glycosylation* (enzymatic addition of sugar molecule to a protein substrate to help proteins folding correctly, stabilize proteins, prevent protein degradation), *glycation* (non-enzymatic glycosylation, covalent binding of a protein with sugar molecule), *biotinylation* (covalent binding of biotin to a protein target), *neddylation* (conjugation of ubiquitin-like protein NEDD8 to protein substrates), *succinylation* (addition of succinyl group to a lysine residue), *sulfation* (addition of a sulfate group to a tyrosine residue).

1.3. Micro RNAs and Epigenetic

Small noncoding RNAs (ncRNAs) are functional RNA molecules which do not encode a protein. Even though they are not translated into a protein, they play major roles in cellular and molecular mechanisms. The family of ncRNAs include micro RNAs (miRNAs), transfer RNAs (tRNAs), ribosomal RNAs (rRNAs), small interfering RNAs (siRNAs), Piwi-interacting RNAs (piRNAs), and small nucleolar RNA (snoRNAs).

miRNAs are a family of short non-coding RNAs of approximately 19-25 long nucleotides which modulates the expression of mRNA. They play crucial roles in cell proliferation, apoptosis, cell cycle control, differentiation and cellular metabolism. miRNAs are found in serum, plasma, saliva, and other body fluids of humans and also in other organisms such as plants, animals and some viruses. Their regulation of gene expression is at post-transcriptional

level. miRNAs bind to target mRNAs at their 3'-untranslated regions and suppress their translation and/or promote their degradation, thus sequence-specifically control the translation of target mRNAs [39]. In addition, miRNAs bind complementary sequences on DNA and block access for transcription factors, which in turn results in turned off genes. miRNAs may show their functions on many different targets and a gene can be regulated by different miRNAs [40].

A miRNA is produced from a RNA-coding gene or from introns. It is thought that about 40% of all miRNA genes are located in the introns of protein coding or non-coding genes [41]. miRNAs start their life as primary transcripts known as a pri-miRNAs which are transcribed by RNA polymerase II in the nucleus and they carry a 5' cap and a poly-A-tail. The pri-miRNAs are then transcribed to pre-miRNA, a 70 nucleotide long structure, in the microprocessor complex consisting of RNase III enzyme Drosha and dsDNA binding protein DGCR8 (DiGeorge Syndrome Critical Region 8- Pasha in invertebrates). The pre-miRNAs are imported out of the nucleus into the cytoplasm by RAN-GTP complex and exporting 5 (Exp5) protein. This complex plays an important role in translocation of the RNA and proteins through the nuclear pores. When pre-miRNAs are out of the nucleus, Dicer, a RNase III enzyme containing an ATPase/RNA helicase, a DUF283 (Domain of unknown function) domain, a PAZ (Piwi, Argonaut and Zwiille) domain, two catalytic RNase III domains (RIIIa and RIIIb), and a C-terminal double-stranded RNA-binding domain (dsRBD), cleaves the pre-miRNA stem loop into two complementary RNA molecules but only one of these molecules integrate with the RNA-induced silencing complex (RISC), resulting with mature miRNA.

Even though the biogenesis of miRNA has been extensively studied, the regulation of miRNA expression still needs to be clarified. A miRNA can be localized in various regions of the genome. If they are localized between the genes, they are called intragenic, if the localization is in the gene they are called intergenic. In an *in silico* analysis by Sato et al. it has been shown that of 939 miRNAs, 293 (31.2%) were intergenic, whereas 317 (44.4%), 119 (12.7%) and 110 (11.7%) were overlapped by RNA transcripts in the same, opposite and both directions, respectively [42].

There are numerous studies investigating the relationship between miRNA and epigenetic regulations in various diseases such as cancer, diabetic, neurologic and inflammatory related diseases, aging and this number is still growing. Even though the underlying mechanism is not clearly understood, miRNA dysregulation is determined in all of these pathologies. It is thought that miRNAs contribute to formation of diseases via epigenetic regulations, and miRNA modulation is a new therapeutic approach that is extensively investigated in various diseases.

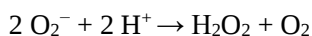
2. OXIDATIVE STRESS

Living organisms mostly rely on oxygen. Oxygen is like a double-edged knife. The cells of living organisms cannot live without oxygen but oxygen and its radical species can also harm cells by reacting with cellular components. Impaired balance between the oxidants and the anti-oxidants may cause a damage to the cells and this is called the oxidative stress. If the body cannot fight against the harmful reactive oxygen species, with another word oxygen free radicals (OFR), such as O_2^- (superoxide radical), $OH\cdot$ (hydroxyl radical), triplet oxygen, triplet carbene ($:CH_2$) and H_2O_2 (hydrogen peroxide), they may cause irreversible damage on the

cellular components such as DNA, proteins, lipids. This damage may even cause stop cellular signaling which is essential for the viability, motility and production of the cell.

According to IUPAC (International Union of Pure and Applied Chemistry) a free radical can be an atom, a molecule or an ion that has unpaired valence electrons [43]. These free radicals are highly reactive with other molecules. As they have an unbalanced structure and lacking of electrons, they are very prone to interact with other molecules in order to steal their electrons from them. They can even react with each other to form dimers, trimers or polymers. The most infamous radicals are superoxide anion (O_2^-) and hydroxyl radical ($OH\cdot$). O_2^- is formed after one electron reduction of dioxygen (O_2). Superoxides are toxic substances and produced by the immune system to be used against foreign substances. In case of a dismutation of superoxide, molecular oxygen (O_2) and hydrogen peroxide (H_2O_2) are formed and this reaction is catalyzed by superoxide dismutase enzyme (SOD-E.C 1.15.1.1).

SOD



The main source of OFR in the cells is the leakage from mitochondrial electron transport chain especially from complex I (NADH dehydrogenase) and III (coenzyme Q and cytochrome C oxidoreductase) [44]. When components of the oxidative phosphorylation located in the inner membrane of the mitochondria are reduced, production of mitochondrial superoxide radical increases. Superoxide radical and hydroxyl radical start the lipid peroxidation in cytoplasm, nucleus, mitochondria and endoplasmic reticulum which increases membrane permeability. Due to effect of OFR proteins, nuclear and mitochondrial DNA are oxidized. Oxidation of proteins may result in functional loss. OFR attacks to DNA causes DNA strand breaks and base modifications. Among the oxidized bases, 8-hydroxydeoxyguanosine (8-OHdG) is the most mutagenic lesion. By mispairing with A residues, 8-OHdG leads to an increased frequency of spontaneous G:C $\rightarrow$ T:A transversion. This mutation is recorded in many mutated proto-oncogenes and tumor suppressor, and 8-OHdG is strongly implicated in the initiation of carcinogenesis.

As mentioned above, free radicals can steal electrons from other atoms, destabilizing and perturbing them. In order to overcome this attack, the body uses antioxidants. Antioxidants are used to neutralize OFR before they harm the cells. They can be classified in two main groups, endogenous and exogenous antioxidants. Endogenous antioxidants can be sub-grouped into two as enzymatic antioxidants and non-enzymatic antioxidants. The first one includes the enzymes such as Superoxide dismutase (SOD), Glutathione S-Transferases (GST), Glutathione Peroxidases (GSH-Px), Catalase (CAT), mitochondrial cytochrome oxidase system, whereas the latter group consists of selenium, melatonin, hemoglobin, glutathione, bilirubin, methionine, retinal, thiols, and many others. Exogenous antioxidants can be classified as vitamins (ascorbic acid, alpha tocoferol, beta carotene, folic acid), drugs (xanthine oxidase inhibitors, NADPH oxidase inhibitors etc.) and nutrient antioxidants (sodium benzoate, catechins, anthocyanins, tannins, lycopene, etc.)

Due to the damaging effects on cellular macromolecules, OFR contribute to pathogenesis of many diseases including cancer, ageing, neurodegenerative diseases, obesity, cardiovascular diseases, neuroinflammation, etc. [45-49]. It has been recently recognized that OFR also modulate epigenetic mechanisms.

3. INTERACTIONS BETWEEN OXIDATIVE STRESS AND EPIGENETIC MECHANISMS

OFR may modulate epigenetic mechanisms. They can effect epigenetic mechanisms in various ways: 1) OFR changes the pattern of DNA methylation 2) Oxidative stress reduces methyl-accepting ability of DNA 3) OFR cause changes in histone modifications [50].

3.1. Oxidative Stress - DNA Methylation

Distruption of DNA methylation patterns by OFR is the best defined in carcinogenesis. Briefly, the hypermethylation of the DNA can cause gene silencing resulting with decreased gene expression, uncontrolled cell proliferation and growth which may end up with cancer. On the other hand, hypomethylation may cause loss of imprinting and chromosomal instability [51, 52]. Hydroxyl radical-induced DNA lesions such as 8-hydroxy-2-deoxyguanosine (8- OHdG) have been shown to contribute to decreased DNA methylation. 8- OHdG adducts interfere with the ability of DNA to function as a substrate for DNMTs. This results in global hypomethylation of the genome which in turn leads to oncogene activation and chromosomal instability. In addition, OFR are responsible for gene silencing by leading aberrant hypermethylation in CpG island-rich promoter region of tumor suppressor genes [53]. Many tumor suppressor genes have been found to be silenced via OFR-mediated aberrant methylation of promoter regions [54-57]. Recent studies have shown a further modification on the methylated CpGs which may make the gene less prone to transcription than those in the methylated state. This further modification is hydroxylation of 5-methylcytosine. Oxidative stress results in hydroxymethylation of DNA, especially in neurons [58]. Regarding other diseases, Patchsung et al. have demonstrated a relationship between increased oxidative stress and hypomethylation of the transposable long-interspersed nuclear element-1 (LINE-1) in patients with bladder cancer. They have found that LINE-1 hypomethylation levels and the number of hypomethylated loci in both the blood- and urine-derived cells are increased which is also causing an increase in oxidative stress in the bladder cancer patients [59]. In patients with myelodysplastic syndrome (MDS), two tumor suppressor genes, P15 and P16, have been found to be hypermethylated due to oxidative stress caused by peroxides, superoxide anion in leukocytes (CD45+ cells) from bone marrows of the patients [60]. This situation may be interpreted as hypermethylation of tumor suppressor genes cause them stop functioning thus increase the chance of cancer progression. Gu et al. have studied histone acetylation and DNA methylation, involved in the transcription of Alzheimer's disease-related genes in human neuroblastoma cells which is treated with hydrogen peroxide. They have found that expression of amyloid- β precursor protein and β -site amyloid- β precursor protein-cleaving enzyme 1 is upregulated due to demethylation in the gene promoters associated with the reduction of methyltransferases. They have also shown down-regulation of HDACs in the same cells. They have concluded that oxidative stress causes an imbalance between DNA methylation and demethylation, and histone acetylation and deacetylation [61]. In fact, both DNMTs and HDACs are vulnerable to oxidative stress due to their protein structure. Increased oxidative stress may inactivate them that results in impaired epigenetic pattern.

3.2. Oxidative Stress - Histone Post-Translational Modifications

Addition of a methyl or phosphate group to histones does not directly affect chemistry of histone. However, acetylation of histone tails directly alter gene expression that higher acetylation of the histones ends up with higher level of transcription. Probably due to this reason, studies examining the effect of oxidative stress on histone post-translational modifications have been focused on acetylation/deacetylation rather than other modifications. OFR- dependent signaling pathways prompt transcriptional and epigenetic dysregulation. Oxidative stress induces various kinase signaling pathways leading to histone acetylation/deacetylation. Activation of upstream kinases, such as nuclear factor- κ B (NF- κ B)-inducing kinase, mitogen and stress-activated kinase-1 by OFR results in downstream chromatin remodeling, which in turn alters the function of gene promoters [62]. In this context, OFR-mediated epigenetic alterations are especially shown to modulate transcription of specific genes that control apoptosis, autophagy, senescence, proliferation, transformation, and differentiation [63]. Although all histone modifying enzymes may be oxidized by OFR, sirtuins are more vulnerable due to their dependency to NAD^+ . When OFR interact with DNA, strand breakage occurs. This is a signal for polyADP ribose polymerase activation, the enzyme that uses NAD^+ . Increased oxidative stress may lead intracellular NAD^+ depletion and decreased sirtuin activity. Since sirtuins have a regulator function on various biological pathways, its low activity results in impaired cellular function. p53 is one of the major targets of SIRT1. Decreasing SIRT1 activity increases p53 acetylation, which in turn promotes apoptosis and senescence [64]. Oxidative stress has been shown to accelerate cellular senescence, due to increased p53 acetylation via decreasing the function of SIRT1 through NAD^+ depletion [65].

3.3. Oxidative Stress – miRNAs

miRNAs are not translated into a protein but play a crucial role in regulation of gene expression via translational repression or mRNA degradation. Although OFR-induced gene regulation has been extensively studied at epigenetic level, the effects of OFR on gene expression regulation by miRNAs is currently uncertain. Certain studies have shown that miRNAs are involved in regulation of cell survival in response to oxidative stress [66, 67]. The activation of apoptosis by Fas death receptor is stimulated by OFR. In an in vitro study Lin et al. have demonstrated that Fas is a functional target gene for miR-23a, forced overexpression of miR-23a reduces cell death induced by H_2O_2 incubation in macular retinal pigment epithelial cells, and this protective effect of miR-23a is disappeared by miR-23a inhibition [68].

In fact, it should be kept in mind that miRNAs are encoded by miRNA genes and their expressions are mostly regulated by epigenetic mechanisms. Epigenetic mechanisms may affect the expression of miRNAs in both physiological and pathologic conditions in a tissue-specific manner [69]. Hyper or hypo methylation of miRNA coding gene promoters leads to altered miRNA expression patterns, and this event may result in different diseases including cancer, glioblastoma, Parkinson's disease and Alzheimer's disease [70, 71]. Furthermore, recent studies have revealed that miRNAs are transcriptionally regulated by a variety of DNA methylation-specific methyl-CpG-binding domain proteins (MBDs) [72]. Besides DNA methylation, histone post-translational modifications and associated changes in chromatin

structure can affect miRNAs at the transcriptional level leading their both up- and down regulation [73].

CONCLUSION

Eukaryotic organisms including human beings are very complex structures and a variety of mechanisms interact with each other. Genetic and epigenetic changes work hand by hand to keep this fabulous factory working. Aberrant epigenetic modifications and regulations in post-transcriptional and post-translational processes make either direct or indirect contributions to the diseases such as cancer, neurodegenerative and inflammatory diseases. Oxidative stress is a state of persistent generation of OFR overwhelming cellular antioxidant defense, and it has long been known to contribute to pathogenesis of various diseases. Although harmful effects of oxidative stress on genes have been discovered a long time ago, OFR-epigenome interaction and its consequences have gained interest in the last decade. OFR-induced epigenetic mechanisms involve changes in DNA methylation profiles, histone acetylations and miRNA expression. Epigenetic modifications are potentially reversible and there is a great potential for the development of effective strategies to cure diseases. Scientist all over the world are studying on different mechanisms and interactions of genes and proteins to find cures for diseases. Understanding the role of oxidative stress on epigenetic mechanisms can help to see light at the end of the tunnel.

REFERENCES

- [1] Berger SL, Kouzarides T, Shiekhhattar R, Shilatifard A. An operational definition of epigenetics. *Genes Dev.* 2009; 23(7):781–3.
- [2] Lister R, Pelizzola M, Dowen RH, Hawkins RD, Hon G, Tonti-Filippini J, Nery JR, Lee L, Ye Z, Ngo QM, Edsall L, Antosiewicz-Bourget J, Stewart R, Ruotti V, Millar AH, Thomson JA, Ren B, Ecker JR. Human DNA methylomes at base resolution show widespread epigenomic differences. *Nature.* 2009; 462(7271):315-22.
- [3] Ghoshal K, Bai S. DNA methyltransferases as targets for cancer therapy. *Drugs Today.* 2007; 43(6):395–422.
- [4] Gopalakrishnan S, Van Emburgh BO, Robertson KD. *Mutat Res.* 2008; 647(1-2):30-8.
- [5] Miremadi A, Oestergaard MZ, Pharoah PD, Caldas C. Cancer genetics of epigenetic genes. *Hum Mol Genet.* 2007; 16:R28–49.
- [6] Jeltsch A, Nellen W, Lyko F. Two substrates are better than one: dual specificities for Dnmt2 methyltransferases. *Trends Biochem Sci.* 2006; 31(6):306-8.
- [7] Hermann A, Schmitt S, Jeltsch A. The human Dnmt2 has residual DNA-(cytosine-C5) methyltransferase activity. *J Biol Chem.* 2003; 278(34):31717–21.
- [8] Schaefer M, Lyko F. Solving the Dnmt2 enigma. *Chromosoma.* 2010; 119(1):35-40.
- [9] Goll MG, Kirpekar F, Maggert KA, Yoder JA, Hsieh CL, Zhang X, Golik KG, Jacobsen SE, Bestor TH. Methylation of tRNA^{Asp} by the DNA methyltransferase homolog Dnmt2. *Science.* 2006; 311(5759):395-8.

- [10] Okano M, Bell DW, Haber DA, Li E. DNA methyltransferases Dnmt3a and Dnmt3b are essential for de novo methylation and mammalian development. *Cell*. 1999; 99(3):247-57.
- [11] Suzuki MM, Bird A. DNA methylation landscapes: provocative insights from epigenomics. *Nat Rev Genet*. 2008; 9(6):465-76.
- [12] Cihák A. Biological effects of 5-azacytidine in eukaryotes. *Oncology*. 1974; 30(5):405-422.
- [13] Vaquero A, Loyola A, Reinberg D. The constantly changing face of chromatin. *Sci Aging Knowledge Environ*. 2003; 2003(14):RE4.
- [14] Bowman GD, Poirier MG. Post-translational modifications of histones that influence nucleosome dynamics. *Chem Rev*. 2015; 115(6):2274-95.
- [15] Tessarz P, Kouzarides T. Histone core modifications regulating nucleosome structure and dynamics. *Nat Rev Mol Cell Biol*. 2014; 15(11):703-8.
- [16] Jack AP, Hake SB. Getting down to the core of histone modifications. *Chromosoma*. 2014; 123(4):355-71.
- [17] Simon M, North JA, Shimko JC, Forties RA, Ferdinand MB, Manohar M, Zhang M, Fishel R, Ottesen JJ, Poirier MG. Histone fold modifications control nucleosome unwrapping and disassembly. *Proc Natl Acad Sci U S A*. 2011; 108(31):12711-6.
- [18] North JA, Javaid S, Ferdinand MB, Chatterjee N, Picking JW, Shoffner M, Nakkula RJ, Bartholomew B, Ottesen JJ, Fishel R, Poirier MG. Phosphorylation of histone H3(T118) alters nucleosome dynamics and remodeling. *Nucleic Acids Res*. 2011; 39(15):6465-74.
- [19] Lee KK, Workman JL. Histone acetyltransferase complexes: one size doesn't fit all. *Nat. Rev. Mol. Cell Biol*. 2007; 8 (4):284-95.
- [20] Bannister AJ, Kouzarides T. Regulation of chromatin by histone modifications. *Cell Res*. 2011; 21(3):381-95.
- [21] Gibson CL, Murphy SP. Benefits of histone deacetylase inhibitors for acute brain injury; a systematic review of animal studies. *J. Neurochem*. 2010; 115:806-813.
- [22] Volmar CH, Wahlestedt C. Histone deacetylases (HDACs) and brain function. *Neuroepigenetics*. 2015; 1:20-27.
- [23] Broide RS, Redwine JM, Aftahi N, Young W, Bloom FE, Winrow CJ. Distribution of histone deacetylases 1-11 in the rat brain. *J. Mol. Neurosci*. 2007; 31:47-58.
- [24] Kahali S, Sarcar B, Prabhu A, Seto E, Chinnaiyan P. Class I histone deacetylases localize to the endoplasmic reticulum and modulate the unfolded protein response. *FASEB J*. 2012; 26(6):2437-45.
- [25] Guan JS, Haggarty SJ, Giacometti E, Dannenberg JH, Joseph N, Gao J. HDAC2 negatively regulates memory formation and synaptic plasticity. *Nature*. 2009; 459; 55-60.
- [26] Graff J, Rei D, Guan JS, Wang WY, Seo J, Hennig KM. An epigenetic blockade of cognitive functions in the neurodegenerating brain. *Nature*. 2012; 483:222-226.
- [27] Lakshmaiah KC, Linu AJS, Apama DL, Smitha CS. Epigenetic therapy of cancer with histone deacetylase inhibitors. *J Cancer Res Thera*. 2014; 10(3):469.
- [28] Ellis L, Hammers H, Pili R. Targeting tumor angiogenesis with Histone Deacetylase Inhibitors. *Cancer Letters*. 2009; 280(2):145.
- [29] Li L, Sun Y, Liu J, Wu X, Chen L, Ma L, Wu P. Histone deacetylase inhibitor sodium butyrate suppresses DNA double strand break repair induced by etoposide more effectively in MCF-7 cells than in HEK293 cells. *BMC Biochem*. 2015; 16:2.

- [30] Campbell GR, Bruckman RS, Chu YL, Spector SA. Autophagy induction by histone deacetylase inhibitors inhibits HIV type 1. *J Biol Chem*. 2015; 290(8):5028-40.
- [31] Smith BC, Denu JM. Chemical mechanisms of histone lysine and arginine modifications. *Biochim Biophys Acta*. 2009; 1789:45–57.
- [32] Schnell JD, Hicke L. Non-traditional functions of ubiquitin and ubiquitin-binding proteins. *J Biol Chem*. 2003; 278(38):35857–60.
- [33] Popovic D, Vucic D, Dikic I. Ubiquitination in disease pathogenesis and treatment. *Nature Med* 2014; 20:1242–1253.
- [34] Komander D, Rape M. The ubiquitin code. *Annu Rev Biochem*. 2012; 81:203–29.
- [35] Dawson TM. Parkin and defective ubiquitination in Parkinson's disease. *J Neural Transm Suppl*. 2006; (70):209-13.
- [36] Guo G, Gui Y, Gao S, Tang A, Hu X, Huang Y, Jia W, Li Z, He M, Sun L. Frequent mutations of genes encoding ubiquitin-mediated proteolysis pathway components in clear cell renal cell carcinoma. *Nat Genet*. 2011; 44(1):17-9.
- [37] Dalle-Donne I, Aldini G, Carini M, Colombo R, Rossi R, Milzani A. Protein carbonylation, cellular dysfunction, and disease progression. *J Cell Mol Med*. 2006; 10(2):389–406.
- [38] Sharma R, Nakamura A, Takahashi R, Nakamoto H, Goto S. Carbonyl modification in rat liver histones: decrease with age and increase by dietary restriction, *Free Radic Biol Med*. 2006; 40:1179–1184.
- [39] Kim, VN. Small RNAs: classification, biogenesis, and function, *Mol Cells*. 2005; 19:1–15.
- [40] Friedman RC, Farh KK, Burge CB, Bartel DP; Farh BB. Most mammalian mRNAs are conserved targets of microRNAs. *Genome Res*. 2009; 19 (1): 92–105.
- [41] Rodriguez A, Griffiths-Jones S, Ashurst JL, Bradley A, Griffiths J, Ashurst B. Identification of mammalian microRNA host genes and transcription units. *Genome Res*. 2004; 14(10A): 1902–10.
- [42] Sato F, Tsuchiya S, Meltzer SJ, Shimizu K. MicroRNAs and epigenetics. *FEBS J*. 2011; 278(10):1598-609.
- [43] IUPAC Gold Book. Revised nomenclature for radicals, ions, radical ions and related species. *Pure&App/Chern*. 1993; 65(6):1357-1455.
- [44] Muller FL, Liu Y, van Remmen H. Complex III releases superoxide to both sides of the inner mitochondrial membrane. *J Biol Chem*. 2004; 279:49064–49073.
- [45] Mikhed Y, Daiber A, Steven S. Mitochondrial Oxidative Stress, Mitochondrial DNA Damage and Their Role in Age-Related Vascular Dysfunction. *Int J Mol Sci*. 2015; 16(7):15918-15953.
- [46] Zawia NH, Lahiri DK, Cardozo-Pelaez F. Epigenetics, oxidative stress, and Alzheimer disease. *Free Radic Biol Med*. 2009; 46(9):1241-9.
- [47] Kwok JB. Role of epigenetics in Alzheimer's and Parkinson's disease. *Epigenomics*. 2010; 2(5):671-82.
- [48] Prattichizzo F, Giuliani A, Ceka A, Rippo MR, Bonfigli AR, Testa R, Procopio AD, Olivieri F. Epigenetic mechanisms of endothelial dysfunction in type 2 diabetes. *Clin Epigenetics*. 2015; 7(1):56.
- [49] Yara S, Lavoie JC, Levy E. Oxidative stress and DNA methylation regulation in the metabolic syndrome. *Epigenomics*. 2015; 7(2):283-300.

- [50] Ziech D, Franco R, Pappa A. Reactive oxygen species (ROS)--induced genetic and epigenetic alterations in human carcinogenesis. *Mut Res-Gen Tox En*, 2011; 711:167-73.
- [51] Esteller M. Cancer epigenomics: DNA methylomes and histone modification maps. *Nat Rev Genet*. 2007; 8(4):286-98.
- [52] Feinberg AP, Tycko B. The history of cancer epigenetics. *Nat Rev Cancer*. 2004; 4(2):143-53.
- [53] Franco R, Schoneveld O, Georgakilas A, Panayiotidis MI. Oxidative stress, DNA methylation and carcinogenesis. *Cancer Lett*. 2008; 266:6-11.
- [54] Ushijima T, Takada EO. Aberrant methylations in cancer cells: where do they come from? *Cancer Sci*, 2005; 96:206-11.
- [55] Ushijima T. Detection and interpretation of altered methylation patterns in cancer cells. *Nat Rev Cancer*. 2005; 5:223-31.
- [56] Toyokuni A. Molecular mechanisms of oxidative stress-induced carcinogenesis: from epidemiology to oxygenomics. *IUMBM Life*. 2008; 60:441-7.
- [57] Georgakilas AG, Aziz K, Ziech D. BRCA1 involvement in toxicological responses and human cancer etiology. *Toxicol Lett*, 2009; 188:77-83.
- [58] Kriaucionis S, Heintz N. The nuclear DNA base 5-hydroxymethylcytosine is present in Purkinje neurons and the brain. *Science*. 2009; 324(5929):929-30.
- [59] Patchsung M, Boonla C, Amnattrakul P, Dissayabutra T, Mutirangura A, Tosukhowong P. Long interspersed nuclear element-1 hypomethylation and oxidative stress: correlation and bladder cancer diagnostic potential. *PLoS One*. 2012; 7(5):e37009.
- [60] Gonçalves AC, Cortesão E, Oliveiros B, Alves V, Espadana AI, Rito L, Magalhães E, Pereira S, Pereira A, Costa JM, Mota-Vieira L, Sarmiento-Ribeiro AB. Oxidative stress levels are correlated with P15 and P16 gene promoter methylation in myelodysplastic syndrome patients. *Clin Exp Med*. 2015. [Epub ahead of print]
- [61] Gu X, Sun J, Li S, Wu X, Li L. Oxidative stress induces DNA demethylation and histone acetylation in SH-SY5Y cells: potential epigenetic mechanisms in gene transcription in A β production. *Neurobiol Aging*. 2013; 34(4):1069-79.
- [62] Sundar IK, Yao H, Rahman I. Oxidative stress and chromatin remodeling in chronic obstructive pulmonary disease and smoking-related diseases. *Antioxidant Redox Signal*. 2013; 18(15):1956-71.
- [63] Hoshino S, Yoshida M, Inoue K, Yano Y, Yanagita M, Mawatari H, Yamane H, Kijima T, Kumagai T, Osaki T, Tachiba I, Kawase I. Cigarette smoke extract induces endothelial cell injury via JNK pathway. *Biochem Biophys Res Commun*. 2005; 329:58-63.
- [64] Vaziri H, Dessain SK, Ng Eaton E, Imai SI, Frye RA, Pandita TK, Guarente L, Weinberg RA. hSIR2(SIRT1) functions as an NAD-dependent p53 deacetylase. *Cell*. 2001; 107:149-159.
- [65] Furukawa A, Tada-Oikawa S, Kawanishi S, Oikawa S. H₂O₂ accelerates cellular senescence by accumulation of acetylated p53 via decrease in the function of SIRT1 by NAD⁺ depletion. *Cell Physiol Biochem*. 2007; 20:45-54.
- [66] Cheng Y, Liu X, Zhang S, Lin Y, Yang J, Zhang C. MicroRNA-21 protects against the H(2)O(2)-induced injury on cardiac myocytes via its target gene PDCD4. *J Mol Cell Cardiol*. 2009; 47:5-14.
- [67] Tang Y, Zheng J, Sun Y, Wu Z, Liu Z, Huang G. MicroRNA-1 regulates cardiomyocyte apoptosis by targeting Bcl-2. *Int Heart J*. 2009; 50:377-387.

- [68] Lin H1, Qian J, Castillo AC, Long B, Keyes KT, Chen G, Ye Y. Effect of miR-23 on oxidant-induced injury in human retinal pigment epithelial cells. *Invest Ophthalmol Vis Sci.* 2011; 52(9):6308-14.
- [69] Sato F, Tsuchiya S, Meltzer SJ, Shimizu K. MicroRNAs and epigenetics. *FEBS J.* 2011; 278:1598-1609).
- [70] Baer C, Claus R, Frenzel LP, Zucknick M, Park YJ, Gu L, Weichenhan D, Fischer M, Pallasch CP. Extensive promoter DNA hypermethylation and hypomethylation is associated with aberrant microRNA expression in chronic lymphocytic leukemia. *Cancer Res.* 2012; 72:3775-3785.
- [71] Li J, Harris RA, Cheung, SW, Coarfa C, Jeong M, Goodell MA, White LD, Patel A, Kang SH. Genomic hypomethylation in the human germline associates with selective structural mutability in the human genome. *PLoS Genet.* 2012; 8; e1002692.
- [72] Saito Y, Saito H.. MicroRNAs in cancers and neurodegenerative disorders. *Front Genet.* 2012; 3:194).
- [73] Brait M, Sidransky D. Cancer epigenetics: above and beyond. *Toxicol Mech Methods.* 2011; 21:275e288.

Chapter 61

EPIGENETIC MODIFICATIONS AND THEIR POTENTIAL ROLE IN TUMORIGENESIS

***Muthu K. Shanmugam¹, PhD, Frank Arfuso², PhD
and Gautam Sethi^{1,2,*}, PhD***

¹ Department of Pharmacology, Yong Loo Lin School of Medicine,
National University of Singapore, Singapore

² School of Biomedical Sciences, CHIRI Biosciences Research Precinct,
Curtin University, Western Australia, Australia

ABSTRACT

Cancer is one of the deadliest malignancies that have plagued mankind for decades. Epigenetic mechanism(s) play a central role in the homeostasis of normal cell proliferation and differentiation. Global epigenetic modifications are frequently associated with cancer initiation, progression, as well as metastasis. These changes include DNA methylation, histone lysine methylation/demethylation, acetylation/ deacetylation, including methylation and acetylation of non-histone proteins, and can alter the expression of various oncogenic signaling cascades, which in-turn can lead to uncontrolled proliferation. In this chapter, we primarily focus on the major epigenetic changes that occur in oncogenes, tumor suppressor genes, transcription factors and cancer stem cells, which in turn mediate tumor growth. These modifications are controlled by regulatory enzymes such as DNA methyltransferase, histone acetyltransferases, histone deacetylase, lysine acetyltransferase, and arginine and lysine methyl transferases. In addition, we also describe a few selected pharmacological agents that can modulate the action of these enzymes and display significant potential for cancer therapy.

Keywords: cancer, epigenetics, transcription factors, oncogenes

* Corresponding Author's Email: gautam_sethi@nuhs.edu.sg.

ABBREVIATIONS

ALL	acute lymphocytic leukemia
AML	acute myeloid leukemia
APC	adenomatosis polyposis coli
ASH2L	absent, small or homeotic discs-like 2
AR	androgen receptor
BRCA1	breast cancer 1, early onset
COX	cyclooxygenase
CML	chronic myelogenous leukemia
CARM1	coactivator associated arginine methyltransferase 1
Cdk9	cyclin dependent kinase 9
CDKN2	cyclin dependent kinase 2
CpG	cytosine-phosphate-guanine
CREB	c-AMP response element binding protein
CSC	cancer stem cell
CDH1	E-cadherin
CREB	cAMP response element binding
CBP	cAMP response element binding (CREB) protein
CTL	cytotoxic T-Lymphocytes
DAPK1	death associated protein kinase 1
DLBCL	diffuse large B-cell lymphoma
DNMT	DNA methyltransferase
DR	death receptor
EZH2	enhancer of zeste 2 polycomb repressive complex 2 subunit
ERRB2	human epidermal growth factor receptor 2
EMT	epithelial-to-mesenchymal transitions
ER	estrogen receptor
FDA	food and drug administration
FBXL11	F-box and leucine-rich repeat protein 11
GNAT5	Gcn5-related N-acetyltransferase
GSTP1	glutathione S-transferase-pi
GM-CSF	granulocyte macrophage colony stimulating factor
GSK3 β	glycogen synthase kinase 3 β
HCC	hepatocellular carcinoma
HIF1	hypoxia-inducible factor
HRE	HIF1 α binding to the HIF1 α -response element
H3K27me3	trimethylated lysine 27 on histone H3
H3K4me	
1/2/3	mono-di-tri-methylated histone H3 at lysine 4
H3K64me3	trimethylated histone 3 at lysine 64
H3K9ac	acetylated histone H3 at lysine 9
H3K9me	
1/2/3	mono-di-tri-methylated histone H3 at lysine 9
H4K20me3	trimethylated histone 4 at lysine 20

HMG	high mobility group
HSC	hematopoietic stem cells
IGF	insulin like growth factor
IKK	I κ B kinase
IL	interleukin
JAK	janus kinase
JmJc/N	jumonji-domain-containing protein histone lysine demethylases
KAT	histone lysine acetyltransferases
KATi	histone lysine acetyltransferases inhibitor
KDAC	histone lysine deacetylase
KDACi	histone lysine deacetylase inhibitor
KMT	histone lysine methyltransferases
KDMs	histone lysine demethylases
MBD	methyl-binding domain
MBD2	methyl-CpG-binding domain protein 2
MBD3	methyl-CpG binding domain protein 3
MHC	major histocompatibility complex
MLL	histone methyl transferase
MOZ	monocytic leukaemia zinc finger protein
MKP1	mitogen activated protein kinase (MAPK) phosphatase-1
NSD1	nuclear receptor binding Su(var)3–9, enhancer of zeste and Trithorax (SET) domain protein 1
iNOS	inducible nitric oxide synthase
NF- κ B	nuclear factor kappaB
NFAT	nuclear factor of activated T cell
PG	prostaglandins
PPAR- γ	proliferator-activated receptor gamma
PTEN	phosphatase and tensin homolog
PARP-1	poly-ADP-ribose-polymerase 1
PCAF	p300/CBP associated factor
PcG	polycomb group of proteins
PR	progesterone receptor
PRC2	polycomb repressive complex 2
PRMT	protein arginine methyl transferases
PHF	plant homeodomain (PHD)-finger protein
PKMT	protein lysine methyl transferase
PKDM	protein lysine demethylases
RASSF1	ras association domain-containing gene
RelA	v-rel avian reticuloendotheliosis viral oncogene homolog A
SAM	s-adenosyl methionine
SMYD	SET and MYND-domain containing
SIRT	sirtuin
Skp2	S-phase kinase associated protein 2
STAT3	signal transducer and activator of transcription
TLR	toll like receptor

TGF β	transforming growth factor beta
TIMP-3	tissue inhibitor of metalloproteinase 3
TNF α	tumor necrosis factor alpha
TSA	trichostatin A
VHL	Von Hippel–Lindau

1. INTRODUCTION

Historically, epigenetics is defined as heritable changes in gene expression patterns that occur without any changes to the DNA but involve changes that are adequate to regulate activation or repression of gene expression. The term epigenetics was first introduced by developmental biologist Conrad H. Waddington in 1942. The latest revised definition of epigenetics includes the transient changes that occur during gene expression [1]. Epigenetics is an important physiological process that occurs all through the development and differentiation of any organisms that have the same set of DNA sequences, and drives them to diversify into different cell types [2]. In addition to its pivotal role in the development of an organism, deregulated epigenetics mechanisms are often encountered in the development and severity of disease. It is now increasingly apparent that chronic inflammation-driven diseases occur due to both genetic and epigenetic alterations [3, 4]. It was Rudolf Virchow in 1870 who linked inflammation with cancer, diabetes, obesity, arthritis, allergy, and atherosclerosis [5]. From then on the connection between inflammation and cancer development has been frequently observed in chronic inflammatory bowel disease and colorectal cancer [6], Barrett's esophagitis and esophageal cancer [7], and in virus-induced cancers such as human papilloma virus-induced cervical cancer, hepatitis B or C virus-induced liver cancer [8]. It has been estimated that 15% of all cancers are linked to inflammation [6]. In addition, environmental factors such as exposure to radiation are also key contributors in cancer progression [7]. The first and foremost defense response to any type of injury to the human body is the recruitment of immune cells such as leukocytes, lymphocytes, neutrophils, monocytes, and macrophages. These cells reduce the inflammatory reaction in an acute attack. However, in chronic inflammation, pro-inflammatory molecules are produced and secreted by these cells, which produces a cascading effect and thus, has been implicated in the development of various major human diseases including malignancies [9]. Thus, excessive production of pro-inflammatory molecules such as cytokines, chemokines, prostaglandins (PGs), nitric oxide (NO), and leukotrienes deregulate the cellular signaling and transcriptional mechanism(s), which ultimately leads to the development of neoplasms [10]. Thus, changes to the epigenome mediated by various enzymes are now been recognized as being important mediators in the activation or silencing of genes, the impact of which are as significant as heritable permanent genetic mutations in the DNA [11]. The impact of epigenetic alterations in disease progression has led pharmaceutical companies to target these histone and non-histone modifying enzymes for anti-cancer drug development [12-14].

In the eukaryotic cell nucleus, DNA is wrapped around the core histone octamer H2A, H2B, H3 and H4, thus forming the fundamental unit of chromatin, the nucleosome [15]. Post-translation modifications on core histone tails, along with cytosine methylation of the DNA, determines the accessibility of the chromatin and thereby enhancing the ability of transcription

factors to bind to DNA and initiate gene transcription [15]. These transient changes are unstable and disappear very rapidly, and can be easily modulated by any external stimulus [15, 16]. In contrast to transient change, permanent change to the DNA often leads to underdeveloped or defective organs and to the development of diseases. These heritable permanent DNA changes often occur at specific dinucleotide sites along the genome such as the cytosine 5' of guanine (CpG) sites. It has been observed that 40% of genes contain clusters of CpG dinucleotide upstream of the transcription start site and are often found to be 70-80% methylated [17, 18]. Escalation in DNA methylation is often seen in chronic inflammation-driven diseases, viral infections, and cancers. Thus, aberrant DNA methylation now serves as a biomarker for the early detection of cancer and can also be related to the therapeutic effectiveness of drugs affecting DNA methylation and histone acetylation [4]. Epigenetic modifications of chromatin and DNA have in recent times been acknowledged as being either the expressive or suppressive factors in directly controlling gene transcription [18]. The post-translation modification on histone proteins in chromatin and methylation of DNA are regulated by two distinct pathways [18]. Environmental factors influence the epigenome at all times during the development of human life, especially nutritional factors, which can have profound effects in the expression of specific genes by epigenetic modifications [19]. The epigenome is characterized by global hypomethylation, but distinct CpG islands located in the regulatory regions of genes can be hypermethylated, leading to repression of gene expression [20]. Based on the levels of methylation status of CpG islands, monoallelic silencing and other epigenetic regulatory mechanisms have been reported in key inflammatory response genes [11, 18]. Histone hypermethylation silences genes involved in cell cycle control (*CDKN2*) [21], apoptosis (*DAPK1*) [22], DNA repair (*BRCA1*) [23], metastasis (*TIMP3*) [24], drug resistance, tumor heterogeneity, stem cell-like state [25], differentiation, epithelial-mesenchymal transition, metabolism [26], and VEGFR2 [27]. This global phenomenon of CpG hypermethylation is a common biological pathway to switching off genes. While DNA mutations have been linked to chronic inflammation-driven cancers, it is also observed that epigenetic alterations can contribute to familial cancer risk [28]. Chronic inflammation mediates the initiation of a variety of diseases including cancer progression, which can be explained on the basis of epigenetic dysregulation [29-32]. Recent studies have indicated that DNA hypermethylation and histone modifications play important roles in gene silencing [33].

Selective silencing of DNA methyltransferase leads to acute fatty acid-induced, non-CpG methylation of proliferator-activated receptor gamma (PPAR- γ) co-activator 1 alpha promoter [34]. In this chapter, we primarily focus on the major epigenetic changes that occur in oncogenes, tumor suppressor genes, transcription factors, and cancer stem cells, which in turn mediate tumor growth. These modifications are controlled by regulatory enzymes such as DNA methyltransferase, histone methylases, histone demethylases, histone lysine acetyltransferases, histone deacetylases, and arginine and lysine methyl transferases. In addition, we also describe a few selected pharmacological agents that can modulate the action of these enzymes and display significant potential for cancer therapy.

2. THE DYSREGULATION OF EPIGENETIC FACTORS IN VARIOUS CANCERS

2.1. The Deregulation of DNA Methylation in Cancer

DNA methylation is a widespread epigenetic modification in eukaryotic cells. DNA methylation occurs at the 5' end of the CpG dinucleotide of the cytosine ring, with S-adenosyl-L-methionine as its methyl donor [35, 36]. This reaction is catalyzed by a family of closely related DNA methyltransferases (DNMT), DNMT1, DNMT3A, and DNMT3B [37]. DNMT1 is the most abundant DNMT in mammals and maintains the methylation status on CpG dinucleotides [38]. In normal eukaryotic cells, the CpG islands are usually found in the unmethylated state to maintain euchromatin structure, and thereby allowing gene expression. However, during cancer development, many of the genes are hypermethylated at their CpG islands and silence gene transcription by changing the open euchromatic structure to a compact heterochromatic structure that is repressive for transcription [39]. DNMT1 predominantly adds a methyl group to DNA when one strand is already methylated (hemi-methylated) [40]. Once methylated they recruit the methyl-CpG-binding domain (MBD) proteins such as Kaiso, MeCP2, and members of the MBD family, which recognize methylated CpG islands and initiate chromatin silencing [41].

In cancer, promoter DNA hypermethylation has been observed for over two decades with the subsequent silencing of tumor suppressor genes [42, 43]. Global hypomethylation 'shores' also occur in cancer cells and are associated with genomic instability [44]. Both hypomethylation and hypermethylation have been seen across different types of cancers [45]. In cancer cells, hypermethylation are often observed in the promoter regions of tumor suppressor genes such as retinoblastoma gene (RB1), phosphatase and tensin homolog (PTEN), and DLC-1, glutathione S-transferase-pi (GSTP1), and E-cadherin (CDH1), thereby resulting in loss-of-function in colon cancers [46, 47]. Hypermethylation targeting of specific genes such as p16, p15, MINT1, MINT2, MINT31, MLH1, MSH2, MGMT, DAPK, APC, LKB1, IGF2, and COX2 has also been detected in colon cancer [47-49]. Global methylations have been observed in tumors of breast, prostate, lung, as well as in hematological malignancies [50, 51]. Overexpression of DNMT1 is widely reported in leukemias [51], and mutation and hypermethylation of this gene is observed in a variety of hematological malignancies [52]. Hypermethylation of genes such as GSTP1, RASSF1A, APC, MGMT, CCNA1, CDH1, RUNX3, ESR2, CDKN2A, RAR β 2, IGFBP7, and SFRP2 are all implicated in prostate cancer development and progression [53, 54], including E-cadherin, CD44, galectins [55], ER, androgen receptor, retinoic acid receptor, and genes involved in apoptosis [56], tissue inhibitors of metalloproteinases (TIMPs) [57], and cyclinD kinases [58]. In breast cancer, more than 100 genes have been hypermethylated in all type of cell processes including cell cycle regulation genes CCND2, p16ink4A/CDKN2A [59], apoptosis genes such as APC, TWIST, and HOXA5 [60, 61], metastasis, invasion genes such as RAR β 2, CDH1, ER, BRCA1, CCND2, p16 and TWIST, angiogenic genes VEGFR2, eNOS and in cell signaling pathways [62]. The estrogen receptor (ER) α and progesterone receptor (PR), which are critical in hormone regulation, are also frequently methylated in breast cancer [63]. DNA hypomethylation is also frequently reported in breast cancer, especially on FEN1 BCSG1, PLAUR, IGF2, and CDH3 [64, 65]. In addition, recent research has shown that miRNA that function as tumor suppressors can be

silenced via DNA methylation [66]. Lower levels of genome-wide DNA hypomethylation have been observed in metastatic prostate cancer in contrast to non-metastatic prostate cancer [67, 68]. Gene-specific hypomethylation also occurs in urokinase plasminogen activator (uPa), cell proliferation genes, and in various cellular processes such as invasion and metastasis [69], cancer/testis antigen gene [70], hydroxylation of estrogens [71], and X-chromosome inactivation [72]. In pancreatic carcinoma, DNA hypomethylation occurs in the promoter regions of the RASSF1A gene and K-Ras mutations [73]. Similarly, global hypomethylation and hypermethylation are also observed in lung cancer [74]. DNA hypomethylation in H19, IGF2, and MEST genes results in abnormal lung cancer growth. In NSCLC cancer, testis antigen and melanoma-associated antigen were upregulated due to hypomethylation [75]. Hypermethylation of tumor suppressor genes such as *p53*, *CDKN2*, *p16INK4* [21], *MGMT* [76], *DAPK*, *caspase 8*, *ARF*, *FAS*, *TRAILR1* [76, 77], *RASSF1A*, *NORE1A*, *G0S2*, *cadherins*, *TIMP3* [78, 79] may also result in their silencing.

2.2. The Deregulation of Histone Lysine Methylation in Cancer

ϵ -*N*-methyl-lysine was first found in a bacterial flagellar protein in 1959 [80] and subsequently, in 1964, this post-translational modification (PTM) was also identified in histones [81]. Other PTMs include acetylation, ubiquitination, phosphorylation, sumoylation, and ribosylation at histone tails [82-84]. Several reports have indicated that PTM do influence the density of chromatin and thus accessibility to DNA and gene transcription [84]. Core histones H2A, H2B, H3, and H4 combine to form the basic structure of the histone octamer, the nucleosome [84, 85]. Mono, di-, or tri-methylation of histones takes place in the amino acid of the N-terminal tail of histones that are prone to covalent modifications. It has been shown that methylation predominantly occurs in histone H3 or H4 lysine residues such as H3K4, H3K9, H3K27, H3K36, and H4K20, and one lysine residue located in H3 (H3K79) [86, 87]. Nuclear receptor binding Su (var)3-9, Enhancer of zeste and Trithorax (SET) domain protein 1 (NSD1) is the largest protein of the NSD family of histone lysine methyltransferases. It is located on chromosome 5q35 and consists of 2696 aa [88]. In the multi-domain NSDs, it is the SET domain that is responsible for methyltransferase activity. The NSD protein lysine methyltransferases (KMTs) family is composed of three members, NSD1 (KMT3B), NSD2 (WHSC1/MMSET), and NSD3 (WHSC1L1), which regulate gene expression through methylation of lysine 36 on histone H3 (H3K36) [89]. An accumulation of evidence has unveiled the role of NSDs in a variety of cancers due to genetic alterations or aberrant overexpression. Numerous studies have reported the role of NSD1 with a distinct *HOX* gene expression pattern in acute myeloid leukemia (AML) [90-93]. In addition to NSD1, NSD2 has also been implicated in cancer pathogenesis. NSD2 mRNA was found to be overexpressed in 13 types of cancers [94], and overexpression of NSD2 protein in 3774 tumor samples of colon, small cell lung, skin, multiple myeloma, and bladder cancer [88, 95]. In breast, prostate, oligodendroglioma, and head and neck cancers, NSD2 overexpression was associated with higher tumor grade [96]. A high level of NSD2 is an indicator of poor overall and disease free survival in hepatocellular carcinoma and endometrial cancer [97, 98]. In addition, NSD3 has been identified as the fourth most frequently amplified methyltransferase in breast cancer [99, 100]. However, histone tri-methylation on the lysine 4 residue of histone 3 (H3K4me3) leads to gene activation while tri-methylation on lysine 9 (H3K9me3) or on lysine 27 (H3K9me3)

leads to gene repression [19]. It is of note that histone methylation often leads to a condensed chromatin structure and thereby suppress gene expression [101]. EZH2 (KMT6) catalyzes the trimethylation of histone H3K27me3 [102] and initiates gene silencing via local chromatin reorganization. Several studies suggest that EZH2 plays a critical role in initiation, progression, and metastasis of prostate, breast, bladder, ovarian, renal carcinoma, lung, liver, brain, gastric, esophageal, pancreatic, and melanoma cancer, and also in cancer stem cells [102, 103]. In solid tumors, a hypoxic microenvironment induces hypoxia-inducible factor (HIF1) α binding to the HIF1 α -response element (HRE) and transactivates EZH2 transcription [104]. There are also other methyltransferases that are overexpressed in variety of cancers, such as KMT2A in human lymphoid and myeloid leukemias [105, 106], KMT1A in colon cancer [107], KMT1C in lung carcinoma [108, 109], Ash2L in hematological malignancies, hepatocellular carcinoma (HCC) and gastrointestinal cancers [110, 111], and SET and MYND-domain containing (SMYD)-3 in breast, colon, and HCC [112, 113]. Other lysine methyltransferases that are either mutated or downregulated in cancers are menin, KMT3B, and KMT8 [39, 114, 115].

2.3. The Deregulation of Histone Lysine Demethylases in Cancer

Lysine demethylases are a diverse group of demethylases. The lysine-specific histone demethylase (LSD) family and JmjC-domain containing histone demethylases (JHDMs) protein family have been identified as the two major families of demethylases that regulate histone demethylation, which are responsible for removing the methylation mark on lysines [82, 116-124]. LSD1 (KDM1A) and LSD2 (KDM1B) are flavin-dependent monoamine oxidases that can only demethylate mono- and di-methyl lysine and not tri-methyl lysine residues [117, 122, 125]. LSD1 demethylates H3K4me1/2 and H3K9me1/2, and non-histone proteins such as p53, E2F1, and DNMT1 [124]. LSD1 is overexpressed in prostate, breast, bladder, lung, colon, and neuroblastoma cancers [126], and is under expressed in HCC [111, 127]. LSD1 is also associated with mixed lineage leukemia (MLL) [128], and sustains the growth of MLL-AF9 acute myeloid leukemia cells by preventing differentiation and apoptosis [129], and serves as a biomarker for aggressive malignancies and tumor relapse [130]. It is of note that LSD1 upregulates cyclin A2 (*CCNA2*) and human epidermal growth factor receptor 2 (*ERBB2*) in aggressive breast cancer [130]. LSD2 is highly overexpressed in breast cancer and in invasive tumors, in contrast, a decreased expression of LSD2 is seen in leukemia and in seminoma [131]. The JmjC domain containing demethylases can demethylate mono-, di-, and tri-methyl lysine residues [120, 121, 124, 132]. Around 30 JmjC domain containing proteins have been identified in the human genome [133] and are divided into six sub-families based on histone lysine sites and demethylation states. They are KDM2, KDM3, KDM4, KDM5, KDM6, and plant homeo domain (PHD)-finger protein (PHF) [134]. JmjC domain-containing proteins belong to the Fe (II) and 2-oxoglutarate (2-OG)-dependent dioxygenases. They demethylate di- or tri-methylation marks on H3K4, H3K9, H3K27, H3K36, and non-histone proteins [124]. The KDM2 subfamily is comprised of KDM2A and KDM2B, which can demethylate H3K36A/B and the non-histone protein nuclear factor kappa-light-chain-enhancer of activated B cells (NF- κ B) [135]. KDM2A/B is also involved in cellular processes such as ribosomal RNA transcription, cell proliferation, apoptosis, differentiation, and carcinogenesis [136-138]. Overexpression of KDM2A/B is observed in bladder cancer, T-acute lymphocytic leukemia

(T-ALL), B-ALL, AML, breast, pancreatic adenocarcinoma, and seminoma. In contrast, reduced expression of KDM2A/B is observed in prostate cancer and glioblastoma [124, 139, 140]. In the human genome, the KDM4 family is sub-divided into two major groups based on their domain structure. KDM4A, KDM4B, and KDM4C form the first group while KDM4D, KDM4E, and KDM4F form the second group. The first group contains two PHD domains and two tudor domains in addition to the JmjC and JmjN domains, whereas the second group has only JmjC and JmjN domains [141]. KDM4A/B/C demethylates both histone and non-histone protein methylation such as WIZ, CDYL1, and G9a [142]. Therefore, KDM4 demethylases act as either repressors or activators of transcription and are involved in regulating various cellular processes including cell cycle and DNA damage repair, as well as tumor initiation and progression [124, 143, 144]. Numerous evidence suggests that overexpression of KDM4 demethylase is involved in breast, prostate, lymphoma, gastric, lung, bladder, medulloblastomas, and colorectal cancers, whereas mutational deletion or knockdown of KDM4A/B/C inhibits cancer cell proliferation and growth [133, 144-146]. The KDM5 family contains KDM5A, KDM5B, KDM5C, and KDM5D. This enzyme removes the methylation marks on H3K4me_{2/3} [132]. KDM5s are also implicated in cancer development. KDMs are either overexpressed (KDM5A) or downregulated (KDM3B) in breast cancers [147]. Differences in KDMs have been shown to be associated with aberrant histone methylation and demethylation of H3K4 and shorter relapse free survival [147]. The abundance of methyl moieties on the lysine residue determines cancer prognosis. For example, increased H3K4me₂ is found in cancerous cells but not in normal cells, and is a biomarker for tumor recurrence in prostate cancer [148-150]. Overexpression of KDM5A has been observed in lung and gastric cancers [151, 152], while overexpression of KDM5B was detected in bladder, AML, lung, breast, chronic myelogenous leukemia (CML), cervical, renal, metastatic prostate cancer, and testicular cancer [153]. Knockdown of KDM5B by siRNA results in inhibition of cancer growth [153, 154]. The KDM6 subfamily contains KDM6A, KDM6B, and UTY (KDM6C), which all exhibit high levels of catalytic activity on H3K27me_{2/3} [155, 156]. Several oncogenic mutations on KDM6A are observed in the cells of multiple myeloma, renal cell carcinoma, and esophageal squamous cell carcinoma [157, 158]. In addition, KDM6B can activate the tumor suppressor *p16INK4A* and *p14ARF*; hence, any mutation or deregulation in these genes might lead to carcinogenesis [159].

3. HISTONE LYSINE ACETYLATION IN CANCER

In addition to the other post-translational modifications, the physiological roles of N-ε-lysine protein acetylation have been identified in tumorigenesis [160]. Some histone acetylation marks have been associated with compaction of chromatin [161], DNA repair [162], protein stability, and protein-protein interaction [82], and lysine acetylation has emerged as the ubiquitous post-translational modification that is found across the entire proteome [163, 164]. Public repositories such as phosphositeplus database show approximately 15,000 lysine acetylation sites in human cells, with prominent roles in signaling networks [165]. Acetylation was initially identified 45 years ago primarily in histones, but lysine acetylation also occurs in non-histone proteins such as transcription factors, chromatin regulators, cytoplasmic proteins

involved in the regulation of cytoskeleton dynamics, and in proteins involved with energy metabolism and endocytosis [163, 164, 166, 167].

3.1. The Deregulation of Histone Lysine Acetyltransferases in Tumor Proliferation and Metastasis

Access to DNA that is wrapped around histones occurs via post-translational modifications such as lysine acetylation. Approximately 10-20 lysine residues get acetylated per histone by lysine acetyltransferases (KATs), which add an acetyl group to the lysine residue on histone. Acetylation of histone increases the negative charge on DNA, thus facilitating the DNA accessibility to proteins involved in replication, transcription, or DNA repair [168-170]. The most understood set of KATs are GNAT5 (Gcn5-related N-acetyltransferase), PCAF (p300/CBP associated factor), Hat1, Elp3, and Hpa2, all of which are involved in transcription initiation. cAMP response element binding (CREB) protein (CBP) and p300 are the two most abundant KATs, both proteins are regulators of RNA polymerase II-mediated transcription, and are ~300 kDa proteins possessing structural homologues with high sequence similarity [171]. Additionally, both KATs acetylate at multiple sites on each of the four core histones H2A, H2B, H3, and H4, and regulate human growth and development [172]. Inactivation or mutation on either of both KATs leads to abnormal organ development and diseases in humans, while insufficiency in p300 leads to aberrant levels of acetylation in multiple cancers [173-177]. The germ line mutation of CBP results in Rubinstein-Taybi syndrome with increased susceptibility to childhood malignancies [171]. Therefore, it is vital to understand the divergent functions such as specificity and selectivity for lysine residues on core histones in treating the disease that they cause. Constitutive activation of transcription factors are observed in all types of cancers. In particular, acetylation of K9, K14, K18, and K56 of histone H3, and K5, K8, K13, and K16 of histone H4 are strongly involved in transcriptional activation [15, 178]. Histone lysine acetyltransferase is known to play important roles in normal and malignant hematopoiesis by acetylating histone and non-histone proteins. In normal hematopoietic stem/progenitor cells, myeloid progenitor cells, and lymphoid cells, protein acetylation by KATs such as p300/CBP, MOZ, HBO1, and GCN5 regulates hematopoietic stem cell (HSC) self-renewal, proliferation, and their differentiation into committed hematopoietic progenitors [179]. Several studies have identified aberrant KAT activity on oncogenes and tumor suppressor proteins that have been shown to be involved in the development of hematological malignancies. In AML, T-cell leukemia, lymphoma, B-cell ALL, and CML leukemogenic proteins physically interact with the KATs p300/CBP, PCAF, MOZ, Tip60, and GCN5, and target cell cycle genes, oncogenes, and tumor suppressor genes such as *Id1*, *p21*, *Egr1*, *c-Myc*, *p53*, *p53*, *RAR β* , *PU.1*, *Syk*, *Btk*, and *AML1* that promote malignant transformation of cells [179]. Over a dozen human lysine acetyltransferases have been identified. Two noted KATs are MOZ (monocytic leukemia zinc finger protein; a.k.a. MYST3 and KAT6A) and its paralog MORF (a.k.a. MYST4 and KAT6B) form tetrameric complexes with BRPF1 (bromodomain- and PHD finger-containing protein 1) and two small non-catalytic subunits. These two acetyltransferases and BRPF1 are found to be recurrently mutated in leukemia, non-hematologic malignancies, and in multiple developmental disorders. The BRPF1 gene is mutated in childhood leukemia and adult medulloblastoma [180]. In addition to hematologic malignancies, the MOZ and MORF gene is mutated in esophageal adenocarcinoma [181] and

metastatic medulloblastoma [182]. The MORF gene is altered in leiomyomata [183, 184], breast cancer [185], and castration-resistant prostate cancer [186]. Moreover, both the MOZ and MORF genes have been identified as very common KATs amplified in different types of cancers [187], suggesting that they have an oncogenic role. Deletion of CBP/p300 specifically reduces acetylation on H3K18ac and H3K27ac in mouse embryonic fibroblasts [188]. Mice heterozygous for CBP spontaneously develop myelodysplastic and/or myeloproliferative tumors [189]. Genetic mutations/deletions of CBP/p300 are correlated with the development of transitional cell bladder cancer [190], follicular lymphoma, and DLBCL [191], relapsed ALL [192] and colon cancer [193]. Overexpression of CBP/p300 is a poor prognostic indicator in small-cell lung cancer [194], and is associated with higher grade tumor with distant metastasis of nasopharyngeal tumor [195]. In melanoma patients, decreased expression of p300 is associated with poor prognosis [196]. CBP/p300 can also induce tumorigenesis independently by acetylating S-phase kinase associated protein 2 (Skp2) [197]. Reduced expression of Tip60 is an indicator for poor prognosis in primary and metastatic melanoma patients [198], and it is also downregulated in colon, lung, head and neck carcinoma, metastatic prostate cancer, and lymphoma [199-204]. In hormone refractory or androgen independent prostate cancer, Tip60 accumulates in the nucleus [202]. In castration resistant prostate cancer LNCaP derived CxR cells; Tip60 is upregulated in the nucleus, which results in the activation of transcription even in the absence of androgen [205]. ER α and PRs have been implicated in breast cancer initiation and progression, and increased activation of ER α is a risk factor for the development of breast cancer by increasing cell proliferation and growth of abnormal breast tissue. ER α is a transcription factor that is activated by hormonal and growth factor signals. Several reports have highlighted the importance of ER α post-translational modification. It has been shown that ER α is acetylated by co-activator p300 at lysine in the hinge/ligand domain at K229, K299, K302, and K303 [206]. In atypical breast hyperplasia samples, lysine to arginine substitution was found at residue K303R in ER α , with increased estradiol dependent activation [207]. Another study demonstrated that, in the presence of p160 and p300 co-activators, ER α was acetylated at K266 and K268 [208]. In breast cancer cells where the tumor suppressor gene BRCA1 is mutated, p300/CBP acetylates ER α and drives cell proliferation, while reintroduction of BRCA1 downregulates p300 and not CBP [209]. Prostate cancer is one of the most frequently diagnosed cancers in men and a leading cause of death worldwide [210, 211]. The androgen receptor (AR) plays a crucial role in prostate cancer growth and progression, and mutations in the AR often lead to tumors becoming resistant to androgen ablation therapy [212]. Post-translational modifications of the AR result in enhanced activity. The AR is acetylated in response to dehydrotestosterone. The AR is acetylated at a conserved lysine residue at K630, K632, and K633, which increases its transcriptional activity [213-217]. It was found that a gain of function mutation increased cell proliferation and colony size in soft agar, and accelerated prostate tumor growth *in vivo* [214]. Histone acetylation was detected in prostate cancer patient tumors and was a positive predictor for tumor recurrence [218]. Deacetylation of ER α and AR by SIRT1 inactivates both the receptors' activity, and thus serves as an alternative approach in the prevention and therapy of breast and prostate cancer [212]. Oncogenic KATs p300 and TIP60 acetylate the AR in a ligand-independent manner and induce transcription of target genes involved in cell proliferation, metastasis, and apoptosis [173, 219]. In resected prostate cancer patients treated with endocrine therapy, p300 and CBP expression is increased. Inhibition of p300 but not CBP using targeted siRNA showed a significant decrease in proliferation rate and

a significant increase in the induction of apoptosis [173]. This might explain why p300 is upregulated in highly metastatic prostate cancer and is an indicator of poor prognosis.

3.2. The Deregulation of Histone Lysine Deacetylases (KDACs) in Cancer

Addition or removal of the acetyl group is catalyzed by acetylases and deacetylases respectively. Aberrant expression due to mutation or deletions of various KDACs is commonly encountered in chronic inflammation associated diseases, in particular cancer. Thus, KDACs are an attractive target for cancer therapy. Numerous studies have indicated global histone acetylation is deregulated in human cancers, and absence of the acetylation mark in H4K16 and H4K20 is critical in establishing the tumor phenotype [220]. The KDACs family of enzymes is divided into four classes and 18 members, based on their sequence homology to yeast KDAC [221, 222]. KDACs deacetylate the acetyl moiety on lysines and thereby can interact with numerous transcription factors to regulate a large number of genes [221]. KDACs are transcriptional repressors because they condense the chromosomes and make them less accessible to transcription initiation [223]. Class 1 KDAC1, 2, 3, and 8 are large protein complexes that silence genes by removing the acetyl moiety on lysine residues on histone tails. High expression of KDAC1, 2, and 3 is seen in urothelial bladder cancer [224]; it is also associated with shorter survival of colorectal cancer patients, and is highly expressed in prostate cancer. Interestingly, KDAC2 is the single prognostic factor that is associated with shorter prostate-specific antigen relapse time following radical prostatectomy [225]. In pancreatic cancers, this class of KDAC is associated with dedifferentiation and rapid proliferation of cells [226-228]. KDAC1 was found to be overexpressed in several organ specific cancers such as gastric, HCC, lung, prostate, breast, and pancreatic cancer samples derived from patients and is a predictor of poor prognosis [228, 229]. Other studies have also reported the overexpression of KDAC1, 2, and 3 in renal cell carcinoma, gastric cancer, and Hodgkin's lymphoma [228, 230, 231]. KDAC8 overexpression is associated with high grade tumor and poor overall survival in childhood neuroblastoma [232]. In a different study analyzing breast tumors, KDAC1 and KDAC3 expression correlates with ER and PR expression, pointing to KDAC1 as an independent prognostic marker [233]. In a tissue microarray representing 44 categories of malignant and borderline mesenchymal tumors, KDAC2 was highly expressed compared to KDAC1, indicating that KDAC2 more likely contributes to the pathogenesis of the disease [234]. KDAC2 is also aberrantly expressed in lung cancer tissues [235]. Class II KDACs are sub-divided into two groups: class IIa consists of KDAC4, KDAC5, KDAC7, and KDAC9; and class IIb consists of KDAC6 and KDAC10. Numerous studies both *in vitro* and *in vivo* have established an association between class IIa KDAC expression and tumor development. KDAC4 is overexpressed in Walsdenstrom's macroglobulinemia [236], breast, renal, and colorectal cancer [237]. In addition, KDAC4 dysfunction is also implicated in cancer development. For example, KDAC4 homozygous deletion was observed in melanoma cells [238], and KDAC4 mutations have also been observed in breast cancer [239]. In renal cell carcinoma, KDAC4 is associated with repression of *Pax7* and regulates HIF1 α transcriptional activity. KDAC5 and KDAC9 are overexpressed and are poor prognostic indicators in medulloblastoma patients, demonstrating an intricate relationship in their expression and poor survival [240]. KDAC5 and KDAC3 overexpression is correlated with copy number gains in hepatocellular carcinoma [241]. In colorectal cancer, KDAC5 and KDAC7 are overexpressed

and interact with the transcription factor GATA-1 [226, 227]. In pancreatic cancer, KDAC7 is overexpressed [242, 243]. In ALL, high levels of KDAC7 and KDAC9 are associated with poor prognosis [244]. In contrast, KDAC7 is not overexpressed in breast, renal, and bladder cancer, and myeloproliferative neoplasms [245]. KDAC9 has been reported to be overexpressed in cervical cancer [246], whereas in glioblastoma, KDAC9 is downregulated [247]. Class IIb KDAC6 is overexpressed in breast and oral squamous cell carcinoma [228]. Deacetylation is also catalyzed by another class of enzymes called sirtuins (Class III KDACs). Sirtuins are NAD⁺ dependent enzymes [248-250]. Sirtuins consist of seven members (Sir1-7) homologous to the yeast KDAC silent information regulator 2 (Sir2) that was cloned and characterized in 1984 as a gene required for maintaining silent chromatin in yeast [251].

Table 1. List of synthetic inhibitors of histone lysine deacetylase either approved or undergoing clinical trials

Compound	KDAC targets	Current status	References
Compound 6j and SEN196	SIRT2	Phase II	[439]
Panobinostat (LBH-589)	KDAC1, 2, 3 and 6	FDA approved	[440]
Vorinostat (SAHA)	KDAC1, 2, 3 and 6	FDA approved	[439]
Romidepsin (FK228)	KDAC1, 2, 3 and 8	FDA approved	[439]
Valproic acid	Class I and Class IIa	FDA approved	[438]
Belinostat (PXD101)	KDAC1, 2, and 3	Phase II	[438]
Entinostat (MS275)	KDAC1, 2, 3 and 8	Phase II	[438]
Mocetinostat (MGCD01030)	KDAC1, 2 and 3	Phase II	[438]
Givinostat (ITF2357)	KDACs	Phase II	[438]
Practinostat (SB939)	KDACs	Phase II	[438]
Chidamide (CS055/HBI-8000)	KDACs	Phase II	[438]
Quisinostat (JNJ26481585)	KDACs	Phase II	[438]
CHR-5154	Macrophage targeted KDACs	Phase I	[441]
CUDC-907	KDACs	Phase I	[441]
Abexinostat (PCI-24781)	KDAC1, 2, 3, 6 and 10	Phase II	[438]
Trichostatin A	KDAC1, 2, 3 and 6	Preclinical	[442]

Sirtuins are regulators of various cellular and pathological processes such as cell proliferation and differentiation, cell survival, stress response and DNA damage, genome stability, metabolism, energy homeostasis, organ development, aging, and cancer. Sirtuins are overexpressed in variety of cancers. SIRT1 is upregulated in prostate, acute myeloid leukemia, and colorectal cancer. SIRT3 and SIRT7 are overexpressed in breast cancer whereas SIRT2 is downregulated in gastric cancer and glioma [252, 253]. KDAC11 belongs to the Class 4 KDACs, and it is overexpressed in colon, prostate, breast, and ovarian cancer [254]. In Hodgkin lymphoma, KDAC11 regulates OX40, a mediator of anti-tumor immunity [228]. However, inhibition of KDACs activity by sodium butyrate or trichostatin A induces apoptosis by upregulating pro-apoptotic protein Bad in human glioma T98G, U251MG, and U877MG cells [255]). There are several FDA approved KDACs inhibitors that are used to treat multiple myeloma, hematological malignancies and solid tumors (Table 1).

4. HISTONE ACETYLATION “CODE” AS A DIAGNOSTIC MARKER FOR CANCER PROGRESSION

Heritable epigenetic marks on histones, such as acetylation, methylation, ubiquitination, and phosphorylation, are maintained with high fidelity during cell division. This phenomenon of histone modifications that act in a dependable manner is called the “histone code” [116, 256]. These post-translational modifications present on the histone tail “codes” are decoded by various cellular machinery regulating cellular processes [256]. Histone post-translational modifications have been detected in almost all cancers types. Loss of H4K16 acetylation and H4K20Me3 is one of the hallmarks of cancers [220]. H3K4Me, H3K9Me2, H3K9Me3, and H3Ac were all reduced in prostate cancer compared to normal prostate tissue [150]. Furthermore, a decrease in H3K4Me2, H3K9Me2, and H3K18Ac levels are all biomarkers of lung cancer, kidney cancer, and pancreatic adenocarcinoma [257, 258]. In lung cancer, aberrant hyperacetylation of H4K5 and H4K8 was observed compared to normal lung epithelium, where hypoacetylation was observed [259]. In addition, H4K20Me3 was shown to be a biomarker for early detection of squamous cell carcinoma [259]. Barlesi et al., 2007 showed that low levels of H2AK5Ac indicated poor prognosis in lung cancer patients [260]. Similarly, adenocarcinoma patients with reduced H3K9Ac exhibited better survival [260]. DNA mismatch repair (MMR) is observed in several cancers and is characterized by frequent alterations in simple repetitive DNA sequences, which is a phenomenon called microsatellite instability and plays a role in apoptosis. In a recent review by Guo Min Li et al., 2013, it was reported that the H3K6Me3 histone biomarker plays an important role in the regulation of MMR *in vivo* in hereditary nonpolyposis colorectal cancer or Lynch syndrome [261]. Taken together, these studies show that histone post-translational modifications can serve as biomarkers for the early diagnosis of cancer and as prognostic predictors.

5. EPIGENETIC MODIFICATIONS IN ONCOGENES AND TUMOR SUPPRESSOR GENES

Oncogenes or proto-oncogenes transcribe oncogenic proteins that alter normal cell proliferation to an abnormal and uncontrolled proliferation of transformed cells [262]. Oncogenes were discovered three decades ago and have, over the years, provided valuable information in understanding the genetic landscape of cancers, differentiation, and apoptosis [262]. One of the most commonly encountered oncogenes in cancer is the *Myc* oncogene. *Myc* regulates a complex inflammatory milieu [263] and, upon activation in B cells, it rapidly induces synthesis and release of IL1 β [264]. The pleiotropic effects of oncogenes include creating a pro-tumor microenvironment, and a persistent, constitutive activation of pro-inflammatory transcription factors [262, 265, 266]. To date approximately 20 transforming oncogenes including *Ras*, *Raf*, *Myc*, *c-Src*, *EGFR*, and *HER-2*, and a number of tumor suppressor genes such as *p53*, *VHL*, and *PTEN* are directly implicated in angiogenesis [267-269]. For example, vascular endothelial growth factor production is upregulated in cancer cells expressing a mutant *Ras* oncogene [268]. Global hypomethylation of genomic DNA is one of the main contributors in the development of aggressive cancers by upregulating oncogenes such as *c-Myc* or *h-Ras* [20]. In addition, promoter hypomethylation can also cause activation of

aberrant expression of oncogenes and induce loss of imprinting in some loci. The tumor suppressor gene, *MASPIN* (also known as *SERPINB5*) becomes hypermethylated in breast and prostate cancer cells [270], whereas *MASPIN* is hypomethylated in other tumor types. *MASPIN* hypomethylation and its subsequent expression increases with the degree of dedifferentiation of some types of cancer cells [271, 272]. The *Myc* oncogene is deregulated in most cancers, where it exhibits abnormally high levels of expression. In a recent study by Wasylishen AR et al., 2014, they show that *Myc* signaling is regulated by six lysine residues near the highly conserved *Myc* homology box IV. These lysine residues are important for the negative regulation of *Myc*-induced transformation and tumorigenesis, which may expose novel therapeutic strategies to target *Myc* in cancer [273]. This region has been previously associated with *Myc* acetylation [274, 275].

The Ras small GTPase family of proteins regulates cellular responses to external stimuli by post-translational modifications. Disruptions to these signals through mutations of *Ras* are involved in 90% of pancreatic cancer and 50% of colon cancer. *Ras* has been shown to be acetylated at lysine 104 and is a negative regulatory modification of *Ras* oncogenicity [276]. In hematological malignancies such as AML and ALL, low frequency mutation in the KAT genes leads to poor prognosis in patients [277]. Chromosomal translocations during mitosis and meiosis generate fusion proteins and chimeric proteins with oncogenic potential. In some rare cases, KAT gene translocations form chimeric proteins that retain KAT catalytic activity and bromodomains. For example, this is seen in mixed lineage leukemia (MLL)-CBP [t(11;16)(q23;p13)] or in MLL-p300 [t(11;22)(q23;q13)] fusions [278]. In addition, monocytic leukemia zinc finger protein (MOZ) fusions such as the MOS-CBP [t(8;16)(p11;p13)] or MOZ-p300 [t(8;22)(p11;q13)] are also reported for AML [174, 279, 280]. These fusions are frequently encountered in patients who have been treated with topoisomerase II inhibitors for the treatment of other cancers, and who can develop secondary therapy-related leukemia [280]. Other oncogenic fusions, in particular the most frequent in AMLs, AML1-ETO [t(8;21)(q22;q22)], require p300 mediated acetylation to induce leukemogenesis [281]. This clearly indicates the importance of transcriptional co-activators p300/CBP and their oncogenic role in cancer progression. In another oncogenic fusion between nuclear receptor co-activator TIF2 and MOZ is associated with AML [inv (8)(p11q13)] and recruits CBP to function as an oncoprotein [180, 282, 283]. Indeed, altered expression of the KATs is found in various cancers [237]. Overexpression of p300 is observed in HCC [284], breast cancer [285], prostate cancer [219], and non-small cell lung cancer [286]. In human skin cancer, upregulation of the enzyme ornithine decarboxylase leads to histone H4 acetylation and skin tumor progression [287]. All these studies suggest KATs as being an integral part of carcinogenesis. p300/CBP acetylates lysine residues K372, K373, and K382 at the carboxy terminal of p53. In addition, DNA damage induces acetylation of K320 and K373, which regulates interactions at DNA binding sites of pro-apoptotic genes [288].

Tumor suppressor *p53* is the most extensively studied gene that is mutated in almost all types of cancers [289]. Under normal physiological conditions, *p53* induces cell cycle arrest, induces apoptosis, and upsets other cellular processes such as senescence and differentiation [290]. The frequent post-translation modification observed in *p53* is acetylation by p300 on 5 lysine residues at the C-terminal regulatory region [291]. Acetylation of *p53* is triggered by ultraviolet light and ionizing radiation [292]). In addition, the *Ras* oncogene and other stress factors also induce acetylation of *p53* [293]. PCAF acetylates K320, while TIP60 (HIV Tat-interacting protein; 60 kDa) and MOF (males absent on the first)-two members of the MYST

family-acetylated K120 are located within the DNA binding domain of *p53* [294, 295]. K120 acetylation induces apoptotic target genes but with minimal or no effect on cell cycle genes. Acetylation of *p53* neutralizes the positive charge on lysine and impairs its ability to form hydrogen bonds. Acetylation also regulates both loss of function and gain of function such as enhancing DNA binding [296] and preventing nonspecific DNA binding by the *p53* C-terminal domain [297]. In addition, acetylation forms docking sites for transcription co-activators, where K382 increases *p53* affinity to CBP bromodomain [298, 299] whereas deacetylation of K373 and K382 enables interaction with the tandem bromodomains of TAF1, a TFIID subunit [299]. Acetylation of *p53* blocks ubiquitination and K320 acetylation but PCAF prevents ubiquitination by E4F1 [300]. Remarkably, both acetylation and ubiquitination activate *p53*. Ubiquitination of K370, K372, K373, K381, and K382 indicates *p53* nuclear export and proteasomal degradation [296]. Thereby, acetylation of these residues stabilizes *p53* and promotes its nuclear localization. Interestingly, in human cancer, p300 and PCAF acetylation sites are not mutated, except K120 which is mutated in human cancer and is linked with cell cycle control and apoptosis [290, 294, 295, 301]. PC4 is a transcriptional coactivator that is acetylated by p300. PC4 interacts with *p53*, and both proteins become acetylated, which enhance their DNA binding ability and leads to the expression of *p53* responsive genes, especially during DNA damage [302, 303].

p53 is also methylated at lysine residues [304]. Lysine mono-methylation of *p53* at K372 by protein lysine methyl transferase (PKMT) and NSD1 regulates its stability and transcriptional activity [305]. This methylation pattern is restricted to the nucleus and is not seen in any of the cytosolic fractions [305]. Subsequently, Huang et al., 2006, showed that PKMT SMYD2 methylates *p53* at K370 and impairs the expression of *CDKN1A* [306]. Inhibiting SMYD2 by small interfering RNA activates *p53*-mediated apoptosis of cancer cells [306]. Interestingly, mono- or di-methylation at K370 of *p53* has contrasting functions. Mono-methylation results in repression of *p53* activity, whereas di-methylation increases the association of *p53* with the co-activator tandem tudor domain containing the *p53*-binding protein1 [126]. LSD1 preferentially demethylates di-methylated *p53* at K370 [126]. The complex nature of methylation or demethylation by SYMD2 or LSD1 respectively can promote oncogenesis, as evidenced by the overexpression of both the enzymes [126]. In another study by Shi et al., 2007, they show that PKMT SETD8 monomethylates *p53* at K382 and suppresses *p53* transcriptional activity in cancer cells [307], indicating that methylation is also an important regulator of *p53* function.

E2F is a family of transcription factors that modulates important cellular events, including cell cycle progression, apoptosis, and DNA damage response, and also functions as a tumor suppressor gene [153, 308]. RB1 is a cell cycle regulator. Both *E2F* and *RB1* genes are exposed to PTMs in several types of cancer [309]. *RB1* is methylated at K810 by SMYD2, which has been identified as being pivotal for cell cycle regulation [310-312]. Methylated *RB1* subsequently enhances phosphorylation of serine 807 and 811 [312]. Unmethylated *RB1* is a suppressor of *E2F* transcriptional activities but, upon RB1 methylation at K810, it transactivates E2F transcriptional activity, and promotes RB1 phosphorylation and cell cycle progression. Furthermore, MYPT1 K442 is a target for methylation. Demethylation of MYPT1 by SETD7 and LSD1 enhances polyubiquitination and subsequent degradation [313], and thereby releases E2F, which transcribes genes that are involved in S-phase to promote cell cycle progression [313]. Taken together, methylation and demethylation regulates cell cycle progression by modulation RB1 activity. Deregulated *E2F* has been detected in a variety of

human cancers [308]. In p53 deficient cells, methylation of K185 or E2F1 by SETD7 induces polyubiquitination and degradation of E2F1, and thus prevents E2F1 accumulation upon DNA damage [314]. In addition, LSD1 also demethylates *E2F1* at K185 and maintains an adequate amount of unmethylated *E2F1* in cancer cells [314].

6. EPIGENETIC MODIFICATIONS IN CELLULAR SIGNALING PATHWAYS

6.1. NF- κ B Signaling Pathway

The master transcription factor NF- κ B was first discovered by David Baltimore in 1986. It is present in the cytoplasm in a heterodimeric form bound to its inhibitor, and upon phosphorylation of its inhibitor, NF- κ B translocates to the nucleus of the cell and binds to the promoter region in the immunoglobulin kappa chain in B cells [315]. NF- κ B regulates multiple biological functions, including inflammation, immunity, cell proliferation, and apoptosis [316]. NF- κ B is activated by many diverse agents such as interleukins, cytokines, T and B-cell mitogens, gram negative bacterial lipopolysaccharide, viruses and viral proteins, double-stranded DNA, and chemical and physical stress [317-319]. NF- κ B regulates more than 400 genes involved in inflammation, immunoregulation, tumor cell proliferation, invasion, metastasis and angiogenesis, chemoresistance and radioresistance [320-322]. Activation of an inflammatory response is mainly mediated through NF- κ B, which is susceptible to post-translational modifications such as acetylation, lysine methylation, and arginine methylation, in addition to its primary phosphorylation modification [323]. Acetylation modification has been shown to be a major player in inflammatory gene expression [324] in T-cells and monocytes [325-327]. Acetylation occurs in the lysine residue of RelA, which regulates NF- κ B transcription activation, DNA binding affinity, I κ B- α assembly, and subcellular localization [328, 329]. Inhibitory kappa kinase (IKK α), in association with polymerase II and CREB binding protein (CBP), binds to the NF- κ B dependent promoter, where it acetylates H3K9 [330] and also phosphorylates H3Ser10 [331]. This IKK α -dependent and cytokine-mediated phosphorylation is essential for CBP-mediated acetylation of H3K14 [332]. One of the common features in cytokine-mediated inflammation and in autocrine and paracrine activation is that it leads to increased NF- κ B driven inflammatory gene expression [333]. However, glucocorticoids and KDAC2 can reverse NF- κ B activation and thus the inflammatory reaction [334]. Poly (ADP-ribose) polymerase -1 (PARP-1), in association with NF- κ B, contributes to the progression of chronic inflammation driven diseases [335]. PARP-1 is a promoter-specific co-activator of NF- κ B *in vivo* and is independent of its enzymatic activity [336]. PARP-1 directly interacts with p300, p50, and p65, and synergistically activates NF- κ B [337]. Under inflammatory conditions, p300 can acetylate PARP-1 at specific lysine residues in a variety of cell types. NF- κ B mediated H3 and H4 acetylation induced by IL1 β , TNF α , or endotoxins can induce the production of granulocyte colony stimulation factor (GM-CSF) [338]. Thus, it is obvious that reversible acetylation of NF- κ B, PARP1, and histones plays a prominent role in regulating gene expression during inflammation. Another important post-translational modification is methylation. An altered lysine methylation pattern is often

observed during inflammatory diseases. The H3K4 methyltransferase SET7/9 can also regulate NF- κ B(p65) to its promoters, thereby activating inflammatory gene expression [339]. In addition, arginine methyltransferases PRMT1 and PRMT4 are transcriptional co-activators of NF- κ B, although they have not been directly implicated in inflammation [340-342]. Recent studies have implicated NF- κ B hyperactivity upon deletion of H3K9 methyl marks from its promoter region, especially in hyperglycemic memory [343, 344], suggesting a strong link between epigenetic modification and NF- κ B in the development of chronic inflammation-driven disease [345, 346]. Constitutive activation of transcription factors is observed in all types of cancers. In particular, acetylation of K9, K14, K18, and K56 of histone H3, and acetylation of K5, K8, K13, and K16 of histone H4 are strongly involved in transcriptional activation [15, 178]. The master transcription factor NF- κ B is another non-histone protein and a key regulator of the cellular immune response that is also regulated by reversible acetylation at K218, K221, and K310. Acetylation of NF- κ B regulates different functions such as DNA binding affinity, I κ B α assembly and subcellular localization, and in the recruitment of NF- κ B co-activators such as BRD4 and PTEFB [347], which stimulates the transcription of NF- κ B target genes. In turn, acetylated NF- κ B is deacetylated by KDAC3 [348]. Acetylated NF- κ B and its subsequent recruitment of bromodomain-containing factor Brd4, which is a member of the bromodomain extraterminal (BET) family, leads to sustained activation of NF- κ B in adult T-cell leukemia [349]. Acetylation of p65 at K122 and K123 reduces its affinity for DNA and increases I κ B interaction, while suppressing downstream gene expression [350]. In addition, acetylation of p50 at K431, K440, and K441 enhances transcriptional activation [351]. In breast cancer, the ER along with NF- κ B plays an important role in cell proliferation and survival. Pradhan et al., 2012 show that there is a major increase in NF- κ B dependent histone acetylation around the estrogen response element in the BIRC3 promoter region in aggressive luminal B breast tumors [352], which is in contrast to earlier reports that suggested repressive interactions between ER and NF- κ B [352].

The importance of protein lysine methylation in the pathogenesis of cancer is broadly divided into 5 categories. First, lysine methylation affects other post-translational modifications directly or indirectly. Second, lysine methylation regulates protein-protein interaction. Third, lysine methylation inhibits polyubiquitination of lysine residues. Fourth, lysine methylation regulates subcellular localization of proteins. Fifth, lysine methylation regulates promoter binding affinity of transcription factors, thereby controlling the expression levels of target genes [353, 354]. Lysine di-methylation of RelA by PKMT and NSD1 increases its promoter binding ability and transcriptional activity [88, 135, 355]. Upon TNF α stimulation, SETD7 methylates K37 RelA in the nucleus and increases the promoter binding affinity of RelA. Lu et al., 2013 reported that NSD1 methylates RelA at two sites, K218me1 and K221me2, to increase NF- κ B transcription of target genes, while any mutations on these two sites diminish its ability to initiate transcription [355]. In addition, SETD6 methylates RelA at K310 in primary immune cells and completely abolishes the NF- κ B mediated inflammatory response [356]. RelA methylation is demethylated by PKDM F-box and leucine-rich repeat protein 11 (FBXL11). Overexpression of FBXL11 inhibits NF- κ B activation and impedes the growth of HT29 colon cancer cells [135]. Yang et al., 2012 found that, in castration-resistance prostate cancer, NSD2 interacts with NF- κ B and acetyltransferase p300 and di or tri-methylates H3K36 at the

promoter regions of NF- κ B target genes [357]. Overall, transcriptional activation or inhibition of NF- κ B is intricately linked to the site of methylation on RelA.

6.2. STAT3 Signaling Cascade

Signal transducer and activator of transcription (STAT) is another important transcription factor that is constitutively activated in cancer cells. The primary route of signaling in the STAT signaling pathway is via cytokine ligand-induced receptor dimerization, which activates receptor-associated tyrosine kinases and in-turn phosphorylates STATs, induces dimerization, nuclear translocation, DNA binding, and cytokine responsive genes [358]. Acetylation of STAT3 is a cytokine-induced, post-translational modification. STAT1, STAT2, STAT3, STAT5, and STAT6 have been shown to be acetylated by CBP/p300 [359]. Of all the STAT family of transcription factors, the most prominent is STAT3. STAT3 gets acetylated by histone acetyl transferase CBP/p300 at K685. Acetylation enhances its DNA binding affinity, increases transcriptional activation, promotes protein–protein interaction, and modulates dimerization [360]. Deacetylation by KDAC3, and to a lesser extent by KDAC1 and KDAC2, can lead to loss of DNA binding and suppression of transcription [360]. Thus, the acetylation and deacetylation reaction is a novel signaling mechanism that regulates the IL6/STAT3 signaling pathway in hepatocellular carcinoma. In addition to K685 acetylation, STAT3 also gets acetylated by p300 on aa49 and aa87 at its NH2 terminal domain, but the exact role of this in transcriptional initiation is still unclear [361]. When lysine residues were substituted by arginine at K679, K685, and K707, it reduced STAT3 phosphorylation compared to a K685R mutant alone, signifying the importance of the regulatory role of multiple acetylations [362].

In another report by Ohbayashi et al., 2007, they showed that IL-6 or leukemia inhibitory factor induced acetylation of STAT3 at K685 in 293T and Hep3B cells. This effect was abolished by the PI3K inhibitor, LY294002 [363]. STAT1 also gets acetylated and binds to the NF- κ B subunit p65, and negatively regulates NF- κ B target genes in cancer cells [364]. In the majority of published studies, cytokines have been shown to be the major inducers of acetylation; however, KDAC inhibitors have also been shown to hyperacetylate STAT3. In diffuse, large B-cell lymphoma (DLBCL), the KDAC inhibitor LBH589 leads to STAT3 hyperacetylation and dephosphorylation, and thereby inhibiting STAT3 transcription in DLBCL cells [365]. Acetylation of the transcription factor E2F1, which regulates cell cycle progression, at the N-terminal promotes DNA binding activity [366, 367] while acetylation attenuates Foxo1 and p65 transcription factors' ability to bind to DNA [350, 368]. The innate immune response activated by toll like receptors (TLRs) upon exposure to lipopolysaccharide often leads to chronic inflammation. Stimulation of TLRs induces expression of mitogen activated protein kinase (MAPK) phosphatase-1 (MKP1), which is acetylated at a key lysine residue K57, and promotes the dephosphorylation of p38MAPK and c-jun N-terminal kinase (JNK), resulting in the attenuated production of pro-inflammatory cytokines [369]. STAT3 is a latent transcription factor that is constitutively activated in numerous cancer cells [370]. STAT3 is di-methylated by SETD7 at K140 following phosphorylation, dimerization, nuclear translocation, and promoter binding; however, this reversible methylation is demethylated by LSD1 [371]. This transient dimethylation has a minimal repressive effect and contributes to specific transcriptional regulation. Furthermore, STAT3 is trimethylated at K180 by EZH2,

which increases the phosphorylation of STAT3 in glioblastoma cells [372]. Interestingly, inhibiting EZH2 directly inhibits Polycomb gene expression and indirectly blocks STAT3 activation, which serves as a strategic target for glioblastoma therapy.

6.3. Wnt/ β -Catenin Signaling Pathway

The Wnt signaling pathway is evolutionarily conserved and plays important roles in the development of organ systems. The important mediators in this pathway are adenomatous polyposis coli (APC), glycogen synthase kinase 3 β (GSK3 β), axin, and the transcriptional cofactor β -catenin (encoded by *CTNNB1*) [373]. Under normal circumstances, β -catenin levels remain low due to constitutive phosphorylation of β -catenin by GSK3 β and subsequent polyubiquitination and degradation of β -catenin [374]. Wnt ligand binding to the frizzled (Fz) receptor activates the Disheveled (Dvl) adaptor protein, which in turn inhibits GSK3 β activity and thereby inhibits β -catenin degradation. β -catenin translocates to the nucleus where it binds to the T-cell factor (Tcf)/lymphoid-enhancing factor (Lef) family of transcription factors and induces the expression of genes such as *CMYC* and *CCND1* [375]. A high level of nuclear β -catenin is often observed in a variety of cancers [374]. Aberrant Wnt signaling is a very common phenomenon in colon and hepatocellular carcinoma [376]. Epigenetic silencing of genes involved in Wnt signaling (by DNA methylation), and histone modification in tumor suppressor genes is a major contributor in aberrant Wnt/ β -catenin signaling in cancers [377]. Promoter hypermethylation of Wnt inhibitors such as secreted frizzled-related protein (SFRP), Wnt inhibitory factor 1 (WIF1), HDPR1, and Dickkopf 3 (DKK3) is seen in a variety of tumors [378-380]. Schiefer et al., 2014 recently demonstrated that hypermethylation of the SFRP family was a common feature of human glioblastoma multiforme [381].

Overexpression of Wnt1, 2, 3A, and 5A leads to oncogenic activation of canonical Wnt signaling in cancers of the breast, colon, lung, and prostate [382, 383]. Wnt5A, a tumor suppressor, is silenced by methylation in hematological malignancies. Inactivation of Wnt7A and Wnt9A by promoter hypermethylation has been reported in ALL, pancreatic, and colon cancer [384, 385]. Promoter hypermethylation of SFRP1 and WIF1 is a poor prognostic indicator in breast cancer and acute promyelocytic leukemia respectively [386-388]. In gastric cancer, DKK3 hypermethylation is strongly correlated with a shorter survival time for patients [389]. Epigenetic silencing of *CDH1* and *APC* by promoter methylation leads to aberrant activation of Wnt/ β -catenin signaling in invasive ductal carcinoma of the breast [390]. NSD2 methyltransferase may activate the Wnt signaling pathway through interaction with β -catenin [391]. NSD2 was shown to interact with β -catenin and its co-activators IQ motif containing GTPase activating protein 1 (IQGAP1) and T-cell lymphoma invasion and metastasis 1 (TIAM1) proteins and increase motility, adhesion, and migration properties [88]. Inhibiting NSD2 concomitantly decreases the expression of cyclin D1 (*CCND1*), a target of the β -catenin/Tcf4 complex, through H3K39 trimethylation [88]. Interestingly, Ge et al., 2009 reported that PCAF acetylation stabilizes β -catenin in colon cancer cells [392].

7. EPIGENETIC MODIFICATION IN CANCER STEM CELLS

Cancer stem cells (CSCs) represent a small sub-population in the whole tumor population that represents the tumor initiating clone. They are very similar to normal stem cells in that they have unlimited self-renewal capacities with the potential to differentiate into many cancer cell types. CSCs are resistant to chemotherapy and may be responsible for the distant site metastasis of cancers [393, 394]. EZH2 catalyzes methylation of histone H3K27me₃ to cause the silencing of tumor suppressor genes [395, 396]. Extensive reports have shown that EZH2 plays a critical role in cancer initiation, progression, and metastasis, and also in cancer stem cell biology [102, 397]. EZH2 plays a pivotal role in the maintenance and self-renewal of embryonic and adult stem cells [398–400]. EZH2 has been implicated in the maintenance of CSC, promotion of breast cancer progression [401], and glioblastoma CSC self-renewal and progression [402, 403]. In hematological malignancies, DNA hypermethylation by DNMTs of oncogenic pathway regulatory genes may contribute to the differentiation of hematopoietic stem cells to leukemia stem cells (LSCs) and leukemogenesis [51, 52, 404–406].

Activating mutations of the DNMT3A gene is a poor prognosis indicator in AML, with leukemogenic mutations of Flt3-ITD, NPM1, PTPN11, and IDH2 genes being associated with intermediate risk [407]. In addition, other mechanisms mediated by KDMs and KATs play key roles in the development and maintenance of LSCs [52]. In AML-CSCs and non-CSCs there is a difference in the histone methylation pattern, especially in H3K4me₃ and H3K27me₃, whereas it is the same in DNA methylation patterns [408]. In glioblastoma multiforme (GBM)-CSCs, the loss of polycomb mark H3K27me₃ from the promoters is the primary cause for the activation of multiple transcription factors [409]. In particular, activation of the achaetescute complex homolog-like 1 (Ascl1) transcription factor causes activation of the Wnt signaling pathway, which is required for the maintenance of stemness and its tumorigenicity [410, 411].

8. EPIGENETIC MODIFICATIONS ASSOCIATED WITH CANCER CHEMO- AND RADIO-RESISTANCE

Epigenetic alterations are now increasingly recognized as major factors in tumor biology and are integral for their role in treatment resistance. Epigenetic modulators have made a profound impact in preventing, delaying, or reversing chemo-resistance, as seen with the use of natural product molecules or synthetic molecules that can reverse DNA methylation or histone deacetylation [412, 413]. The emergence of a CSC population is now recognized as one of the causative factors contributing to multi-therapy resistance, which is further compounded by new mutations arising in genes associated with oncogenic signaling pathways [414–416]. For example, *K-Ras* mutation in multiple tumor types, *APC* in colon cancer, and *VHL* in renal cancer empower the cancer cells with self-renewal and tumorigenic properties [25]. CBP/p300 is implicated in resistance in certain cancer types. P300 expression is reduced in doxorubicin-resistant bladder cancer cells and knockdown of p300 leads to bladder cancer cells being resistant to doxorubicin [417].

In addition, CBP/p300 is required in the interleukin-4-mediated castration resistance of prostate cancer cells. Downregulation of CBP/p300 abolished interleukin-4-mediated AR activation [418]. Overall, the changes in genetic alterations, expression level, and subcellular

localization of CBP/ p300 are associated with progression, prognosis, and resistance of many types of cancers. Acquired drug resistance is the most common phenomenon in cancer therapy [419].

Overexpression of the hMLH1 DNA mismatch repair enzyme in ovarian cancer confers resistance to platinum drugs; 25% of patients acquired hMLH1 methylation during chemotherapy, and hMLH1 methylation is poor prognosis indicator [420].

Treatment with the DNMT inhibitor azacitidine reactivates hMLH1, reverses resistance [421, 422], and sensitizes ovarian cancer cells to platinum drugs *in vitro* [423]. The methylation status of the cells serves as a biomarker to monitor progression in patients under treatment with platinum drugs [413, 424, 425]. KAT Tip60 is overexpressed in cisplatin-resistant human lung cancer cells, and knockdown of Tip60 expression sensitizes the cells to cisplatin therapy [426]. ABCB1 (MDR1) is a multidrug resistance p-glycoprotein that is often induced in traditional chemotherapy. However, upon chemotherapy, hypermethylated promoter regions decrease in these cells while, in contrast, epigenetic methylation is implicated in the development of multidrug resistance [427, 428]. These findings suggest that epigenetic methylation independently could contribute to multidrug resistance and not as a part of accumulating genetic mutations in cancer development. In cisplatin resistant ovarian cancer cells, KDAC4 is overexpressed and promotes cancer cell survival by deacetylating the transcription factor STAT1 [429]. The promoter methylation-specific silencing of gene O-6-methylguanine-DNA methyltransferase (*MGMT*) in glioma serves as a biomarker to the alkylating agent carmustine or temozolomide [430-433]. Romidepsin, a KDAC inhibitor, increases sensitivity to EGFR tyrosine kinase inhibitors in lung cancer [434]. In a combination therapy with the KDAC inhibitors entinostat and erlotinib, patients experienced disease progression due to erlotinib resistance that was due to epigenetic changes [435]. In another combination therapy, all-trans-retinoic acid with the LSD1/KDM1A inhibitor tranylcypromine has been shown to upregulate the myeloid differentiation pathway [129, 436, 437].

9. PHARMACOLOGICAL INHIBITORS OF EPIGENETIC MODIFICATIONS

Especially in malignant epithelial and hematological tumors, deregulation of cellular signaling pathways is often due to overexpression of epigenetic modifying enzymes that either methylate or acetylate the DNA or the lysine residues on the histone. KDAC inhibitors are classified into groups based on their chemical structure, including hydroxamic acids (trichostatin A, vorinostat), carboxylic acids (valproate, butyrate), aminobenzamides (entinostat, mocetinostat), cyclic peptides (apicidin, romidepsin), epoxyketones (trapoxins), and hybrid molecules [438]. Table 1 lists the KDAC inhibitors that have been approved by the FDA and the other drugs that are in clinical trials for the treatment of cancers. Table 2 depicts the DNA methyltransferase inhibitors that are in clinical trials. Table 3 indicates the histone methyltransferase and histone demethylase inhibitors currently undergoing clinical trials. The histone acetyl transferase inhibitors currently in use are presented in Table 4.

Table 2. A list of selected pharmacological DNA methyltransferase inhibitors

Compound	Mechanism(s) of action	Current status	Reference
5-Azacytidine (Vidaza)	Nucleoside DNMT inhibitor	FDA approved	[441]
Decitabine (Dacogen)	Nucleoside DNMT inhibitor	FDA approved	
5-fluor-2'-deoxycytidine (FdCyd) with tetrahydrouridine (THU)	FdCyd metabolite inhibits DNMT	In Phase I and Phase II clinical trials	
ASTX-727	Nucleoside DNMT inhibitor	In phase I clinical trials	
CC-486	Nucleoside DNMT inhibitor	In phase I-III clinical trials	
Hydralazine	Non-nucleoside DNMT inhibitor	In Phase I-III clinical trials	
MG-98	Non-nucleoside DNMT inhibitor	In phase I clinical trials	
RRx-001	Non-nucleoside DNMT inhibitor	In Phase I and phase II clinical trials	
SGI-110	Nucleoside DNMT inhibitor	In phase I and Phase II clinical trials	
Zebularine	Nucleoside DNMT inhibitor	Preclinical	
SGI-1027	Non-nucleoside DNMT inhibitor	Preclinical	
RG108	Non-nucleoside DNMT inhibitor	Preclinical	
Psammaplin A	Non-nucleoside DNMT inhibitor	Preclinical	
Procain and procainamide	Non-nucleoside DNMT inhibitor	Preclinical	
CP-4200	Nucleoside DNMT inhibitor	Preclinical	
Epigallocatechin-3-gallate	Nucleoside DNMT inhibitor	Preclinical	

Table 3. A list of selected Histone lysine methyl transferase (KMTi) and histone lysine demethylase inhibitors (KDMi)

Compound	Target	Reference
Chaetocin	KMTi	[442]
DZNep	KMTi	
BIX-01294	KMTi	
AMI-1	KMTi	
AMI-5	KMTi	
Allantodapsone	KMTi	
RM-65	KMTi	
Tranylcypromine	KMTi	
Pargyline	KMTi	
Phenelzine	KMTi	

Table 4. A list of selected Histone lysine acetyltransferase inhibitors (KATi)

Compound	Target(s)	References
Garcinol	PCAF/p300	[442, 443]
Isogarcinol/LTK14	PCAF/p300	[439, 443]
Curcumin	CBP/p300	[439, 443]
Anacardic acid	PCAF/p300	[439, 443]
Lys-CoA	p300	[439]
Lys-CoA-CH ₃	PCAF	[439]
Ph-Lys-CoA	P300	[439]

PERSPECTIVES AND CONCLUSION

Epigenetic changes such as DNA methylation and histone modification by enzyme DNA methylases, histone lysine methyltransferases, demethylases, histone lysine acetyltransferases and deacetylases control critical gene transcription mechanisms. Their deregulation is often seen in diseases and malignancies. Detailed investigations of the changes that regulate disease progression are essential, and epigenetic alterations constitute valuable therapeutic targets. Large scale screening of the epigenome has helped in identifying several new epigenetic targets that have enormous potential in epigenetic-specific therapy. Several KDAC and DNA methyltransferase inhibitors have been approved by the FDA for the treatment of cancers. While natural product agents have shown significant progress in preclinical studies, well defined human clinical trials are essential to validate their potential as a therapeutic modality. Overall, existing evidence(s) clearly suggest that drugs targeting the epigenome have made significant inroads in cancer therapy, especially in the management of hematological malignancies and in other aggressive solid tumors.

ACKNOWLEDGMENT

This work was supported by NUHS Bench-to- Bedside grant to GS.

CONFLICT OF INTEREST

The authors declare no competing financial interests.

REFERENCES

- [1] Hirst M, Marra MA. Epigenetics and human disease. *Int J Biochem Cell Biol.* 2009;41(1):136-46.
- [2] Feinberg AP. Phenotypic plasticity and the epigenetics of human disease. *Nature.* 2007;447(7143):433-40.

- [3] Egger G, Liang G, Aparicio A, Jones PA. Epigenetics in human disease and prospects for epigenetic therapy. *Nature*. 2004;429(6990):457-63.
- [4] Esteller M. Epigenetics in cancer. *N Engl J Med*. 2008;358(11):1148-59.
- [5] Heidland A, Klassen A, Rutkowski P, Bahner U. The contribution of Rudolf Virchow to the concept of inflammation: what is still of importance? *J Nephrol*. 2006;19 Suppl 10:S102-9.
- [6] Balkwill F, Mantovani A. Inflammation and cancer: back to Virchow? *Lancet*. 2001;357(9255):539-45.
- [7] Schottenfeld D, Beebe-Dimmer J. Chronic inflammation: a common and important factor in the pathogenesis of neoplasia. *CA Cancer J Clin*. 2006;56(2):69-83.
- [8] Saigo K, Yoshida K, Ikeda R, Sakamoto Y, Murakami Y, Urashima T, et al. Integration of hepatitis B virus DNA into the myeloid/lymphoid or mixed-lineage leukemia (MLL4) gene and rearrangements of MLL4 in human hepatocellular carcinoma. *Hum Mutat*. 2008;29(5):703-8.
- [9] Aggarwal BB, Shishodia S, Sandur SK, Pandey MK, Sethi G. Inflammation and cancer: how hot is the link? *Biochem Pharmacol*. 2006;72(11):1605-21.
- [10] Surh YJ, Kundu JK, Na HK, Lee JS. Redox-sensitive transcription factors as prime targets for chemoprevention with anti-inflammatory and antioxidative phytochemicals. *J Nutr*. 2005;135(12 Suppl):2993S-3001S.
- [11] Portela A, Esteller M. Epigenetic modifications and human disease. *Nat Biotechnol*. 2010;28(10):1057-68.
- [12] Eliseeva ED, Valkov V, Jung M, Jung MO. Characterization of novel inhibitors of histone acetyltransferases. *Mol Cancer Ther*. 2007;6(9):2391-8.
- [13] Selvi BR, Mohankrishna DV, Ostwal YB, Kundu TK. Small molecule modulators of histone acetylation and methylation: a disease perspective. *Biochim Biophys Acta*. 2010;1799(10-12):810-28.
- [14] Wagner JM, Hackanson B, Lubbert M, Jung M. Histone deacetylase (HDAC) inhibitors in recent clinical trials for cancer therapy. *Clin Epigenetics*. 2010;1(3-4):117-36.
- [15] Berger SL. The complex language of chromatin regulation during transcription. *Nature*. 2007;447(7143):407-12.
- [16] Bird A. Perceptions of epigenetics. *Nature*. 2007;447(7143):396-8.
- [17] Bird A. DNA methylation patterns and epigenetic memory. *Genes Dev*. 2002;16(1):6-21.
- [18] Wilson AG. Epigenetic regulation of gene expression in the inflammatory response and relevance to common diseases. *J Periodontol*. 2008;79(8 Suppl):1514-9.
- [19] Bishop KS, Ferguson LR. The interaction between epigenetics, nutrition and the development of cancer. *Nutrients*. 2015;7(2):922-47.
- [20] Das PM, Singal R. DNA methylation and cancer. *J Clin Oncol*. 2004;22(22):4632-42.
- [21] Otterson GA, Khleif SN, Chen W, Coxon AB, Kaye FJ. CDKN2 gene silencing in lung cancer by DNA hypermethylation and kinetics of p16INK4 protein induction by 5-aza 2'deoxyctidine. *Oncogene*. 1995;11(6):1211-6.
- [22] Katzenellenbogen RA, Baylin SB, Herman JG. Hypermethylation of the DAP-kinase CpG island is a common alteration in B-cell malignancies. *Blood*. 1999;93(12):4347-53.
- [23] Dobrovic A, Simpfendorfer D. Methylation of the BRCA1 gene in sporadic breast cancer. *Cancer Res*. 1997;57(16):3347-50.

- [24] Kang SH, Choi HH, Kim SG, Jong HS, Kim NK, Kim SJ, et al. Transcriptional inactivation of the tissue inhibitor of metalloproteinase-3 gene by dna hypermethylation of the 5'-CpG island in human gastric cancer cell lines. *Int J Cancer*. 2000;86(5):632-5.
- [25] Easwaran H, Tsai HC, Baylin SB. Cancer epigenetics: tumor heterogeneity, plasticity of stem-like states, and drug resistance. *Mol Cell*. 2014;54(5):716-27.
- [26] Li L, Li W. Epithelial-mesenchymal transition in human cancer: comprehensive reprogramming of metabolism, epigenetics, and differentiation. *Pharmacol Ther*. 2015;150:33-46.
- [27] Chen LJ, Wei SY, Chiu JJ. Mechanical regulation of epigenetics in vascular biology and pathobiology. *J Cell Mol Med*. 2013;17(4):437-48.
- [28] Fleming JL, Huang TH, Toland AE. The role of parental and grandparental epigenetic alterations in familial cancer risk. *Cancer Res*. 2008;68(22):9116-21.
- [29] Brigati C, Noonan DM, Albini A, Benelli R. Tumors and inflammatory infiltrates: friends or foes? *Clin Exp Metastasis*. 2002;19(3):247-58.
- [30] Coussens LM, Werb Z. Inflammation and cancer. *Nature*. 2002;420(6917):860-7.
- [31] Mantovani A, Allavena P, Sica A, Balkwill F. Cancer-related inflammation. *Nature*. 2008;454(7203):436-44.
- [32] Rosin MP, Anwar WA, Ward AJ. Inflammation, chromosomal instability, and cancer: the schistosomiasis model. *Cancer Res*. 1994;54(7 Suppl):1929s-33s.
- [33] Vaissiere T, Sawan C, Herceg Z. Epigenetic interplay between histone modifications and DNA methylation in gene silencing. *Mutat Res*. 2008;659(1-2):40-8.
- [34] Barres R, Osler ME, Yan J, Rune A, Fritz T, Caidahl K, et al. Non-CpG methylation of the PGC-1alpha promoter through DNMT3B controls mitochondrial density. *Cell Metab*. 2009;10(3):189-98.
- [35] Herman JG, Baylin SB. Gene silencing in cancer in association with promoter hypermethylation. *N Engl J Med*. 2003;349(21):2042-54.
- [36] Weber M, Schubeler D. Genomic patterns of DNA methylation: targets and function of an epigenetic mark. *Curr Opin Cell Biol*. 2007;19(3):273-80.
- [37] Okano M, Xie S, Li E. Cloning and characterization of a family of novel mammalian DNA (cytosine-5) methyltransferases. *Nat Genet*. 1998;19(3):219-20.
- [38] Bernstein BE, Meissner A, Lander ES. The mammalian epigenome. *Cell*. 2007;128(4):669-81.
- [39] Chen QW, Zhu XY, Li YY, Meng ZQ. Epigenetic regulation and cancer (review). *Oncol Rep*. 2014;31(2):523-32.
- [40] Jeltsch A. Molecular enzymology of mammalian DNA methyltransferases. *Curr Top Microbiol Immunol*. 2006;301:203-25.
- [41] McCabe MT, Brandes JC, Vertino PM. Cancer DNA methylation: molecular mechanisms and clinical implications. *Clin Cancer Res*. 2009;15(12):3927-37.
- [42] Galm O, Herman JG, Baylin SB. The fundamental role of epigenetics in hematopoietic malignancies. *Blood Rev*. 2006;20(1):1-13.
- [43] Weisenberger DJ, Velicescu M, Preciado-Lopez MA, Gonzales FA, Tsai YC, Liang G, et al. Identification and characterization of alternatively spliced variants of DNA methyltransferase 3a in mammalian cells. *Gene*. 2002;298(1):91-9.
- [44] Schoofs T, Muller-Tidow C. DNA methylation as a pathogenic event and as a therapeutic target in AML. *Cancer Treat Rev*. 2011;37 Suppl 1:S13-8.

- [45] Hansen KD, Timp W, Bravo HC, Sabunciyan S, Langmead B, McDonald OG, et al. Increased methylation variation in epigenetic domains across cancer types. *Nat Genet.* 2011;43(8):768-75.
- [46] Schuebel KE, Chen W, Cope L, Glockner SC, Suzuki H, Yi JM, et al. Comparing the DNA hypermethylome with gene mutations in human colorectal cancer. *PLoS Genet.* 2007;3(9):1709-23.
- [47] Yuan BZ, Durkin ME, Popescu NC. Promoter hypermethylation of DLC-1, a candidate tumor suppressor gene, in several common human cancers. *Cancer Genet Cytogenet.* 2003;140(2):113-7.
- [48] Fang JY, Lu R, Mikovits JA, Cheng ZH, Zhu HY, Chen YX. Regulation of hMSH2 and hMLH1 expression in the human colon cancer cell line SW1116 by DNA methyltransferase 1. *Cancer Lett.* 2006;233(1):124-30.
- [49] Goel A, Arnold CN, Niedzwiecki D, Carethers JM, Dowell JM, Wasserman L, et al. Frequent inactivation of PTEN by promoter hypermethylation in microsatellite instability-high sporadic colorectal cancers. *Cancer Res.* 2004;64(9):3014-21.
- [50] Yang L, Rau R, Goodell MA. DNMT3A in haematological malignancies. *Nat Rev Cancer.* 2015;15(3):152-65.
- [51] Trowbridge JJ, Sinha AU, Zhu N, Li M, Armstrong SA, Orkin SH. Haploinsufficiency of Dnmt1 impairs leukemia stem cell function through derepression of bivalent chromatin domains. *Genes Dev.* 2012;26(4):344-9.
- [52] Vecchio L, Seke Etet PF, Kipanyula MJ, Krampera M, Nwabo Kamdje AH. Importance of epigenetic changes in cancer etiology, pathogenesis, clinical profiling, and treatment: what can be learned from hematologic malignancies? *Biochim Biophys Acta.* 2013;1836(1):90-104.
- [53] Ngollo M, Dagdemir A, Karsli-Ceppioglu S, Judes G, Pajon A, Penault-Llorca F, et al. Epigenetic modifications in prostate cancer. *Epigenomics.* 2014;6(4):415-26.
- [54] Reibenwein J, Pils D, Horak P, Tomicek B, Goldner G, Worel N, et al. Promoter hypermethylation of GSTP1, AR, and 14-3-3sigma in serum of prostate cancer patients and its clinical relevance. *Prostate.* 2007;67(4):427-32.
- [55] Woodson K, Hayes R, Wideroff L, Villaruz L, Tangrea J. Hypermethylation of GSTP1, CD44, and E-cadherin genes in prostate cancer among US Blacks and Whites. *Prostate.* 2003;55(3):199-205.
- [56] Phe V, Cussenot O, Roupret M. Interest of methylated genes as biomarkers in urothelial cell carcinomas of the urinary tract. *BJU Int.* 2009;104(7):896-901.
- [57] Pulukuri SM, Patibandla S, Patel J, Estes N, Rao JS. Epigenetic inactivation of the tissue inhibitor of metalloproteinase-2 (TIMP-2) gene in human prostate tumors. *Oncogene.* 2007;26(36):5229-37.
- [58] Henrique R, Costa VL, Cerveira N, Carvalho AL, Hoque MO, Ribeiro FR, et al. Hypermethylation of Cyclin D2 is associated with loss of mRNA expression and tumor development in prostate cancer. *J Mol Med (Berl).* 2006;84(11):911-8.
- [59] Parrella P, Poeta ML, Gallo AP, Prencipe M, Scintu M, Apicella A, et al. Nonrandom distribution of aberrant promoter methylation of cancer-related genes in sporadic breast tumors. *Clin Cancer Res.* 2004;10(16):5349-54.
- [60] Swift-Scanlan T, Vang R, Blackford A, Fackler MJ, Sukumar S. Methylated genes in breast cancer: associations with clinical and histopathological features in a familial breast cancer cohort. *Cancer Biol Ther.* 2011;11(10):853-65.

- [61] Fackler MJ, McVeigh M, Evron E, Garrett E, Mehrotra J, Polyak K, et al. DNA methylation of RASSF1A, HIN-1, RAR-beta, Cyclin D2 and Twist in in situ and invasive lobular breast carcinoma. *Int J Cancer*. 2003;107(6):970-5.
- [62] Li S, Rong M, Iacopetta B. DNA hypermethylation in breast cancer and its association with clinicopathological features. *Cancer Lett*. 2006;237(2):272-80.
- [63] Yan L, Yang X, Davidson NE. Role of DNA methylation and histone acetylation in steroid receptor expression in breast cancer. *J Mammary Gland Biol Neoplasia*. 2001;6(2):183-92.
- [64] Dietrich D, Lesche R, Tetzner R, Krispin M, Dietrich J, Haedicke W, et al. Analysis of DNA methylation of multiple genes in microdissected cells from formalin-fixed and paraffin-embedded tissues. *J Histochem Cytochem*. 2009;57(5):477-89.
- [65] Yballe CM, Vu TH, Hoffman AR. Imprinting and expression of insulin-like growth factor-II and H19 in normal breast tissue and breast tumor. *J Clin Endocrinol Metab*. 1996;81(4):1607-12.
- [66] Lehmann U, Hasemeier B, Romermann D, Muller M, Langer F, Kreipe H. [Epigenetic inactivation of microRNA genes in mammary carcinoma]. *Verh Dtsch Ges Pathol*. 2007;91:214-20.
- [67] Kim SJ, Kelly WK, Fu A, Haines K, Hoffman A, Zheng T, et al. Genome-wide methylation analysis identifies involvement of TNF-alpha mediated cancer pathways in prostate cancer. *Cancer Lett*. 2011;302(1):47-53.
- [68] Bedford MT, van Helden PD. Hypomethylation of DNA in pathological conditions of the human prostate. *Cancer Res*. 1987;47(20):5274-6.
- [69] Pakneshan P, Szyf M, Rabbani SA. Hypomethylation of urokinase (uPA) promoter in breast and prostate cancer: prognostic and therapeutic implications. *Curr Cancer Drug Targets*. 2005;5(7):471-88.
- [70] Cho B, Lee H, Jeong S, Bang YJ, Lee HJ, Hwang KS, et al. Promoter hypomethylation of a novel cancer/testis antigen gene CAGE is correlated with its aberrant expression and is seen in premalignant stage of gastric carcinoma. *Biochem Biophys Res Commun*. 2003;307(1):52-63.
- [71] Tokizane T, Shiina H, Igawa M, Enokida H, Urakami S, Kawakami T, et al. Cytochrome P450 1B1 is overexpressed and regulated by hypomethylation in prostate cancer. *Clin Cancer Res*. 2005;11(16):5793-801.
- [72] Laner T, Schulz WA, Engers R, Muller M, Florl AR. Hypomethylation of the XIST gene promoter in prostate cancer. *Oncol Res*. 2005;15(5):257-64.
- [73] Dammann R, Schagdarsurengin U, Liu L, Otto N, Gimm O, Dralle H, et al. Frequent RASSF1A promoter hypermethylation and K-ras mutations in pancreatic carcinoma. *Oncogene*. 2003;22(24):3806-12.
- [74] Rauch TA, Zhong X, Wu X, Wang M, Kernstine KH, Wang Z, et al. High-resolution mapping of DNA hypermethylation and hypomethylation in lung cancer. *Proc Natl Acad Sci USA*. 2008;105(1):252-7.
- [75] Glazer CA, Smith IM, Ochs MF, Begum S, Westra W, Chang SS, et al. Integrative discovery of epigenetically derepressed cancer testis antigens in NSCLC. *PLoS One*. 2009;4(12):e8189.
- [76] Paz MF, Avila S, Fraga MF, Pollan M, Capella G, Peinado MA, et al. Germ-line variants in methyl-group metabolism genes and susceptibility to DNA methylation in normal tissues and human primary tumors. *Cancer Res*. 2002;62(15):4519-24.

- [77] Pereira MA, Tao L, Liu Y, Li L, Steele VE, Lubet RA. Modulation by budesonide of DNA methylation and mRNA expression in mouse lung tumors. *Int J Cancer*. 2007;120(5):1150-3.
- [78] Katayama H, Hiraki A, Fujiwara K, Matsuo K, Maeda T, Chikamori K, et al. Aberrant promoter methylation profile in pleural fluid DNA and clinicopathological factors in patients with non-small cell lung cancer. *Asian Pac J Cancer Prev*. 2007;8(2):221-4.
- [79] Licchesi JD, Westra WH, Hooker CM, Herman JG. Promoter hypermethylation of hallmark cancer genes in atypical adenomatous hyperplasia of the lung. *Clin Cancer Res*. 2008;14(9):2570-8.
- [80] Ambler RP, Rees MW. Epsilon-N-Methyl-lysine in bacterial flagellar protein. *Nature*. 1959;184:56-7.
- [81] Murray K. The Occurrence of Epsilon-N-Methyl Lysine in Histones. *Biochemistry*. 1964;3:10-5.
- [82] Kouzarides T. Chromatin modifications and their function. *Cell*. 2007;128(4):693-705.
- [83] Cheung P, Allis CD, Sassone-Corsi P. Signaling to chromatin through histone modifications. *Cell*. 2000;103(2):263-71.
- [84] Luger K. Structure and dynamic behavior of nucleosomes. *Curr Opin Genet Dev*. 2003;13(2):127-35.
- [85] Luger K, Mader AW, Richmond RK, Sargent DF, Richmond TJ. Crystal structure of the nucleosome core particle at 2.8 Å resolution. *Nature*. 1997;389(6648):251-60.
- [86] Zhang Y, Reinberg D. Transcription regulation by histone methylation: interplay between different covalent modifications of the core histone tails. *Genes Dev*. 2001;15(18):2343-60.
- [87] Margueron R, Trojer P, Reinberg D. The key to development: interpreting the histone code? *Curr Opin Genet Dev*. 2005;15(2):163-76.
- [88] Vougiouklakis T, Hamamoto R, Nakamura Y, Saloura V. The NSD family of protein methyltransferases in human cancer. *Epigenomics*. 2015:1-12.
- [89] Wagner EJ, Carpenter PB. Understanding the language of Lys36 methylation at histone H3. *Nat Rev Mol Cell Biol*. 2012;13(2):115-26.
- [90] Wang GG, Cai L, Pasillas MP, Kamps MP. NUP98-NSD1 links H3K36 methylation to Hox-A gene activation and leukaemogenesis. *Nat Cell Biol*. 2007;9(7):804-12.
- [91] Hollink IH, van den Heuvel-Eibrink MM, Arentsen-Peters ST, Pratercorona M, Abbas S, Kuipers JE, et al. NUP98/NSD1 characterizes a novel poor prognostic group in acute myeloid leukemia with a distinct HOX gene expression pattern. *Blood*. 2011;118(13):3645-56.
- [92] Thol F, Kolking B, Hollink IH, Damm F, van den Heuvel-Eibrink MM, Michel Zwaan C, et al. Analysis of NUP98/NSD1 translocations in adult AML and MDS patients. *Leukemia*. 2013;27(3):750-4.
- [93] Fasan A, Haferlach C, Alpermann T, Kern W, Haferlach T, Schnittger S. A rare but specific subset of adult AML patients can be defined by the cytogenetically cryptic NUP98-NSD1 fusion gene. *Leukemia*. 2013;27(1):245-8.
- [94] Toyokawa G, Cho HS, Iwai Y, Yoshimatsu M, Takawa M, Hayami S, et al. The histone demethylase JMJD2B plays an essential role in human carcinogenesis through positive regulation of cyclin-dependent kinase 6. *Cancer Prev Res (Phila)*. 2011;4(12):2051-61.

- [95] Hudlebusch HR, Skotte J, Santoni-Rugiu E, Zimling ZG, Lees MJ, Simon R, et al. MMSET is highly expressed and associated with aggressiveness in neuroblastoma. *Cancer Res.* 2011;71(12):4226-35.
- [96] Saloura V, Cho HS, Kiyotani K, Alachkar H, Zuo Z, Nakakido M, et al. WHSC1 promotes oncogenesis through regulation of NIMA-related kinase-7 in squamous cell carcinoma of the head and neck. *Mol Cancer Res.* 2015;13(2):293-304.
- [97] Xiao M, Yang S, Chen J, Ning X, Guo L, Huang K, et al. Overexpression of MMSET in endometrial cancer: a clinicopathologic study. *J Surg Oncol.* 2013;107(4):428-32.
- [98] Zhou P, Wu LL, Wu KM, Jiang W, Li JD, Zhou LD, et al. Overexpression of MMSET is correlation with poor prognosis in hepatocellular carcinoma. *Pathol Oncol Res.* 2013;19(2):303-9.
- [99] Angrand PO, Apiou F, Stewart AF, Dutrillaux B, Losson R, Chambon P. NSD3, a new SET domain-containing gene, maps to 8p12 and is amplified in human breast cancer cell lines. *Genomics.* 2001;74(1):79-88.
- [100] Liu L, Kimball S, Liu H, Holowatyj A, Yang ZQ. Genetic alterations of histone lysine methyltransferases and their significance in breast cancer. *Oncotarget.* 2015;6(4):2466-82.
- [101] Pan Y-XZ, Y.; Chen, D. Epigenetic Mechanisms of Colon Cancer Prevention: What Can Nutrition Do? In *Molecular Mechanisms and Physiology of Disease*. New York, NY, USA: *Springer*; 2014. p. pp. 277–353.
- [102] Volkel P, Dupret B, Le Bourhis X, Angrand PO. Diverse involvement of EZH2 in cancer epigenetics. *Am J Transl Res.* 2015;7(2):175-93.
- [103] Muller J, Hart CM, Francis NJ, Vargas ML, Sengupta A, Wild B, et al. Histone methyltransferase activity of a Drosophila Polycomb group repressor complex. *Cell.* 2002;111(2):197-208.
- [104] Chang CJ, Yang JY, Xia W, Chen CT, Xie X, Chao CH, et al. EZH2 promotes expansion of breast tumor initiating cells through activation of RAF1-beta-catenin signaling. *Cancer Cell.* 2011;19(1):86-100.
- [105] Canaani E, Nakamura T, Rozovskaia T, Smith ST, Mori T, Croce CM, et al. ALL-1/MLL1, a homologue of Drosophila TRITHORAX, modifies chromatin and is directly involved in infant acute leukaemia. *Br J Cancer.* 2004;90(4):756-60.
- [106] Liu H, Takeda S, Kumar R, Westergard TD, Brown EJ, Pandita TK, et al. Phosphorylation of MLL by ATR is required for execution of mammalian S-phase checkpoint. *Nature.* 2010;467(7313):343-6.
- [107] Tell R, Rivera CA, Eskra J, Taglia LN, Blunier A, Wang QT, et al. Gastrin-releasing peptide signaling alters colon cancer invasiveness via heterochromatin protein 1Hsbeta. *Am J Pathol.* 2011;178(2):672-8.
- [108] Chen MW, Hua KT, Kao HJ, Chi CC, Wei LH, Johansson G, et al. H3K9 histone methyltransferase G9a promotes lung cancer invasion and metastasis by silencing the cell adhesion molecule Ep-CAM. *Cancer Res.* 2010;70(20):7830-40.
- [109] Watanabe H, Soejima K, Yasuda H, Kawada I, Nakachi I, Yoda S, et al. Deregulation of histone lysine methyltransferases contributes to oncogenic transformation of human bronchoepithelial cells. *Cancer Cell Int.* 2008;8:15.
- [110] Wang J, Zhou Y, Yin B, Du G, Huang X, Li G, et al. ASH2L: alternative splicing and downregulation during induced megakaryocytic differentiation of multipotential leukemia cell lines. *J Mol Med (Berl).* 2001;79(7):399-405.

- [111] Magerl C, Ellinger J, Braunschweig T, Kremmer E, Koch LK, Holler T, et al. H3K4 dimethylation in hepatocellular carcinoma is rare compared with other hepatobiliary and gastrointestinal carcinomas and correlates with expression of the methylase Ash2 and the demethylase LSD1. *Hum Pathol.* 2010;41(2):181-9.
- [112] Xi Y, Formentini A, Nakajima G, Kornmann M, Ju J. Validation of biomarkers associated with 5-fluorouracil and thymidylate synthase in colorectal cancer. *Oncol Rep.* 2008;19(1):257-62.
- [113] Wang XQ, Miao X, Cai Q, Garcia-Barcelo MM, Fan ST. SMYD3 tandem repeats polymorphism is not associated with the occurrence and metastasis of hepatocellular carcinoma in a Chinese population. *Exp Oncol.* 2007;29(1):71-3.
- [114] Morishita M, di Luccio E. Cancers and the NSD family of histone lysine methyltransferases. *Biochim Biophys Acta.* 2011;1816(2):158-63.
- [115] Nakshatri H, Appaiah HN, Anjanappa M, Gilley D, Tanaka H, Badve S, et al. NF-kappaB-dependent and -independent epigenetic modulation using the novel anti-cancer agent DMAPT. *Cell Death Dis.* 2015;6:e1608.
- [116] Strahl BD, Allis CD. The language of covalent histone modifications. *Nature.* 2000;403(6765):41-5.
- [117] Shi Y, Lan F, Matson C, Mulligan P, Whetstine JR, Cole PA, et al. Histone demethylation mediated by the nuclear amine oxidase homolog LSD1. *Cell.* 2004;119(7):941-53.
- [118] Shi Y. Histone lysine demethylases: emerging roles in development, physiology and disease. *Nat Rev Genet.* 2007;8(11):829-33.
- [119] Tsukada Y, Fang J, Erdjument-Bromage H, Warren ME, Borchers CH, Tempst P, et al. Histone demethylation by a family of JmjC domain-containing proteins. *Nature.* 2006;439(7078):811-6.
- [120] Klose RJ, Kallin EM, Zhang Y. JmjC-domain-containing proteins and histone demethylation. *Nat Rev Genet.* 2006;7(9):715-27.
- [121] Klose RJ, Yamane K, Bae Y, Zhang D, Erdjument-Bromage H, Tempst P, et al. The transcriptional repressor JHDM3A demethylates trimethyl histone H3 lysine 9 and lysine 36. *Nature.* 2006;442(7100):312-6.
- [122] Karytinis A, Forneris F, Profumo A, Ciossani G, Battaglioli E, Binda C, et al. A novel mammalian flavin-dependent histone demethylase. *J Biol Chem.* 2009;284(26):17775-82.
- [123] Liu Y, Liu K, Qin S, Xu C, Min J. Epigenetic targets and drug discovery: part 1: histone methylation. *Pharmacol Ther.* 2014;143(3):275-94.
- [124] Liu K, Liu Y, Lau JL, Min J. Epigenetic targets and drug discovery Part 2: Histone demethylation and DNA methylation. *Pharmacol Ther.* 2015;151:121-40.
- [125] Dong C, Zhang H, Xu C, Arrowsmith CH, Min J. Structure and function of dioxygenases in histone demethylation and DNA/RNA demethylation. *IUCrJ.* 2014;1(Pt 6):540-9.
- [126] Hayami S, Kelly JD, Cho HS, Yoshimatsu M, Unoki M, Tsunoda T, et al. Overexpression of LSD1 contributes to human carcinogenesis through chromatin regulation in various cancers. *Int J Cancer.* 2011;128(3):574-86.
- [127] Suikki HE, Kujala PM, Tammela TL, van Weerden WM, Vessella RL, Visakorpi T. Genetic alterations and changes in expression of histone demethylases in prostate cancer. *Prostate.* 2010;70(8):889-98.

- [128] Guenther MG, Jenner RG, Chevalier B, Nakamura T, Croce CM, Canaani E, et al. Global and Hox-specific roles for the MLL1 methyltransferase. *Proc Natl Acad Sci U S A*. 2005;102(24):8603-8.
- [129] Harris WJ, Huang X, Lynch JT, Spencer GJ, Hitchin JR, Li Y, et al. The histone demethylase KDM1A sustains the oncogenic potential of MLL-AF9 leukemia stem cells. *Cancer Cell*. 2012;21(4):473-87.
- [130] Lim S, Janzer A, Becker A, Zimmer A, Schule R, Buettner R, et al. Lysine-specific demethylase 1 (LSD1) is highly expressed in ER-negative breast cancers and a biomarker predicting aggressive biology. *Carcinogenesis*. 2010;31(3):512-20.
- [131] Katz TA, Vasilatos SN, Harrington E, Oesterreich S, Davidson NE, Huang Y. Inhibition of histone demethylase, LSD2 (KDM1B), attenuates DNA methylation and increases sensitivity to DNMT inhibitor-induced apoptosis in breast cancer cells. *Breast Cancer Res Treat*. 2014;146(1):99-108.
- [132] Rotili D, Mai A. Targeting Histone Demethylases: A New Avenue for the Fight against Cancer. *Genes Cancer*. 2011;2(6):663-79.
- [133] Cloos PA, Christensen J, Agger K, Helin K. Erasing the methyl mark: histone demethylases at the center of cellular differentiation and disease. *Genes Dev*. 2008;22(9):1115-40.
- [134] Pedersen MT, Helin K. Histone demethylases in development and disease. *Trends Cell Biol*. 2010;20(11):662-71.
- [135] Lu T, Jackson MW, Wang B, Yang M, Chance MR, Miyagi M, et al. Regulation of NF-kappaB by NSD1/FBXL11-dependent reversible lysine methylation of p65. *Proc Natl Acad Sci U S A*. 2010;107(1):46-51.
- [136] Tanaka Y, Okamoto K, Teye K, Umata T, Yamagiwa N, Suto Y, et al. JmjC enzyme KDM2A is a regulator of rRNA transcription in response to starvation. *Embo J*. 2010;29(9):1510-22.
- [137] Huang Y, Liu Y, Yu L, Chen J, Hou J, Cui L, et al. Histone demethylase KDM2A promotes tumor cell growth and migration in gastric cancer. *Tumour Biol*. 2015;36(1):271-8.
- [138] Wagner KW, Alam H, Dhar SS, Giri U, Li N, Wei Y, et al. KDM2A promotes lung tumorigenesis by epigenetically enhancing ERK1/2 signaling. *J Clin Invest*. 2013;123(12):5231-46.
- [139] Curtis C, Shah SP, Chin SF, Turashvili G, Rueda OM, Dunning MJ, et al. The genomic and transcriptomic architecture of 2,000 breast tumours reveals novel subgroups. *Nature*. 2012;486(7403):346-52.
- [140] Tzatsos A, Paskaleva P, Ferrari F, Deshpande V, Stoykova S, Contino G, et al. KDM2B promotes pancreatic cancer via Polycomb-dependent and -independent transcriptional programs. *J Clin Invest*. 2013;123(2):727-39.
- [141] Katoh M, Katoh M. Identification and characterization of JMJD2 family genes in silico. *Int J Oncol*. 2004;24(6):1623-8.
- [142] Ponnaluri VK, Vavilala DT, Putty S, Gutheil WG, Mukherji M. Identification of non-histone substrates for JMJD2A-C histone demethylases. *Biochem Biophys Res Commun*. 2009;390(2):280-4.
- [143] Berry WL, Shin S, Lightfoot SA, Janknecht R. Oncogenic features of the JMJD2A histone demethylase in breast cancer. *Int J Oncol*. 2012;41(5):1701-6.

- [144] Berry WL, Janknecht R. KDM4/JMJD2 histone demethylases: epigenetic regulators in cancer cells. *Cancer Res.* 2013;73(10):2936-42.
- [145] Berdel B, Nieminen K, Soini Y, Tengstrom M, Malinen M, Kosma VM, et al. Histone demethylase GASC1--a potential prognostic and predictive marker in invasive breast cancer. *BMC Cancer.* 2012;12:516.
- [146] Rui L, Emre NC, Kruhlak MJ, Chung HJ, Steidl C, Slack G, et al. Cooperative epigenetic modulation by cancer amplicon genes. *Cancer Cell.* 2010;18(6):590-605.
- [147] Paolicchi E, Crea F, Farrar WL, Green JE, Danesi R. Histone lysine demethylases in breast cancer. *Crit Rev Oncol Hematol.* 2013;86(2):97-103.
- [148] Bianco-Miotto T, Chiam K, Buchanan G, Jindal S, Day TK, Thomas M, et al. Global levels of specific histone modifications and an epigenetic gene signature predict prostate cancer progression and development. *Cancer Epidemiol Biomarkers Prev.* 2010;19(10):2611-22.
- [149] Chen Z, Wang L, Wang Q, Li W. Histone modifications and chromatin organization in prostate cancer. *Epigenomics.* 2010;2(4):551-60.
- [150] Ellinger J, Kahl P, von der Gathen J, Rogenhofer S, Heukamp LC, Gutgemann I, et al. Global levels of histone modifications predict prostate cancer recurrence. *Prostate.* 2010;70(1):61-9.
- [151] Zeng J, Ge Z, Wang L, Li Q, Wang N, Bjorkholm M, et al. The histone demethylase RBP2 Is overexpressed in gastric cancer and its inhibition triggers senescence of cancer cells. *Gastroenterology.* 2010;138(3):981-92.
- [152] Qi L, Zhu F, Li SH, Si LB, Hu LK, Tian H. Retinoblastoma binding protein 2 (RBP2) promotes HIF-1 α -VEGF-induced angiogenesis of non-small cell lung cancer via the Akt pathway. *PLoS One.* 2014;9(8):e106032.
- [153] Hayami S, Yoshimatsu M, Veerakumarasivam A, Unoki M, Iwai Y, Tsunoda T, et al. Overexpression of the JmjC histone demethylase KDM5B in human carcinogenesis: involvement in the proliferation of cancer cells through the E2F/RB pathway. *Mol Cancer.* 2010;9:59.
- [154] Yamamoto S, Wu Z, Russnes HG, Takagi S, Peluffo G, Vaske C, et al. JARID1B is a luminal lineage-driving oncogene in breast cancer. *Cancer Cell.* 2014;25(6):762-77.
- [155] Agger K, Christensen J, Cloos PA, Helin K. The emerging functions of histone demethylases. *Curr Opin Genet Dev.* 2008;18(2):159-68.
- [156] Walport LJ, Hopkinson RJ, Vollmar M, Madden SK, Gileadi C, Oppermann U, et al. Human UTY(KDM6C) is a male-specific N-methyl lysyl demethylase. *J Biol Chem.* 2014;289(26):18302-13.
- [157] van Haaften G, Dalgliesh GL, Davies H, Chen L, Bignell G, Greenman C, et al. Somatic mutations of the histone H3K27 demethylase gene UTX in human cancer. *Nat Genet.* 2009;41(5):521-3.
- [158] Dalgliesh GL, Furge K, Greenman C, Chen L, Bignell G, Butler A, et al. Systematic sequencing of renal carcinoma reveals inactivation of histone modifying genes. *Nature.* 2010;463(7279):360-3.
- [159] Agger K, Cloos PA, Rudkjaer L, Williams K, Andersen G, Christensen J, et al. The H3K27me3 demethylase JMJD3 contributes to the activation of the INK4A-ARF locus in response to oncogene- and stress-induced senescence. *Genes Dev.* 2009;23(10):1171-6.

- [160] Stasevich TJ, Hayashi-Takanaka Y, Sato Y, Maehara K, Ohkawa Y, Sakata-Sogawa K, et al. Regulation of RNA polymerase II activation by histone acetylation in single living cells. *Nature*. 2014;516(7530):272-5.
- [161] Shogren-Knaak M, Ishii H, Sun JM, Pazin MJ, Davie JR, Peterson CL. Histone H4-K16 acetylation controls chromatin structure and protein interactions. *Science*. 2006;311(5762):844-7.
- [162] Celic I, Masumoto H, Griffith WP, Meluh P, Cotter RJ, Boeke JD, et al. The sirtuins hst3 and Hst4p preserve genome integrity by controlling histone h3 lysine 56 deacetylation. *Curr Biol*. 2006;16(13):1280-9.
- [163] Choudhary C, Kumar C, Gnad F, Nielsen ML, Rehman M, Walther TC, et al. Lysine acetylation targets protein complexes and co-regulates major cellular functions. *Science*. 2009;325(5942):834-40.
- [164] Zhao S, Xu W, Jiang W, Yu W, Lin Y, Zhang T, et al. Regulation of cellular metabolism by protein lysine acetylation. *Science*. 2010;327(5968):1000-4.
- [165] Hornbeck PV, Kornhauser JM, Tkachev S, Zhang B, Skrzypek E, Murray B, et al. PhosphoSitePlus: a comprehensive resource for investigating the structure and function of experimentally determined post-translational modifications in man and mouse. *Nucleic Acids Res*. 2012;40(Database issue):D261-70.
- [166] Kouzarides T. Acetylation: a regulatory modification to rival phosphorylation? *Embo J*. 2000;19(6):1176-9.
- [167] You L, Nie J, Sun WJ, Zheng ZQ, Yang XJ. Lysine acetylation: enzymes, bromodomains and links to different diseases. *Essays Biochem*. 2012;52:1-12.
- [168] Unnikrishnan A, Gafken PR, Tsukiyama T. Dynamic changes in histone acetylation regulate origins of DNA replication. *Nat Struct Mol Biol*. 2010;17(4):430-7.
- [169] Vo N, Goodman RH. CREB-binding protein and p300 in transcriptional regulation. *J Biol Chem*. 2001;276(17):13505-8.
- [170] Capell BC, Berger SL. Genome-wide epigenetics. *J Invest Dermatol*. 2013;133(6):e9.
- [171] Henry RA, Kuo YM, Andrews AJ. Differences in specificity and selectivity between CBP and p300 acetylation of histone H3 and H3/H4. *Biochemistry*. 2013;52(34):5746-59.
- [172] Kalkhoven E. CBP and p300: HATs for different occasions. *Biochem Pharmacol*. 2004;68(6):1145-55.
- [173] Santer FR, Hoschele PP, Oh SJ, Erb HH, Bouchal J, Cavarretta IT, et al. Inhibition of the acetyltransferases p300 and CBP reveals a targetable function for p300 in the survival and invasion pathways of prostate cancer cell lines. *Mol Cancer Ther*. 2011;10(9):1644-55.
- [174] Iyer NG, Ozdag H, Caldas C. p300/CBP and cancer. *Oncogene*. 2004;23(24):4225-31.
- [175] Gayther SA, Batley SJ, Linger L, Bannister A, Thorpe K, Chin SF, et al. Mutations truncating the EP300 acetylase in human cancers. *Nat Genet*. 2000;24(3):300-3.
- [176] Bryan EJ, Jokubaitis VJ, Chamberlain NL, Baxter SW, Dawson E, Choong DY, et al. Mutation analysis of EP300 in colon, breast and ovarian carcinomas. *Int J Cancer*. 2002;102(2):137-41.
- [177] Ozdag H, Batley SJ, Forsti A, Iyer NG, Daigo Y, Boutell J, et al. Mutation analysis of CBP and PCAF reveals rare inactivating mutations in cancer cell lines but not in primary tumours. *Br J Cancer*. 2002;87(10):1162-5.

- [178] Grunstein M. Histone acetylation in chromatin structure and transcription. *Nature*. 1997;389(6649):349-52.
- [179] Sun XJ, Man N, Tan Y, Nimer SD, Wang L. The Role of Histone Acetyltransferases in Normal and Malignant Hematopoiesis. *Front Oncol*. 2015;5:108.
- [180] Yang XJ, Ullah M. MOZ and MORF, two large MYSTic HATs in normal and cancer stem cells. *Oncogene*. 2007;26(37):5408-19.
- [181] Dulak AM, Stojanov P, Peng S, Lawrence MS, Fox C, Stewart C, et al. Exome and whole-genome sequencing of esophageal adenocarcinoma identifies recurrent driver events and mutational complexity. *Nat Genet*. 2013;45(5):478-86.
- [182] Wu X, Northcott PA, Dubuc A, Dupuy AJ, Shih DJ, Witt H, et al. Clonal selection drives genetic divergence of metastatic medulloblastoma. *Nature*. 2012;482(7386):529-33.
- [183] Moore SD, Herrick SR, Ince TA, Kleinman MS, Dal Cin P, Morton CC, et al. Uterine leiomyomata with t(10;17) disrupt the histone acetyltransferase MORF. *Cancer Res*. 2004;64(16):5570-7.
- [184] Panagopoulos I, Gorunova L, Bjerkehagen B, Heim S. Novel KAT6B-KANSL1 fusion gene identified by RNA sequencing in retroperitoneal leiomyoma with t(10;17)(q22;q21). *PLoS One*. 2015;10(1):e0117010.
- [185] Lynch H, Wen H, Kim YC, Snyder C, Kinarsky Y, Chen PX, et al. Can unknown predisposition in familial breast cancer be family-specific? *Breast J*. 2013;19(5):520-8.
- [186] Grasso CS, Wu YM, Robinson DR, Cao X, Dhanasekaran SM, Khan AP, et al. The mutational landscape of lethal castration-resistant prostate cancer. *Nature*. 2012;487(7406):239-43.
- [187] Zack TI, Schumacher SE, Carter SL, Cherniack AD, Saksena G, Tabak B, et al. Pan-cancer patterns of somatic copy number alteration. *Nat Genet*. 2013;45(10):1134-40.
- [188] Jin Q, Yu LR, Wang L, Zhang Z, Kasper LH, Lee JE, et al. Distinct roles of GCN5/PCAF-mediated H3K9ac and CBP/p300-mediated H3K18/27ac in nuclear receptor transactivation. *Embo J*. 2011;30(2):249-62.
- [189] Zimmer SN, Lemieux ME, Karia BP, Day C, Zhou T, Zhou Q, et al. Mice heterozygous for CREB binding protein are hypersensitive to gamma-radiation and invariably develop myelodysplastic/myelo proliferative neoplasm. *Exp Hematol*. 2012;40(4):295-306 e5.
- [190] Gui Y, Guo G, Huang Y, Hu X, Tang A, Gao S, et al. Frequent mutations of chromatin remodeling genes in transitional cell carcinoma of the bladder. *Nat Genet*. 2011;43(9):875-8.
- [191] Pasqualucci L, Dominguez-Sola D, Chiarenza A, Fabbri G, Grunn A, Trifonov V, et al. Inactivating mutations of acetyltransferase genes in B-cell lymphoma. *Nature*. 2011;471(7337):189-95.
- [192] Mullighan CG, Zhang J, Kasper LH, Lerach S, Payne-Turner D, Phillips LA, et al. CREBBP mutations in relapsed acute lymphoblastic leukaemia. *Nature*. 2011;471(7337):235-9.
- [193] Ionov Y, Matsui S, Cowell JK. A role for p300/CREB binding protein genes in promoting cancer progression in colon cancer cell lines with microsatellite instability. *Proc Natl Acad Sci U S A*. 2004;101(5):1273-8.
- [194] Gao Y, Geng J, Hong X, Qi J, Teng Y, Yang Y, et al. Expression of p300 and CBP is associated with poor prognosis in small cell lung cancer. *Int J Clin Exp Pathol*. 2014;7(2):760-7.

- [195] Liao ZW, Zhou TC, Tan XJ, Song XL, Liu Y, Shi XY, et al. High expression of p300 is linked to aggressive features and poor prognosis of nasopharyngeal carcinoma. *J Transl Med.* 2012;10:110.
- [196] Rotte A, Bhandaru M, Cheng Y, Sjoestroem C, Martinka M, Li G. Decreased expression of nuclear p300 is associated with disease progression and worse prognosis of melanoma patients. *PLoS One.* 2013;8(9):e75405.
- [197] Inuzuka H, Gao D, Finley LW, Yang W, Wan L, Fukushima H, et al. Acetylation-dependent regulation of Skp2 function. *Cell.* 2012;150(1):179-93.
- [198] Chen G, Cheng Y, Tang Y, Martinka M, Li G. Role of Tip60 in human melanoma cell migration, metastasis, and patient survival. *J Invest Dermatol.* 2012;132(11):2632-41.
- [199] Zhao H, Jin S, Gewirtz AM. The histone acetyltransferase TIP60 interacts with c-Myb and inactivates its transcriptional activity in human leukemia. *J Biol Chem.* 2012;287(2):925-34.
- [200] Chevillard-Briet M, Quaranta M, Grezy A, Mattera L, Courilleau C, Philippe M, et al. Interplay between chromatin-modifying enzymes controls colon cancer progression through Wnt signaling. *Hum Mol Genet.* 2014;23(8):2120-31.
- [201] Gorrini C, Squatrito M, Luise C, Syed N, Perna D, Wark L, et al. Tip60 is a haplo-insufficient tumour suppressor required for an oncogene-induced DNA damage response. *Nature.* 2007;448(7157):1063-7.
- [202] Halkidou K, Gnanapragasam VJ, Mehta PB, Logan IR, Brady ME, Cook S, et al. Expression of Tip60, an androgen receptor coactivator, and its role in prostate cancer development. *Oncogene.* 2003;22(16):2466-77.
- [203] ME LL, Vidal F, Gallardo D, Diaz-Fuertes M, Rojo F, Cuatrecasas M, et al. New p53 related genes in human tumors: significant downregulation in colon and lung carcinomas. *Oncol Rep.* 2006;16(3):603-8.
- [204] Mattera L, Escaffit F, Pillaire MJ, Selves J, Tyteca S, Hoffmann JS, et al. The p400/Tip60 ratio is critical for colorectal cancer cell proliferation through DNA damage response pathways. *Oncogene.* 2009;28(12):1506-17.
- [205] Shiota M, Yokomizo A, Masubuchi D, Tada Y, Inokuchi J, Eto M, et al. Tip60 promotes prostate cancer cell proliferation by translocation of androgen receptor into the nucleus. *Prostate.* 2010;70(5):540-54.
- [206] Wang C, Fu M, Angeletti RH, Siconolfi-Baez L, Reutens AT, Albanese C, et al. Direct acetylation of the estrogen receptor alpha hinge region by p300 regulates transactivation and hormone sensitivity. *J Biol Chem.* 2001;276(21):18375-83.
- [207] Fuqua SA, Wiltshcke C, Zhang QX, Borg A, Castles CG, Friedrichs WE, et al. A hypersensitive estrogen receptor-alpha mutation in premalignant breast lesions. *Cancer Res.* 2000;60(15):4026-9.
- [208] Kim MY, Woo EM, Chong YT, Homenko DR, Kraus WL. Acetylation of estrogen receptor alpha by p300 at lysines 266 and 268 enhances the deoxyribonucleic acid binding and transactivation activities of the receptor. *Mol Endocrinol.* 2006;20(7):1479-93.
- [209] Fan S, Ma YX, Wang C, Yuan RQ, Meng Q, Wang JA, et al. p300 Modulates the BRCA1 inhibition of estrogen receptor activity. *Cancer Res.* 2002;62(1):141-51.
- [210] Amanatullah DF, Reutens AT, Zafonte BT, Fu M, Mani S, Pestell RG. Cell-cycle dysregulation and the molecular mechanisms of prostate cancer. *Front Biosci.* 2000;5:D372-90.

- [211] Jemal A, Murray T, Samuels A, Ghafoor A, Ward E, Thun MJ. Cancer statistics, 2003. *CA Cancer J Clin.* 2003;53(1):5-26.
- [212] Popov VM, Wang C, Shirley LA, Rosenberg A, Li S, Nevalainen M, et al. The functional significance of nuclear receptor acetylation. *Steroids.* 2007;72(2):221-30.
- [213] Fu M, Wang C, Reutens AT, Wang J, Angeletti RH, Siconolfi-Baez L, et al. p300 and p300/cAMP-response element-binding protein-associated factor acetylate the androgen receptor at sites governing hormone-dependent transactivation. *J Biol Chem.* 2000;275(27):20853-60.
- [214] Fu M, Wang C, Wang J, Zhang X, Sakamaki T, Yeung YG, et al. Androgen receptor acetylation governs trans activation and MEKK1-induced apoptosis without affecting in vitro sumoylation and trans-repression function. *Mol Cell Biol.* 2002;22(10):3373-88.
- [215] Fu M, Rao M, Wang C, Sakamaki T, Wang J, Di Vizio D, et al. Acetylation of androgen receptor enhances coactivator binding and promotes prostate cancer cell growth. *Mol Cell Biol.* 2003;23(23):8563-75.
- [216] Fu M, Rao M, Wu K, Wang C, Zhang X, Hessian M, et al. The androgen receptor acetylation site regulates cAMP and AKT but not ERK-induced activity. *J Biol Chem.* 2004;279(28):29436-49.
- [217] Gong J, Zhu J, Goodman OB, Jr., Pestell RG, Schlegel PN, Nanus DM, et al. Activation of p300 histone acetyltransferase activity and acetylation of the androgen receptor by bombesin in prostate cancer cells. *Oncogene.* 2006;25(14):2011-21.
- [218] Seligson DB, Horvath S, Shi T, Yu H, Tze S, Grunstein M, et al. Global histone modification patterns predict risk of prostate cancer recurrence. *Nature.* 2005;435(7046):1262-6.
- [219] Debes JD, Sebo TJ, Lohse CM, Murphy LM, Haugen DA, Tindall DJ. p300 in prostate cancer proliferation and progression. *Cancer Res.* 2003;63(22):7638-40.
- [220] Fraga MF, Ballestar E, Villar-Garea A, Boix-Chornet M, Espada J, Schotta G, et al. Loss of acetylation at Lys16 and trimethylation at Lys20 of histone H4 is a common hallmark of human cancer. *Nat Genet.* 2005;37(4):391-400.
- [221] Ropero S, Esteller M. The role of histone deacetylases (HDACs) in human cancer. *Mol Oncol.* 2007;1(1):19-25.
- [222] Ropero S, Fraga MF, Ballestar E, Hamelin R, Yamamoto H, Boix-Chornet M, et al. A truncating mutation of HDAC2 in human cancers confers resistance to histone deacetylase inhibition. *Nat Genet.* 2006;38(5):566-9.
- [223] Swierczynski S, Klier E, Illig R, Alinger-Scharinger B, Kiesslich T, Neureiter D. Histone deacetylation meets miRNA: epigenetics and post-transcriptional regulation in cancer and chronic diseases. *Expert Opin Biol Ther.* 2015;15(5):651-64.
- [224] Poyet C, Jentsch B, Hermanns T, Schweckendiek D, Seifert HH, Schmidtpeter M, et al. Expression of histone deacetylases 1, 2 and 3 in urothelial bladder cancer. *BMC Clin Pathol.* 2014;14(1):10.
- [225] Weichert W, Roske A, Gekeler V, Beckers T, Stephan C, Jung K, et al. Histone deacetylases 1, 2 and 3 are highly expressed in prostate cancer and HDAC2 expression is associated with shorter PSA relapse time after radical prostatectomy. *Br J Cancer.* 2008;98(3):604-10.
- [226] Witt O, Deubzer HE, Milde T, Oehme I. HDAC family: What are the cancer relevant targets? *Cancer Lett.* 2009;277(1):8-21.

- [227] Millard CJ, Watson PJ, Celardo I, Gordiyenko Y, Cowley SM, Robinson CV, et al. Class I HDACs share a common mechanism of regulation by inositol phosphates. *Mol Cell*. 2013;51(1):57-67.
- [228] Barneda-Zahonero B, Parra M. Histone deacetylases and cancer. *Mol Oncol*. 2012;6(6):579-89.
- [229] Ishihama K, Yamakawa M, Semba S, Takeda H, Kawata S, Kimura S, et al. Expression of HDAC1 and CBP/p300 in human colorectal carcinomas. *J Clin Pathol*. 2007;60(11):1205-10.
- [230] Adams H, Fritzsche FR, Dirnhofer S, Kristiansen G, Tzankov A. Class I histone deacetylases 1, 2 and 3 are highly expressed in classical Hodgkin's lymphoma. *Expert Opin Ther Targets*. 2010;14(6):577-84.
- [231] Fritzsche FR, Weichert W, Roske A, Gekeler V, Beckers T, Stephan C, et al. Class I histone deacetylases 1, 2 and 3 are highly expressed in renal cell cancer. *BMC Cancer*. 2008;8:381.
- [232] Oehme I, Deubzer HE, Wegener D, Pickert D, Linke JP, Hero B, et al. Histone deacetylase 8 in neuroblastoma tumorigenesis. *Clin Cancer Res*. 2009;15(1):91-9.
- [233] Krusche CA, Wulfing P, Kersting C, Vloet A, Bocker W, Kiesel L, et al. Histone deacetylase-1 and -3 protein expression in human breast cancer: a tissue microarray analysis. *Breast Cancer Res Treat*. 2005;90(1):15-23.
- [234] Pacheco M, Nielsen TO. Histone deacetylase 1 and 2 in mesenchymal tumors. *Mod Pathol*. 2012;25(2):222-30.
- [235] Jung KH, Noh JH, Kim JK, Eun JW, Bae HJ, Xie HJ, et al. HDAC2 overexpression confers oncogenic potential to human lung cancer cells by deregulating expression of apoptosis and cell cycle proteins. *J Cell Biochem*. 2012;113(6):2167-77.
- [236] Sun JY, Xu L, Tseng H, Ciccarelli B, Fulciniti M, Hunter ZR, et al. Histone deacetylase inhibitors demonstrate significant preclinical activity as single agents, and in combination with bortezomib in Waldenstrom's macroglobulinemia. *Clin Lymphoma Myeloma Leuk*. 2011;11(1):152-6.
- [237] Ozdag H, Teschendorff AE, Ahmed AA, Hyland SJ, Blenkiron C, Bobrow L, et al. Differential expression of selected histone modifier genes in human solid cancers. *BMC Genomics*. 2006;7:90.
- [238] Stark M, Hayward N. Genome-wide loss of heterozygosity and copy number analysis in melanoma using high-density single-nucleotide polymorphism arrays. *Cancer Res*. 2007;67(6):2632-42.
- [239] Sjoblom T, Jones S, Wood LD, Parsons DW, Lin J, Barber TD, et al. The consensus coding sequences of human breast and colorectal cancers. *Science*. 2006;314(5797):268-74.
- [240] Milde T, Oehme I, Korshunov A, Kopp-Schneider A, Remke M, Northcott P, et al. HDAC5 and HDAC9 in medulloblastoma: novel markers for risk stratification and role in tumor cell growth. *Clin Cancer Res*. 2010;16(12):3240-52.
- [241] Lachenmayer A, Toffanin S, Cabellos L, Alsinet C, Hoshida Y, Villanueva A, et al. Combination therapy for hepatocellular carcinoma: additive preclinical efficacy of the HDAC inhibitor panobinostat with sorafenib. *J Hepatol*. 2012;56(6):1343-50.
- [242] Ouaisi M, Sielezneff I, Silvestre R, Sastre B, Bernard JP, Lafontaine JS, et al. High histone deacetylase 7 (HDAC7) expression is significantly associated with adenocarcinomas of the pancreas. *Ann Surg Oncol*. 2008;15(8):2318-28.

- [243] Weichert W. HDAC expression and clinical prognosis in human malignancies. *Cancer Lett.* 2009;280(2):168-76.
- [244] Moreno DA, Scrideli CA, Cortez MA, de Paula Queiroz R, Valera ET, da Silva Silveira V, et al. Differential expression of HDAC3, HDAC7 and HDAC9 is associated with prognosis and survival in childhood acute lymphoblastic leukaemia. *Br J Haematol.* 2010;150(6):665-73.
- [245] Skov V, Larsen TS, Thomassen M, Riley CH, Jensen MK, Bjerrum OW, et al. Increased gene expression of histone deacetylases in patients with Philadelphia-negative chronic myeloproliferative neoplasms. *Leuk Lymphoma.* 2012;53(1):123-9.
- [246] Choi YW, Bae SM, Kim YW, Lee HN, Kim YW, Park TC, et al. Gene expression profiles in squamous cell cervical carcinoma using array-based comparative genomic hybridization analysis. *Int J Gynecol Cancer.* 2007;17(3):687-96.
- [247] Lucio-Eterovic AK, Cortez MA, Valera ET, Motta FJ, Queiroz RG, Machado HR, et al. Differential expression of 12 histone deacetylase (HDAC) genes in astrocytomas and normal brain tissue: class II and IV are hypoexpressed in glioblastomas. *BMC Cancer.* 2008;8:243.
- [248] Rajendran R, Garva R, Krstic-Demonacos M, Demonacos C. Sirtuins: molecular traffic lights in the crossroad of oxidative stress, chromatin remodeling, and transcription. *J Biomed Biotechnol.* 2011;2011:368276.
- [249] Batta K, Das C, Gadad S, Shandilya J, Kundu TK. Reversible acetylation of non histone proteins: role in cellular function and disease. *Subcell Biochem.* 2007;41:193-212.
- [250] Finkel T, Deng CX, Mostoslavsky R. Recent progress in the biology and physiology of sirtuins. *Nature.* 2009;460(7255):587-91.
- [251] Shore D, Squire M, Nasmyth KA. Characterization of two genes required for the position-effect control of yeast mating-type genes. *Embo J.* 1984;3(12):2817-23.
- [252] Yuan H, Su L, Chen WY. The emerging and diverse roles of sirtuins in cancer: a clinical perspective. *Onco Targets Ther.* 2013;6:1399-416.
- [253] de Oliveira MV, Andrade JM, Paraiso AF, Santos SH. Sirtuins and cancer: new insights and cell signaling. *Cancer Invest.* 2013;31(10):645-53.
- [254] Deubzer HE, Schier MC, Oehme I, Lodrini M, Haendler B, Sommer A, et al. HDAC11 is a novel drug target in carcinomas. *Int J Cancer.* 2013;132(9):2200-8.
- [255] Sawa H, Murakami H, Ohshima Y, Sugino T, Nakajyo T, Kisanuki T, et al. Histone deacetylase inhibitors such as sodium butyrate and trichostatin A induce apoptosis through an increase of the bcl-2-related protein Bad. *Brain Tumor Pathol.* 2001;18(2):109-14.
- [256] Jenuwein T, Allis CD. Translating the histone code. *Science.* 2001;293(5532):1074-80.
- [257] Manuyakorn A, Paulus R, Farrell J, Dawson NA, Tze S, Cheung-Lau G, et al. Cellular histone modification patterns predict prognosis and treatment response in resectable pancreatic adenocarcinoma: results from RTOG 9704. *J Clin Oncol.* 2010;28(8):1358-65.
- [258] Seligson DB, Horvath S, McBrien MA, Mah V, Yu H, Tze S, et al. Global levels of histone modifications predict prognosis in different cancers. *Am J Pathol.* 2009;174(5):1619-28.
- [259] Van Den Broeck A, Brambilla E, Moro-Sibilot D, Lantuejoul S, Brambilla C, Eymin B, et al. Loss of histone H4K20 trimethylation occurs in preneoplasia and influences prognosis of non-small cell lung cancer. *Clin Cancer Res.* 2008;14(22):7237-45.

- [260] Barlesi F, Giaccone G, Gallegos-Ruiz MI, Loundou A, Span SW, Lefesvre P, et al. Global histone modifications predict prognosis of resected non small-cell lung cancer. *J Clin Oncol*. 2007;25(28):4358-64.
- [261] Li GM. Decoding the histone code: Role of H3K36me3 in mismatch repair and implications for cancer susceptibility and therapy. *Cancer Res*. 2013;73(21):6379-83.
- [262] Croce CM. Oncogenes and cancer. *N Engl J Med*. 2008;358(5):502-11.
- [263] Meyer N, Penn LZ. Reflecting on 25 years with MYC. *Nat Rev Cancer*. 2008;8(12):976-90.
- [264] Shchors K, Shchors E, Rostker F, Lawlor ER, Brown-Swigart L, Evan GI. The Myc-dependent angiogenic switch in tumors is mediated by interleukin 1beta. *Genes Dev*. 2006;20(18):2527-38.
- [265] Borrello MG, Degl'Innocenti D, Pierotti MA. Inflammation and cancer: the oncogene-driven connection. *Cancer Lett*. 2008;267(2):262-70.
- [266] Grivennikov SI, Karin M. Inflammation and oncogenesis: a vicious connection. *Curr Opin Genet Dev*. 2010;20(1):65-71.
- [267] Huang J, Sengupta R, Espejo AB, Lee MG, Dorsey JA, Richter M, et al. p53 is regulated by the lysine demethylase LSD1. *Nature*. 2007;449(7158):105-8.
- [268] Rak J, Mitsuhashi Y, Bayko L, Filmus J, Shirasawa S, Sasazuki T, et al. Mutant ras oncogenes upregulate VEGF/VPF expression: implications for induction and inhibition of tumor angiogenesis. *Cancer Res*. 1995;55(20):4575-80.
- [269] Rak J, Mitsuhashi Y, Sheehan C, Tamir A, Vilorio-Petit A, Filmus J, et al. Oncogenes and tumor angiogenesis: differential modes of vascular endothelial growth factor up-regulation in ras-transformed epithelial cells and fibroblasts. *Cancer Res*. 2000;60(2):490-8.
- [270] Futscher BW, O'Meara MM, Kim CJ, Rennels MA, Lu D, Gruman LM, et al. Aberrant methylation of the maspin promoter is an early event in human breast cancer. *Neoplasia*. 2004;6(4):380-9.
- [271] Futscher BW, Oshiro MM, Wozniak RJ, Holtan N, Hanigan CL, Duan H, et al. Role for DNA methylation in the control of cell type specific maspin expression. *Nat Genet*. 2002;31(2):175-9.
- [272] Bettstetter M, Woenckhaus M, Wild PJ, Rummele P, Blaszyk H, Hartmann A, et al. Elevated nuclear maspin expression is associated with microsatellite instability and high tumour grade in colorectal cancer. *J Pathol*. 2005;205(5):606-14.
- [273] Wasylshen AR, Kalkat M, Kim SS, Pandya A, Chan PK, Oliveri S, et al. MYC activity is negatively regulated by a C-terminal lysine cluster. *Oncogene*. 2014;33(8):1066-72.
- [274] Patel JH, Du Y, Ard PG, Phillips C, Carella B, Chen CJ, et al. The c-MYC oncoprotein is a substrate of the acetyltransferases hGCN5/PCAF and TIP60. *Mol Cell Biol*. 2004;24(24):10826-34.
- [275] Zhang K, Faiola F, Martinez E. Six lysine residues on c-Myc are direct substrates for acetylation by p300. *Biochem Biophys Res Commun*. 2005;336(1):274-80.
- [276] Yang MH, Nickerson S, Kim ET, Liot C, Laurent G, Spang R, et al. Regulation of RAS oncogenicity by acetylation. *Proc Natl Acad Sci USA*. 2012;109(27):10843-8.
- [277] Diab A, Zickl L, Abdel-Wahab O, Jhanwar S, Gulam MA, Panageas KS, et al. Acute myeloid leukemia with translocation t(8;16) presents with features which mimic acute promyelocytic leukemia and is associated with poor prognosis. *Leuk Res*. 2013;37(1):32-6.

- [278] Krivtsov AV, Armstrong SA. MLL translocations, histone modifications and leukaemia stem-cell development. *Nat Rev Cancer*. 2007;7(11):823-33.
- [279] Chaffanet M, Gressin L, Preudhomme C, Soenen-Cornu V, Birnbaum D, Pebusque MJ. MOZ is fused to p300 in an acute monocytic leukemia with t(8;22). *Genes Chromosomes Cancer*. 2000;28(2):138-44.
- [280] Rozman M, Camos M, Colomer D, Villamor N, Esteve J, Costa D, et al. Type I MOZ/CBP (MYST3/CREBBP) is the most common chimeric transcript in acute myeloid leukemia with t(8;16)(p11;p13) translocation. *Genes Chromosomes Cancer*. 2004;40(2):140-5.
- [281] Wang L, Gural A, Sun XJ, Zhao X, Perna F, Huang G, et al. The leukemogenicity of AML1-ETO is dependent on site-specific lysine acetylation. *Science*. 2011;333(6043):765-9.
- [282] Deguchi K, Ayton PM, Carapeti M, Kutok JL, Snyder CS, Williams IR, et al. MOZ-TIF2-induced acute myeloid leukemia requires the MOZ nucleosome binding motif and TIF2-mediated recruitment of CBP. *Cancer Cell*. 2003;3(3):259-71.
- [283] Kitabayashi I, Aikawa Y, Nguyen LA, Yokoyama A, Ohki M. Activation of AML1-mediated transcription by MOZ and inhibition by the MOZ-CBP fusion protein. *Embo J*. 2001;20(24):7184-96.
- [284] Li M, Luo RZ, Chen JW, Cao Y, Lu JB, He JH, et al. High expression of transcriptional coactivator p300 correlates with aggressive features and poor prognosis of hepatocellular carcinoma. *J Transl Med*. 2011;9:5.
- [285] Xiao XS, Cai MY, Chen JW, Guan XY, Kung HF, Zeng YX, et al. High Expression of p300 in Human Breast Cancer Correlates with Tumor Recurrence and Predicts Adverse Prognosis. *Chin J Cancer Res*. 2011;23(3):201-7.
- [286] Hou X, Li Y, Luo RZ, Fu JH, He JH, Zhang LJ, et al. High expression of the transcriptional co-activator p300 predicts poor survival in resectable non-small cell lung cancers. *Eur J Surg Oncol*. 2012;38(6):523-30.
- [287] Hobbs CA, Wei G, DeFeo K, Paul B, Hayes CS, Gilmour SK. Tip60 protein isoforms and altered function in skin and tumors that overexpress ornithine decarboxylase. *Cancer Res*. 2006;66(16):8116-22.
- [288] Knights CD, Catania J, Di Giovanni S, Muratoglu S, Perez R, Swartzbeck A, et al. Distinct p53 acetylation cassettes differentially influence gene-expression patterns and cell fate. *J Cell Biol*. 2006;173(4):533-44.
- [289] Vousden KH, Lane DP. p53 in health and disease. *Nat Rev Mol Cell Biol*. 2007;8(4):275-83.
- [290] Soussi T, Ishioka C, Claustres M, Beroud C. Locus-specific mutation databases: pitfalls and good practice based on the p53 experience. *Nat Rev Cancer*. 2006;6(1):83-90.
- [291] Gu W, Roeder RG. Activation of p53 sequence-specific DNA binding by acetylation of the p53 C-terminal domain. *Cell*. 1997;90(4):595-606.
- [292] Sakaguchi K, Herrera JE, Saito S, Miki T, Bustin M, Vassilev A, et al. DNA damage activates p53 through a phosphorylation-acetylation cascade. *Genes Dev*. 1998;12(18):2831-41.
- [293] Ito A, Lai CH, Zhao X, Saito S, Hamilton MH, Appella E, et al. p300/CBP-mediated p53 acetylation is commonly induced by p53-activating agents and inhibited by MDM2. *Embo J*. 2001;20(6):1331-40.

- [294] Sykes SM, Mellert HS, Holbert MA, Li K, Marmorstein R, Lane WS, et al. Acetylation of the p53 DNA-binding domain regulates apoptosis induction. *Mol Cell*. 2006;24(6):841-51.
- [295] Tang Y, Luo J, Zhang W, Gu W. Tip60-dependent acetylation of p53 modulates the decision between cell-cycle arrest and apoptosis. *Mol Cell*. 2006;24(6):827-39.
- [296] Brooks CL, Gu W. Ubiquitination, phosphorylation and acetylation: the molecular basis for p53 regulation. *Curr Opin Cell Biol*. 2003;15(2):164-71.
- [297] Friedler A, Veprintsev DB, Freund SM, von Glos KI, Fersht AR. Modulation of binding of DNA to the C-terminal domain of p53 by acetylation. *Structure*. 2005;13(4):629-36.
- [298] Mujtaba S, He Y, Zeng L, Yan S, Plotnikova O, Sachchidanand, et al. Structural mechanism of the bromodomain of the coactivator CBP in p53 transcriptional activation. *Mol Cell*. 2004;13(2):251-63.
- [299] Li AG, Piluso LG, Cai X, Gadd BJ, Ladurner AG, Liu X. An acetylation switch in p53 mediates holo-TFIID recruitment. *Mol Cell*. 2007;28(3):408-21.
- [300] Le Cam L, Linares LK, Paul C, Julien E, Lacroix M, Hatchi E, et al. E4F1 is an atypical ubiquitin ligase that modulates p53 effector functions independently of degradation. *Cell*. 2006;127(4):775-88.
- [301] Toledo F, Wahl GM. Regulating the p53 pathway: in vitro hypotheses, in vivo veritas. *Nat Rev Cancer*. 2006;6(12):909-23.
- [302] Banerjee S, Kumar BR, Kundu TK. General transcriptional coactivator PC4 activates p53 function. *Mol Cell Biol*. 2004;24(5):2052-62.
- [303] Batta K, Kundu TK. Activation of p53 function by human transcriptional coactivator PC4: role of protein-protein interaction, DNA bending, and posttranslational modifications. *Mol Cell Biol*. 2007;27(21):7603-14.
- [304] Nakamura Y. Isolation of p53-target genes and their functional analysis. *Cancer Sci*. 2004;95(1):7-11.
- [305] Chuikov S, Kurash JK, Wilson JR, Xiao B, Justin N, Ivanov GS, et al. Regulation of p53 activity through lysine methylation. *Nature*. 2004;432(7015):353-60.
- [306] Huang J, Perez-Burgos L, Placek BJ, Sengupta R, Richter M, Dorsey JA, et al. Repression of p53 activity by Smyd2-mediated methylation. *Nature*. 2006;444(7119):629-32.
- [307] Shi X, Kachirskia I, Yamaguchi H, West LE, Wen H, Wang EW, et al. Modulation of p53 function by SET8-mediated methylation at lysine 382. *Mol Cell*. 2007;27(4):636-46.
- [308] Chen HZ, Tsai SY, Leone G. Emerging roles of E2Fs in cancer: an exit from cell cycle control. *Nat Rev Cancer*. 2009;9(11):785-97.
- [309] Goodrich DW. The retinoblastoma tumor-suppressor gene, the exception that proves the rule. *Oncogene*. 2006;25(38):5233-43.
- [310] Saddic LA, West LE, Aslanian A, Yates JR, 3rd, Rubin SM, Gozani O, et al. Methylation of the retinoblastoma tumor suppressor by SMYD2. *J Biol Chem*. 2010;285(48):37733-40.
- [311] Carr SM, Munro S, Kessler B, Oppermann U, La Thangue NB. Interplay between lysine methylation and Cdk phosphorylation in growth control by the retinoblastoma protein. *Embo J*. 2011;30(2):317-27.

- [312] Cho HS, Hayami S, Toyokawa G, Maejima K, Yamane Y, Suzuki T, et al. RB1 methylation by SMYD2 enhances cell cycle progression through an increase of RB1 phosphorylation. *Neoplasia*. 2012;14(6):476-86.
- [313] Cho HS, Suzuki T, Dohmae N, Hayami S, Unoki M, Yoshimatsu M, et al. Demethylation of RB regulator MYPT1 by histone demethylase LSD1 promotes cell cycle progression in cancer cells. *Cancer Res*. 2011;71(3):655-60.
- [314] Kontaki H, Talianidis I. Lysine methylation regulates E2F1-induced cell death. *Mol Cell*. 2010;39(1):152-60.
- [315] Sen R, Baltimore D. Multiple nuclear factors interact with the immunoglobulin enhancer sequences. *Cell* 1986. 46: 705-716. *J Immunol*. 2006;177(11):7485-96.
- [316] Silverman N, Maniatis T. NF-kappaB signaling pathways in mammalian and insect innate immunity. *Genes Dev*. 2001;15(18):2321-42.
- [317] Sethi G, Tergaonkar V. Potential pharmacological control of the NF-kappaB pathway. *Trends Pharmacol Sci*. 2009;30(6):313-21.
- [318] Karin M. Nuclear factor-kappaB in cancer development and progression. *Nature*. 2006;441(7092):431-6.
- [319] Vallabhapurapu S, Karin M. Regulation and function of NF-kappaB transcription factors in the immune system. *Annu Rev Immunol*. 2009;27:693-733.
- [320] Li F., Sethi G. Targeting transcription factor NF-kappaB to overcome chemoresistance and radioresistance in cancer therapy. *Biochim Biophys Acta*. 2010;1805(2):167-80.
- [321] Mantovani A. Molecular pathways linking inflammation and cancer. *Curr Mol Med*. 2010;10(4):369-73.
- [322] Wong ET, Tergaonkar V. Roles of NF-kappaB in health and disease: mechanisms and therapeutic potential. *Clin Sci (Lond)*. 2009;116(6):451-65.
- [323] Cheung N, Chan LC, Thompson A, Cleary ML, So CW. Protein arginine-methyltransferase-dependent oncogenesis. *Nat Cell Biol*. 2007;9(10):1208-15.
- [324] Medzhitov R, Horng T. Transcriptional control of the inflammatory response. *Nat Rev Immunol*. 2009;9(10):692-703.
- [325] Lal G, Zhang N, van der Touw W, Ding Y, Ju W, Bottinger EP, et al. Epigenetic regulation of Foxp3 expression in regulatory T cells by DNA methylation. *J Immunol*. 2009;182(1):259-73.
- [326] Wells AD. New insights into the molecular basis of T cell anergy: anergy factors, avoidance sensors, and epigenetic imprinting. *J Immunol*. 2009;182(12):7331-41.
- [327] Wierda RJ, Geutskens SB, Jukema JW, Quax PH, van den Elsen PJ. Epigenetics in atherosclerosis and inflammation. *J Cell Mol Med*. 2010;14(6A):1225-40.
- [328] Chen L, Fischle W, Verdin E, Greene WC. Duration of nuclear NF-kappaB action regulated by reversible acetylation. *Science*. 2001;293(5535):1653-7.
- [329] Chen LF, Greene WC. Regulation of distinct biological activities of the NF-kappaB transcription factor complex by acetylation. *J Mol Med*. 2003;81(9):549-57.
- [330] Yang SR, Valvo S, Yao H, Kode A, Rajendrasozhan S, Edirisinghe I, et al. IKK alpha causes chromatin modification on pro-inflammatory genes by cigarette smoke in mouse lung. *Am J Respir Cell Mol Biol*. 2008;38(6):689-98.
- [331] Anest V, Hanson JL, Cogswell PC, Steinbrecher KA, Strahl BD, Baldwin AS. A nucleosomal function for IkappaB kinase-alpha in NF-kappaB-dependent gene expression. *Nature*. 2003;423(6940):659-63.

- [332] Yamamoto Y, Verma UN, Prajapati S, Kwak YT, Gaynor RB. Histone H3 phosphorylation by IKK- α is critical for cytokine-induced gene expression. *Nature*. 2003;423(6940):655-9.
- [333] Barnes PJ. Role of HDAC2 in the pathophysiology of COPD. *Annu Rev Physiol*. 2009;71:451-64.
- [334] Ito K, Barnes PJ, Adcock IM. Glucocorticoid receptor recruitment of histone deacetylase 2 inhibits interleukin-1 β -induced histone H4 acetylation on lysines 8 and 12. *Mol Cell Biol*. 2000;20(18):6891-903.
- [335] Aguilar-Quesada R, Munoz-Gamez JA, Martin-Oliva D, Peralta-Leal A, Quiles-Perez R, Rodriguez-Vargas JM, et al. Modulation of transcription by PARP-1: consequences in carcinogenesis and inflammation. *Curr Med Chem*. 2007;14(11):1179-87.
- [336] Hassa PO, Buerki C, Lombardi C, Imhof R, Hottiger MO. Transcriptional coactivation of nuclear factor- κ B-dependent gene expression by p300 is regulated by poly(ADP)-ribose polymerase-1. *J Biol Chem*. 2003;278(46):45145-53.
- [337] Hassa PO, Haenni SS, Buerki C, Meier NI, Lane WS, Owen H, et al. Acetylation of poly(ADP-ribose) polymerase-1 by p300/CREB-binding protein regulates coactivation of NF- κ B-dependent transcription. *J Biol Chem*. 2005;280(49):40450-64.
- [338] Aggarwal BB. Nuclear factor- κ B: the enemy within. *Cancer Cell*. 2004;6(3):203-8.
- [339] Li Y, Reddy MA, Miao F, Shanmugam N, Yee JK, Hawkins D, et al. Role of the histone H3 lysine 4 methyltransferase, SET7/9, in the regulation of NF- κ B-dependent inflammatory genes. Relevance to diabetes and inflammation. *J Biol Chem*. 2008;283(39):26771-81.
- [340] Covic M, Hassa PO, Sacconi S, Buerki C, Meier NI, Lombardi C, et al. Arginine methyltransferase CARM1 is a promoter-specific regulator of NF- κ B-dependent gene expression. *Embo J*. 2005;24(1):85-96.
- [341] Hong H, Kao C, Jeng MH, Eble JN, Koch MO, Gardner TA, et al. Aberrant expression of CARM1, a transcriptional coactivator of androgen receptor, in the development of prostate carcinoma and androgen-independent status. *Cancer*. 2004;101(1):83-9.
- [342] Meyer R, Wolf SS, Obendorf M. PRMT2, a member of the protein arginine methyltransferase family, is a coactivator of the androgen receptor. *J Steroid Biochem Mol Biol*. 2007;107(1-2):1-14.
- [343] Brasacchio D, Okabe J, Tikellis C, Balcerczyk A, George P, Baker EK, et al. Hyperglycemia induces a dynamic cooperativity of histone methylase and demethylase enzymes associated with gene-activating epigenetic marks that coexist on the lysine tail. *Diabetes*. 2009;58(5):1229-36.
- [344] El-Osta A, Brasacchio D, Yao D, Poci A, Jones PL, Roeder RG, et al. Transient high glucose causes persistent epigenetic changes and altered gene expression during subsequent normoglycemia. *J Exp Med*. 2008;205(10):2409-17.
- [345] Taylor CT. Interdependent roles for hypoxia inducible factor and nuclear factor- κ B in hypoxic inflammation. *J Physiol*. 2008;586(Pt 17):4055-9.
- [346] Villeneuve LM, Reddy MA, Lanting LL, Wang M, Meng L, Natarajan R. Epigenetic histone H3 lysine 9 methylation in metabolic memory and inflammatory phenotype of vascular smooth muscle cells in diabetes. *Proc Natl Acad Sci U S A*. 2008;105(26):9047-52.

- [347] Huang B, Yang XD, Zhou MM, Ozato K, Chen LF. Brd4 coactivates transcriptional activation of NF-kappaB via specific binding to acetylated RelA. *Mol Cell Biol.* 2009;29(5):1375-87.
- [348] Greene WC, Chen LF. Regulation of NF-kappaB action by reversible acetylation. *Novartis Found Symp.* 2004;259:208-17; discussion 18-25.
- [349] Wu X, Qi J, Bradner JE, Xiao G, Chen LF. Bromodomain and extraterminal (BET) protein inhibition suppresses human T cell leukemia virus 1 (HTLV-1) Tax protein-mediated tumorigenesis by inhibiting nuclear factor kappaB (NF-kappaB) signaling. *J Biol Chem.* 2013;288(50):36094-105.
- [350] Kiernan R, Bres V, Ng RW, Coudart MP, El Messaoudi S, Sardet C, et al. Post-activation turn-off of NF-kappa B-dependent transcription is regulated by acetylation of p65. *J Biol Chem.* 2003;278(4):2758-66.
- [351] Deng WG, Wu KK. Regulation of inducible nitric oxide synthase expression by p300 and p50 acetylation. *J Immunol.* 2003;171(12):6581-8.
- [352] Pradhan M, Baumgarten SC, Bembins LA, Frasor J. CBP mediates NF-kappaB-dependent histone acetylation and estrogen receptor recruitment to an estrogen response element in the BIRC3 promoter. *Mol Cell Biol.* 2012;32(2):569-75.
- [353] Hamamoto R, Saloura V, Nakamura Y. Critical roles of non-histone protein lysine methylation in human tumorigenesis. *Nat Rev Cancer.* 2015;15(2):110-24.
- [354] Stark GR, Wang Y, Lu T. Lysine methylation of promoter-bound transcription factors and relevance to cancer. *Cell Res.* 2011;21(3):375-80.
- [355] Lu T, Yang M, Huang DB, Wei H, Ozer GH, Ghosh G, et al. Role of lysine methylation of NF-kappaB in differential gene regulation. *Proc Natl Acad Sci U S A.* 2013;110(33):13510-5.
- [356] Levy D, Kuo AJ, Chang Y, Schaefer U, Kitson C, Cheung P, et al. Lysine methylation of the NF-kappaB subunit RelA by SETD6 couples activity of the histone methyltransferase GLP at chromatin to tonic repression of NF-kappaB signaling. *Nat Immunol.* 2011;12(1):29-36.
- [357] Yang P, Guo L, Duan ZJ, Tepper CG, Xue L, Chen X, et al. Histone methyltransferase NSD2/MMSET mediates constitutive NF-kappaB signaling for cancer cell proliferation, survival, and tumor growth via a feed-forward loop. *Mol Cell Biol.* 2012;32(15):3121-31.
- [358] Levy DE, Darnell JE, Jr. Stats: transcriptional control and biological impact. *Nat Rev Mol Cell Biol.* 2002;3(9):651-62.
- [359] Icardi L, De Bosscher K, Tavernier J. The HAT/HDAC interplay: multilevel control of STAT signaling. *Cytokine Growth Factor Rev.* 2012;23(6):283-91.
- [360] Zhuang S. Regulation of STAT signaling by acetylation. *Cell Signal.* 2013;25(9):1924-31.
- [361] Ray S, Boldogh I, Brasier AR. STAT3 NH2-terminal acetylation is activated by the hepatic acute-phase response and required for IL-6 induction of angiotensinogen. *Gastroenterology.* 2005;129(5):1616-32.
- [362] Nie Y, Erion DM, Yuan Z, Dietrich M, Shulman GI, Horvath TL, et al. STAT3 inhibition of gluconeogenesis is downregulated by SirT1. *Nat Cell Biol.* 2009;11(4):492-500.
- [363] Ohbayashi N, Ikeda O, Taira N, Yamamoto Y, Muromoto R, Sekine Y, et al. LIF- and IL-6-induced acetylation of STAT3 at Lys-685 through PI3K/Akt activation. *Biol Pharm Bull.* 2007;30(10):1860-4.

- [364] Kramer OH, Baus D, Knauer SK, Stein S, Jager E, Stauber RH, et al. Acetylation of Stat1 modulates NF-kappaB activity. *Genes Dev.* 2006;20(4):473-85.
- [365] Gupta M, Han JJ, Stenson M, Wellik L, Witzig TE. Regulation of STAT3 by histone deacetylase-3 in diffuse large B-cell lymphoma: implications for therapy. *Leukemia.* 2012;26(6):1356-64.
- [366] Martinez-Balbas MA, Bauer UM, Nielsen SJ, Brehm A, Kouzarides T. Regulation of E2F1 activity by acetylation. *Embo J.* 2000;19(4):662-71.
- [367] Kumar BR, Swaminathan V, Banerjee S, Kundu TK. p300-mediated acetylation of human transcriptional coactivator PC4 is inhibited by phosphorylation. *J Biol Chem.* 2001;276(20):16804-9.
- [368] Matsuzaki H, Daitoku H, Hatta M, Aoyama H, Yoshimochi K, Fukamizu A. Acetylation of Foxo1 alters its DNA-binding ability and sensitivity to phosphorylation. *Proc Natl Acad Sci U S A.* 2005;102(32):11278-83.
- [369] Chi H, Flavell RA. Acetylation of MKP-1 and the control of inflammation. *Sci Signal.* 2008;1(41):pe44.
- [370] Chen Y, Wang J, Wang X, Liu X, Li H, Lv Q, et al. STAT3, a Poor Survival Predictor, Is Associated with Lymph Node Metastasis from Breast Cancer. *J Breast Cancer.* 2013;16(1):40-9.
- [371] Yang J, Huang J, Dasgupta M, Sears N, Miyagi M, Wang B, et al. Reversible methylation of promoter-bound STAT3 by histone-modifying enzymes. *Proc Natl Acad Sci U S A.* 2010;107(50):21499-504.
- [372] Kim E, Kim M, Woo DH, Shin Y, Shin J, Chang N, et al. Phosphorylation of EZH2 activates STAT3 signaling via STAT3 methylation and promotes tumorigenicity of glioblastoma stem-like cells. *Cancer Cell.* 2013;23(6):839-52.
- [373] Serman L, Nikuseva Martic T, Serman A, Vranic S. Epigenetic alterations of the Wnt signaling pathway in cancer: a mini review. *Bosn J Basic Med Sci.* 2014;14(4):191-4.
- [374] Surana R, Sikka S, Cai W, Shin EM, Warriar SR, Tan HJ, et al. Secreted frizzled related proteins: Implications in cancers. *Biochim Biophys Acta.* 2014;1845(1):53-65.
- [375] Staal FJ, Clevers HC. WNT signalling and haematopoiesis: a WNT-WNT situation. *Nat Rev Immunol.* 2005;5(1):21-30.
- [376] Gregorieff A, Clevers H. Wnt signaling in the intestinal epithelium: from endoderm to cancer. *Genes Dev.* 2005;19(8):877-90.
- [377] Ying Y, Tao Q. Epigenetic disruption of the WNT/beta-catenin signaling pathway in human cancers. *Epigenetics.* 2009;4(5):307-12.
- [378] Batra S, Shi Y, Kuchenbecker KM, He B, Reguart N, Mikami I, et al. Wnt inhibitory factor-1, a Wnt antagonist, is silenced by promoter hypermethylation in malignant pleural mesothelioma. *Biochem Biophys Res Commun.* 2006;342(4):1228-32.
- [379] Liu TH, Raval A, Chen SS, Matkovic JJ, Byrd JC, Plass C. CpG island methylation and expression of the secreted frizzled-related protein gene family in chronic lymphocytic leukemia. *Cancer Res.* 2006;66(2):653-8.
- [380] Suzuki H, Watkins DN, Jair KW, Schuebel KE, Markowitz SD, Chen WD, et al. Epigenetic inactivation of SFRP genes allows constitutive WNT signaling in colorectal cancer. *Nat Genet.* 2004;36(4):417-22.
- [381] Schiefer L, Visweswaran M, Perumal V, Arfuso F, Groth D, Newsholme P, et al. Epigenetic regulation of the secreted frizzled-related protein family in human glioblastoma multiforme. *Cancer Gene Ther.* 2014;21(7):297-303.

- [382] Katoh M. Expression and regulation of WNT1 in human cancer: up-regulation of WNT1 by beta-estradiol in MCF-7 cells. *Int J Oncol.* 2003;22(1):209-12.
- [383] Huang CL, Liu D, Nakano J, Ishikawa S, Kontani K, Yokomise H, et al. Wnt5a expression is associated with the tumor proliferation and the stromal vascular endothelial growth factor--an expression in non-small-cell lung cancer. *J Clin Oncol.* 2005;23(34):8765-73.
- [384] Sato N, Fukushima N, Maitra A, Matsubayashi H, Yeo CJ, Cameron JL, et al. Discovery of novel targets for aberrant methylation in pancreatic carcinoma using high-throughput microarrays. *Cancer Res.* 2003;63(13):3735-42.
- [385] Shu J, Jelinek J, Chang H, Shen L, Qin T, Chung W, et al. Silencing of bidirectional promoters by DNA methylation in tumorigenesis. *Cancer Res.* 2006;66(10):5077-84.
- [386] Veeck J, Niederacher D, An H, Klopocki E, Wiesmann F, Betz B, et al. Aberrant methylation of the Wnt antagonist SFRP1 in breast cancer is associated with unfavourable prognosis. *Oncogene.* 2006;25(24):3479-88.
- [387] Veeck J, Noetzel E, Bektas N, Jost E, Hartmann A, Knuchel R, et al. Promoter hypermethylation of the SFRP2 gene is a high-frequent alteration and tumor-specific epigenetic marker in human breast cancer. *Mol Cancer.* 2008;7:83.
- [388] Chim CS, Chan WW, Pang A, Kwong YL. Preferential methylation of Wnt inhibitory factor-1 in acute promyelocytic leukemia: an independent poor prognostic factor. *Leukemia.* 2006;20(5):907-9.
- [389] Yu J, Tao Q, Cheng YY, Lee KY, Ng SS, Cheung KF, et al. Promoter methylation of the Wnt/beta-catenin signaling antagonist Dkk-3 is associated with poor survival in gastric cancer. *Cancer.* 2009;115(1):49-60.
- [390] Prasad CP, Mirza S, Sharma G, Prashad R, DattaGupta S, Rath G, et al. Epigenetic alterations of CDH1 and APC genes: relationship with activation of Wnt/beta-catenin pathway in invasive ductal carcinoma of breast. *Life Sci.* 2008;83(9-10):318-25.
- [391] Toyokawa G, Cho HS, Masuda K, Yamane Y, Yoshimatsu M, Hayami S, et al. Histone lysine methyltransferase Wolf-Hirschhorn syndrome candidate 1 is involved in human carcinogenesis through regulation of the Wnt pathway. *Neoplasia.* 2011;13(10):887-98.
- [392] Ge X, Jin Q, Zhang F, Yan T, Zhai Q. PCAF acetylates {beta}-catenin and improves its stability. *Mol Biol Cell.* 2009;20(1):419-27.
- [393] Chu CS, Lo PW, Yeh YH, Hsu PH, Peng SH, Teng YC, et al. O-GlcNAcylation regulates EZH2 protein stability and function. *Proc Natl Acad Sci U S A.* 2014;111(4):1355-60.
- [394] Crea F, Danesi R, Farrar WL. Cancer stem cell epigenetics and chemoresistance. *Epigenomics.* 2009;1(1):63-79.
- [395] Kanno R, Janakiraman H, Kanno M. Epigenetic regulator polycomb group protein complexes control cell fate and cancer. *Cancer Sci.* 2008;99(6):1077-84.
- [396] Ohm JE, McGarvey KM, Yu X, Cheng L, Schuebel KE, Cope L, et al. A stem cell-like chromatin pattern may predispose tumor suppressor genes to DNA hypermethylation and heritable silencing. *Nat Genet.* 2007;39(2):237-42.
- [397] Widschwendter M, Fiegl H, Egle D, Mueller-Holzner E, Spizzo G, Marth C, et al. Epigenetic stem cell signature in cancer. *Nat Genet.* 2007;39(2):157-8.
- [398] Sparmann A, van Lohuizen M. Polycomb silencers control cell fate, development and cancer. *Nat Rev Cancer.* 2006;6(11):846-56.
- [399] Hurt EM, Farrar WL. Cancer stem cells: the seeds of metastasis? *Mol Interv.* 2008;8(3):140-2.

- [400] Chou RH, Yu YL, Hung MC. The roles of EZH2 in cell lineage commitment. *Am J Transl Res*. 2011;3(3):243-50.
- [401] Economopoulou P, Kaklamani VG, Siziopikou K. The role of cancer stem cells in breast cancer initiation and progression: potential cancer stem cell-directed therapies. *Oncologist*. 2012;17(11):1394-401.
- [402] Lee J, Son MJ, Woolard K, Donin NM, Li A, Cheng CH, et al. Epigenetic-mediated dysfunction of the bone morphogenetic protein pathway inhibits differentiation of glioblastoma-initiating cells. *Cancer Cell*. 2008;13(1):69-80.
- [403] Suva ML, Riggi N, Janiszewska M, Radovanovic I, Provero P, Stehle JC, et al. EZH2 is essential for glioblastoma cancer stem cell maintenance. *Cancer Res*. 2009;69(24):9211-8.
- [404] Pei L, Choi JH, Liu J, Lee EJ, McCarthy B, Wilson JM, et al. Genome-wide DNA methylation analysis reveals novel epigenetic changes in chronic lymphocytic leukemia. *Epigenetics*. 2012;7(6):567-78.
- [405] Lu D, Zhao Y, Tawatao R, Cottam HB, Sen M, Leoni LM, et al. Activation of the Wnt signaling pathway in chronic lymphocytic leukemia. *Proc Natl Acad Sci U S A*. 2004;101(9):3118-23.
- [406] Martin V, Agirre X, Jimenez-Velasco A, Jose-Eneriz ES, Cordeu L, Garate L, et al. Methylation status of Wnt signaling pathway genes affects the clinical outcome of Philadelphia-positive acute lymphoblastic leukemia. *Cancer Sci*. 2008;99(9):1865-8.
- [407] Hou HA, Kuo YY, Liu CY, Chou WC, Lee MC, Chen CY, et al. DNMT3A mutations in acute myeloid leukemia: stability during disease evolution and clinical implications. *Blood*. 2012;119(2):559-68.
- [408] Yamazaki J, Estecio MR, Lu Y, Long H, Malouf GG, Graber D, et al. The epigenome of AML stem and progenitor cells. *Epigenetics*. 2013;8(1):92-104.
- [409] Rheinbay E, Suva ML, Gillespie SM, Wakimoto H, Patel AP, Shahid M, et al. An aberrant transcription factor network essential for Wnt signaling and stem cell maintenance in glioblastoma. *Cell Rep*. 2013;3(5):1567-79.
- [410] Shen H, Laird PW. Interplay between the cancer genome and epigenome. *Cell*. 2013;153(1):38-55.
- [411] Lewis PW, Muller MM, Koletsky MS, Cordero F, Lin S, Banaszynski LA, et al. Inhibition of PRC2 activity by a gain-of-function H3 mutation found in pediatric glioblastoma. *Science*. 2013;340(6134):857-61.
- [412] Juo YY, Gong XJ, Mishra A, Cui X, Baylin SB, Azad NS, et al. Epigenetic therapy for solid tumors: from bench science to clinical trials. *Epigenomics*. 2015;7(2):215-35.
- [413] Baylin SB, Jones PA. A decade of exploring the cancer epigenome - biological and translational implications. *Nat Rev Cancer*. 2011;11(10):726-34.
- [414] Redmond KM, Wilson TR, Johnston PG, Longley DB. Resistance mechanisms to cancer chemotherapy. *Front Biosci*. 2008;13:5138-54.
- [415] Sierra JR, Cepero V, Giordano S. Molecular mechanisms of acquired resistance to tyrosine kinase targeted therapy. *Mol Cancer*. 2010;9:75.
- [416] Turke AB, Zejnullahu K, Wu YL, Song Y, Dias-Santagata D, Lifshits E, et al. Preexistence and clonal selection of MET amplification in EGFR mutant NSCLC. *Cancer Cell*. 2010;17(1):77-88.

- [417] Takeuchi A, Shiota M, Tatsugami K, Yokomizo A, Tanaka S, Kuroiwa K, et al. p300 mediates cellular resistance to doxorubicin in bladder cancer. *Mol Med Rep.* 2012;5(1):173-6.
- [418] Lee SO, Chun JY, Nadiminty N, Lou W, Feng S, Gao AC. Interleukin-4 activates androgen receptor through CBP/p300. *Prostate.* 2009;69(2):126-32.
- [419] Wei SH, Brown R, Huang TH. Aberrant DNA methylation in ovarian cancer: is there an epigenetic predisposition to drug response? *Ann N Y Acad Sci.* 2003;983:243-50.
- [420] Gifford G, Paul J, Vasey PA, Kaye SB, Brown R. The acquisition of hMLH1 methylation in plasma DNA after chemotherapy predicts poor survival for ovarian cancer patients. *Clin Cancer Res.* 2004;10(13):4420-6.
- [421] Plumb JA, Strathdee G, Sludden J, Kaye SB, Brown R. Reversal of drug resistance in human tumor xenografts by 2'-deoxy-5-azacytidine-induced demethylation of the hMLH1 gene promoter. *Cancer Res.* 2000;60(21):6039-44.
- [422] Strathdee G, MacKean MJ, Illand M, Brown R. A role for methylation of the hMLH1 promoter in loss of hMLH1 expression and drug resistance in ovarian cancer. *Oncogene.* 1999;18(14):2335-41.
- [423] Zeller C, Dai W, Steele NL, Siddiq A, Walley AJ, Wilhelm-Benartzi CS, et al. Candidate DNA methylation drivers of acquired cisplatin resistance in ovarian cancer identified by methylome and expression profiling. *Oncogene.* 2012;31(42):4567-76.
- [424] Glasspool RM, Teodoridis JM, Brown R. Epigenetics as a mechanism driving polygenic clinical drug resistance. *Br J Cancer.* 2006;94(8):1087-92.
- [425] Balch C, Huang TH, Brown R, Nephew KP. The epigenetics of ovarian cancer drug resistance and resensitization. *Am J Obstet Gynecol.* 2004;191(5):1552-72.
- [426] Miyamoto N, Izumi H, Noguchi T, Nakajima Y, Ohmiya Y, Shiota M, et al. Tip60 is regulated by circadian transcription factor clock and is involved in cisplatin resistance. *J Biol Chem.* 2008;283(26):18218-26.
- [427] Tada Y, Wada M, Kuroiwa K, Kinugawa N, Harada T, Nagayama J, et al. MDR1 gene overexpression and altered degree of methylation at the promoter region in bladder cancer during chemotherapeutic treatment. *Clin Cancer Res.* 2000;6(12):4618-27.
- [428] Nakayama M, Wada M, Harada T, Nagayama J, Kusaba H, Ohshima K, et al. Hypomethylation status of CpG sites at the promoter region and overexpression of the human MDR1 gene in acute myeloid leukemias. *Blood.* 1998;92(11):4296-307.
- [429] Stronach EA, Alfraidi A, Rama N, Datler C, Studd JB, Agarwal R, et al. HDAC4-regulated STAT1 activation mediates platinum resistance in ovarian cancer. *Cancer Res.* 2011;71(13):4412-22.
- [430] Hegi ME, Diserens AC, Gorlia T, Hamou MF, de Tribolet N, Weller M, et al. MGMT gene silencing and benefit from temozolomide in glioblastoma. *N Engl J Med.* 2005;352(10):997-1003.
- [431] Hegi ME, Liu L, Herman JG, Stupp R, Wick W, Weller M, et al. Correlation of O6-methylguanine methyltransferase (MGMT) promoter methylation with clinical outcomes in glioblastoma and clinical strategies to modulate MGMT activity. *J Clin Oncol.* 2008;26(25):4189-99.
- [432] Esteller M, Hamilton SR, Burger PC, Baylin SB, Herman JG. Inactivation of the DNA repair gene O6-methylguanine-DNA methyltransferase by promoter hypermethylation is a common event in primary human neoplasia. *Cancer Res.* 1999;59(4):793-7.

- [433] Esteller M, Garcia-Foncillas J, Andion E, Goodman SN, Hidalgo OF, Vanaclocha V, et al. Inactivation of the DNA-repair gene MGMT and the clinical response of gliomas to alkylating agents. *N Engl J Med*. 2000;343(19):1350-4.
- [434] Zhang W, Peyton M, Xie Y, Soh J, Minna JD, Gazdar AF, et al. Histone deacetylase inhibitor romidepsin enhances anti-tumor effect of erlotinib in non-small cell lung cancer (NSCLC) cell lines. *J Thorac Oncol*. 2009;4(2):161-6.
- [435] Witta S. Histone deacetylase inhibitors in non-small-cell lung cancer. *J Thorac Oncol*. 2012;7(16 Suppl 5):S404-6.
- [436] Schenk T, Chen WC, Gollner S, Howell L, Jin L, Hebestreit K, et al. Inhibition of the LSD1 (KDM1A) demethylase reactivates the all-trans-retinoic acid differentiation pathway in acute myeloid leukemia. *Nat Med*. 2012;18(4):605-11.
- [437] Lokken AA, Zeleznik-Le NJ. Breaking the LSD1/KDM1A addiction: therapeutic targeting of the epigenetic modifier in AML. *Cancer Cell*. 2012;21(4):451-3.
- [438] West AC, Johnstone RW. New and emerging HDAC inhibitors for cancer treatment. *J Clin Invest*. 2014;124(1):30-9.
- [439] Bojang P, Jr., Ramos KS. The promise and failures of epigenetic therapies for cancer treatment. *Cancer Treat Rev*. 2014;40(1):153-69.
- [440] Fenichel MP. FDA approves new agent for multiple myeloma. *J Natl Cancer Inst*. 2015;107(6):djv165.
- [441] Hamm CA, Costa FF. Epigenomes as therapeutic targets. *Pharmacol Ther*. 2015;151:72-86.
- [442] Rodriguez-Paredes M, Esteller M. Cancer epigenetics reaches mainstream oncology. *Nat Med*. 2011;17(3):330-9.
- [443] Shanmugam MK, Sethi G. Role of Epigenetics in Inflammation-Associated Diseases. In: Kundu TK, editor. *Epigenetics: Development and Disease. Subcellular Biochemistry*. Heidelberg New York London: Springer Dordrecht; 2012. p. 627-58.

Chapter 62

ALZHEIMER'S DISEASE: AN INSIGHT FROM EPIGENETIC PERSPECTIVE

Yildiz Dincer^{1,*} and Zeynep Sezgin^{1,2}

¹Istanbul University, Cerrahpasa Medical Faculty,
Department of Medical Biochemistry, Istanbul, Turkey

²Yeni Yuzyıl University Medical Faculty,
Department of Medical Biochemistry, Istanbul, Turkey

ABSTRACT

Alzheimer's Disease is one of the most common neurodegenerative disorders. Many efforts have been directed to prevent AD due to its rising prevalence and the lack of an effective curative treatment. Epigenetic changes are involved in regulation of gene expression, and may mediate various pathologies. Epigenetic changes are reversible so that can be easily modulated. Modulation of dysregulated epigenetic mechanisms is a promising therapeutic approach for many diseases. There is some evidence for epigenetic dysregulation at various levels contributing to AD pathogenesis. Despite the recent rapid accumulation of knowledge about AD pathogenesis, the role of epigenetic modifications has not been understood exactly. This chapter provides a brief overview about the role of epigenetic changes that are linked to AD pathogenesis and emerging targets for new therapeutic strategies in this field.

Keywords: Alzheimer's disease, Epigenetic modifications, DNA methylation, Histone acetylation, Sirtuins, MicroRNAs

ABBREVIATIONS

ABCA7	ATP-binding cassette subfamily A, member 7
ACE	angiotensin-converting enzyme

* Corresponding Author's Email: yldz.dincer@gmail.com.

AD	Alzheimer's disease
AGE	advanced glycation endproducts
AICD	APP intracellular domain
anti-TNF- α	antitumor necrosis factor- α
APP	amyloid precursor protein
A β	amyloid β -peptid
BACE1/2	β -site APP cleaving enzyme 1/2
BBB	blood-brain barrier
BIN1	bridging integrator 1
CAT	catalase
CD2AP	CD2-associated protein
COX-1/2	cyclooxygenase-1
CR	calorie restriction
CR-1	complement receptor 1
DNMTs	DNA methyltransferases
ECE-1	endothelin-converting enzyme 1
EPHA1	EPH receptor A1
FOXO	forkhead box protein O
GSK-3 β	Glycogen synthase kinase 3 β
HATs	histone acetyltransferases
HDACs	histone deacetylases
IDE	insulin-degrading enzyme
LEARn	latent early-life associated regulation
LRP1 IL-1 β	interleukin-1 β lipoprotein receptor-related protein 1
MCI	Mild Cognitive Impairment
MDA	malondialdehyde
miRNAs	microRNAs
NAD ⁺	nicotinamide adenine dinucleotide
NCT	nicastatin
NEP	neprilysin
NF- κ B	nuclear factor κ -light chain enhancer of activated B cells
NSAIDs	nonsteroidal anti-inflammatory drugs
PBMCs	peripheral blood mononuclear cells
PICALM	phosphatidylinositol binding clathrin assembly protein
PGC-1 α	peroxisome proliferator-activated reseptor- γ coactivator-1 α
PPAR- γ	peroxisome proliferator-activated reseptor- γ
PS1/2	presenilin 1/2
ROS	reactive oxygen species
SAH	S-adenosylhomocysteine
SAM	S-adenosyl methionine
Sir2	silent information regulator factor 2
SIRTs	sirtuins
SOD	superoxide dismutase
3xTg-AD	triple transgenic animal model of AD
8-OHdG	8-hydroxydeoxyguanosine

1. INTRODUCTION

Alzheimer's disease (AD), originally described by the German neurologist Alois Alzheimer in 1907, is the most common form of dementia in elderly people. AD begins with mild memory loss and then progress with severe cognitive impairment and functional decline. The major pathological hallmarks of AD are accumulation of extracellular amyloid plaques and formation of intracellular neurofibrillary tangles. Amyloid plaques are mainly constituted by amyloid β -peptid ($A\beta$) that occurs primarily in 40mer and 42mer forms. Intracellular neurofibrillary tangles contain hyperphosphorylated tau that is a microtubule-associated protein. Extensive loss of synapses and neurons in the temporal and frontal cortices and hippocampi are also hallmarks of the disease. The pathogenesis of AD is highly complex and involves genetic, epigenetic and physiological alterations in the brain that are still not completely understood. Accumulation of $A\beta$ is neurotoxic and leads to neural, synaptic and cognitive dysfunction. In general, AD can be categorized into two types: early-onset familial AD that is seen in patients younger than 65 years, and late-onset sporadic AD that is seen in patients older than 65 years. Both genetic and environmental factors can contribute to development of these AD forms. Early-onset familial AD constitutes 5% of all AD cases and can be explained genetically by mutations in the amyloid precursor protein (APP) gene or in the genes encoding APP processing enzymes. The late-onset sporadic AD (95% of all AD cases) is likely caused by $A\beta$ overproduction and/or dysfunctions in $A\beta$ solubility or aggregation, endocytosis, degradation, transcytosis and removal [1]. With the increase in life expectancy, the prevalence of AD is going up rapidly worldwide. The number of affected individuals has been estimated as more than 100 million by the year 2050 due to the aging of the population (<http://www.alz.co.uk/research/statistics>, http://www.alz.org/downloads/facts_figures_2012.pdf). Progressive loss of memory and other cognitive functions in AD and the lack of effective treatment options to prevent its progression result in a high burden on social and economic resources. Aging is the greatest risk factor for sporadic forms of AD. However, epidemiological studies indicate that genetic alterations and polymorphisms, abnormal immune or inflammatory responses, low educational level, smoking, physical inactivity, depression, midlife hypertension, high cholesterol level, diabetes, APO ϵ 4 allele positivity, traumatic head injury, oxidative stress, certain drugs and hormone replacement are contributing factors for AD. The disruption of epigenetic mechanisms which control expression of AD-related genes is recently discovered another contributory factor. Despite the tremendous progress within the last a few decades in our understanding of AD pathogenesis, the accurate molecular and epigenetic mechanisms of AD have not been revealed completely.

2. PATHOGENESIS OF ALZHEIMER'S DISEASE

Cellular mechanisms provoking anomalies in AD are still under clarification, but $A\beta$ deposition, hyperphosphorylated tau protein, neuroinflammation, oxidative stress, metal dishomeostasis, mitochondrial dysfunction, certain genetic mutations and presence of APO ϵ 4 allele have been shown to play key role in AD pathogenesis. Two main proteins take play a major role in the pathogenesis of AD: $A\beta$ and tau.

2.1. A β Hypothesis

According to this hypothesis, A β and its aggregates are responsible for neurodegeneration and dementia in AD. A β peptides are natural products of brain metabolism, but the balance between A β production and clearance is disrupted in AD. A β is generated by the proteolytic cleavage of APP which is synthesized in ribosomes on endoplasmic reticulum and then transported to the golgi apparatus. APP is a transmembrane protein which is cleaved by a series of secretases: α then γ , or β then γ for processing. Aberrant cleavage of APP leads to A β overproduction and neuritic plaque formation. Proteolytic processing of APP occurs by one of two pathways, the non-amyloidogenic or the amyloidogenic (Figure 1). For *non-amyloidogenic cleavage* of APP, α -secretase cleaves within the A β encoding region, releasing a soluble fragment, sAPP α , into the extracellular space. sAPP α was shown to have neurotrophic and neuroprotective properties [2]. After this first cleavage the remaining C-terminal fragment which includes 83 amino acid residues (C-83) is subsequently cleaved by γ -secretase within the transmembrane region, generating p3 (A β 17-40) and APP intracellular domain (AICD). AICD and p3 (3 kDa) peptide are released into cytosol and extracellular space, respectively. AICD potentially translocates to the nucleus and functions as a transcription factor [3, 4]. As for *amyloidogenic cleavage*, APP is cleaved by β -secretase and then γ -secretase. β -secretases include β -site APP cleaving enzyme 1 (BACE1) and 2 (BACE2). After the cleavage of APP by β -secretases, sAPP β fragment is released whereas remaining C-terminal fragment of 99 amino acid residues (C-99) is still membrane bound. C-99 is further cleaved by γ -secretase, generating AICD and an insoluble fragment, A β , which contains 38–43 amino acid residues. Several peptides in varying lengths can be generated from the cleavages by β and γ -secretases. Majority of these A β fragments is A β 40. The another fragment, A β 42, is less common but more fibrillogenic. It forms insoluble aggregates. A β accumulation in the brain induces oxidative stress and inflammatory response thus leads to neurotoxicity which contributes to impairment of cognitive functions.

The balance between different secretase activities is very important in the maintenance of the physiologic A β levels. Decreased processing of APP into A β has been found to be correlated with decreased activity of BACE1 [5]. In AD, accumulation of A β has been suggested to occur prior to tau pathology, and A β aggregates may be one of a cascade of pathological events ranging from synaptic dysfunction, oxidative stress, mitochondrial dysfunction, loss of calcium regulation and inflammation leading to tau hyperphosphorylation [6]. It has been thought that increased production of A β and/or decreased degradation by A β -degrading enzymes, or reduced clearance across the blood-brain barrier may cause aggregation and accumulation of A β in the brain. Impairment of synaptic functions and related signaling pathways by A β oligomers result in changes in neuronal activities and cause to release of neurotoxic mediators from glial cells.

The degradation and clearance of A β from the brain has been reported to be impaired in patients with AD. A β clearance happens via transporting across the blood-brain barrier (BBB) and metabolic degradation. A β can be transported across the BBB into the blood or cerebrospinal fluid by low-density lipoprotein receptor-related protein 1 (LRP1) [7], AGE advanced glycation endproducts) receptor, ApoE, and β -2-macroglobulin [8, 9]. ApoE binds extracellular A β and forms ApoE/A β complex which is cleared by low-density lipoprotein receptor-related protein (LRP) via the endosomal/ lysosomal pathway [10]. In vitro and in vivo studies have shown that A β undergoes proteolytic degradation by A β -degrading enzymes such

as neprilysin (NEP), insulin-degrading enzyme (IDE), endothelin-converting enzyme (ECE)-1, angiotensin-converting enzyme (ACE), uPA/ tPA-plasmin system, cathepsin D, gelatinase-A and -B, matrix metalloproteinase-9, coagulation factor XIa, antibody light chain c23.5 and hk14, and α 2-macroglobulin complexes [11]. Thus, both disruption of BBB or decreased activities of these A β -degrading enzymes may contribute to AD pathogenesis.

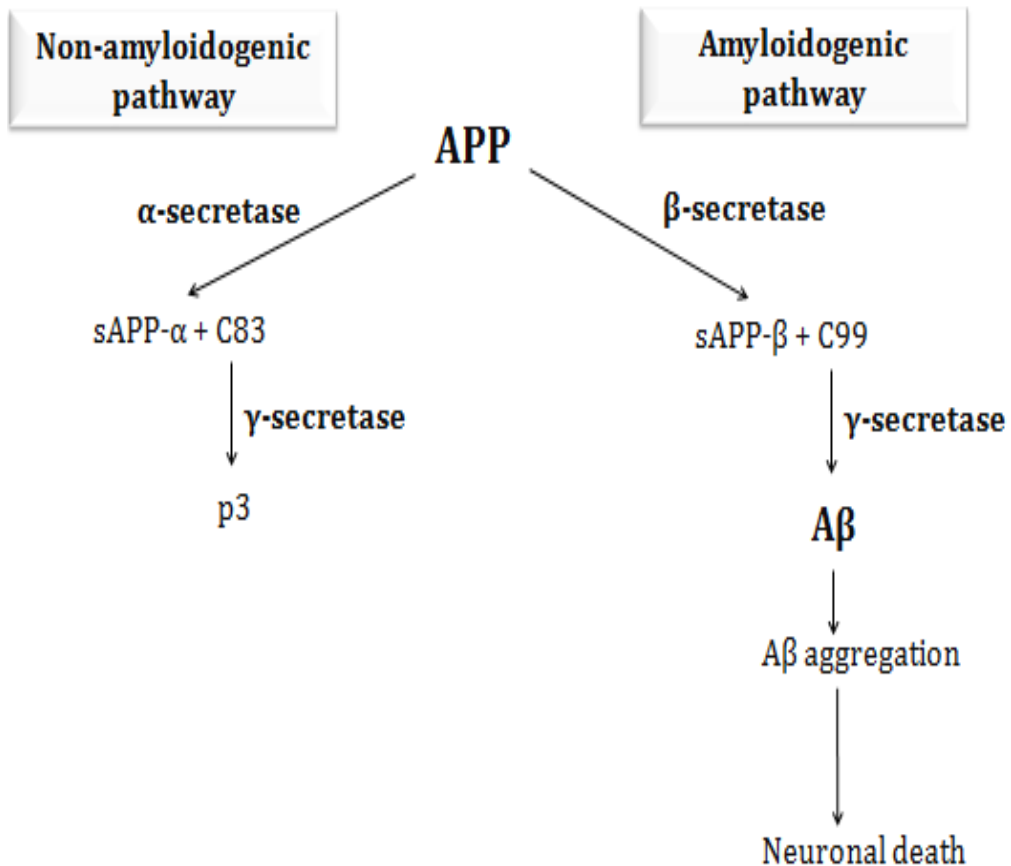


Figure 1. Processing of beta amyloid precursor protein.

2.2. Tau Hypothesis

Tau is an intracellular protein and belongs to a family of microtubule-associated proteins that normally promotes microtubule assembly and stabilization. Tau protein exerts neurotoxic effects when it is hyperphosphorylated. Hyperphosphorylation causes an impairment in its binding ability with tubulin and subsequently results in loss of its normal function. Hyperphosphorylated tau has reduced capacity to promote microtubule assembly. In addition, hyperphosphorylation promotes its polymerization. It aggregates into filaments itself forming neurofibrillary tangles and causes the disorders known as tauopathies [12]. Abnormal hyperphosphorylation of tau is the key player of AD neurodegeneration as the major component of neurofibrillary tangles in AD. Abnormally hyperphosphorylated tau has been isolated from AD brain in 1990s [13]. In vitro studies have shown that tau phosphorylation at different sites

may cause different effects in its biological function as well in pathogenic role. Phosphorylation of tau at Ser262, Thr231, and Ser235 has been shown to inhibit its binding to microtubules by ~35%, ~25%, and ~10%, respectively [14].

Synaptic loss

Synaptic loss is emerged as an early event in the pathogenesis of AD that is confirmed in rodent brain after acute injection of soluble A β oligomers as well in APP transgenic mouse models [15]. Recent gene array investigations have consistently demonstrated that expression levels of neurotransmitter receptor genes and the genes involved in synaptic vesicle trafficking/release, receptor trafficking, postsynaptic density scaffolding, cell adhesion regulating synaptic stability and neuromodulatory systems are altered in early stage of AD [16-19]. Axonal transport defects, oxidative stress, mitochondrial dysfunction, neuroinflammation and possibly A β accumulation in the brain might result in synaptic loss that is observed in AD patients [20].

2.3. Oxidative Stress in Alzheimer's Disease

Oxidative stress is a state that occurs when there is an imbalance between production and clearance of reactive oxygen species (ROS) such as superoxide anion, hydrogen peroxide and hydroxyl radical. ROS have unpaired electrons in separate orbitals in *their* outer *shell* and exhibit a high reactivity. ROS are produced as a byproduct of the normal oxygen metabolism and they have important role in cell signaling and homeostasis. However, in the case of environmental stress such as UV, ionizing radiation, certain drugs and exposure of certain chemicals ROS levels can increase dramatically causing oxidative stress. Interacting with nucleic acids, proteins and lipids, ROS cause structural and functional defects in these molecules. Transition metal ions such as iron and copper act as pro-oxidant. In the abundant presence of these metal ions, ROS production increases via Fenton and Haber-Weiss reactions. There is consistent evidence for presence of extensive oxidative stress in AD brains [21, 22]. Different studies have revealed that the levels of protein carbonyls and 3-nitrotyrosine (protein oxidation markers), 8-hydroxydeoxyguanosine (8-OHdG) and 8-hydroxyguanosine (DNA and RNA oxidation markers), malondialdehyde (MDA), 4-hydroxynonenal, and F2-isoprostanes (lipid peroxidation markers) are elevated in AD brains [23, 24]. Increased production of hydrogen peroxide and nitric oxide, elevated levels of oxidatively modified proteins and lipids have been shown in transgenic mouse models carrying mutations in APP and presenilin genes (PS1 and PS2) [25].

Increased oxidative stress determined in AD is attributed the pro-oxidant role of A β . A β possesses pro-oxidant properties that has an influence on oxidative processes. At high micromolar and nanomolar concentrations, A β aggregates to form fibrils and induces ROS formation, lipid peroxidation, protein oxidation and DNA damage, and subsequently neuronal death. Transition metals including iron and copper are required for its pro-oxidant activity. Direct interaction between A β and transition metal ions may result in increased ROS formation [26]. In an early study, A β treatment has been shown to increase the levels of hydrogen peroxide and lipid peroxides in cell models [27]. Similar results have also been obtained in animal studies. It has been reported that levels of 4-hydroxy-2-nonenal and isoprostanooids, carbonylic proteins are increased in animal models of AD [28]. On the other hand, as a closed circuit,

oxidative stress also promotes A β production by decreasing α -secretase activity while promoting β - and γ -secretase expression and activity [29].

On the other hand, human organism is equipped with efficient antioxidant defence mechanisms. Harmful effects of ROS can be prevented by antioxidant defence system that is constituted by small antioxidant molecules and antioxidant enzymes such as superoxide dismutase (SOD), catalase (CAT) and glutathione peroxidase. Impaired antioxidant defence is the most important factor for oxidative stress. Defective antioxidant defence, elevated oxidative stress and increased A β deposition have been shown in APP overexpressing transgenic mice [30, 31]. Dumont et al. have demonstrated that facilitation of the mitochondrial antioxidant defence improves resistance to A β , decreases plaque formation or increases plaque degradation in a transgenic AD mouse model [31]. Recently, in accordance with Dumont et al. [31], possible role of SOD in AD has been demonstrated in Tg2576 APP-overexpressing AD mouse model. In this model cytoplasmic copper/zinc superoxide dismutase deletion has been found to be associated with increased A β oligomerization, accelerated loss of memory and spatial learning [32]. Aging is a natural and inevitable process and is associated with an prolonged ROS production. Due to the decline in the normal antioxidant defence mechanisms with aging, human body is insufficient in response to ROS-mediated damage [33]. "Free-radical theory by Harman (1956) [34] proposes that ROS production via aerobic respiration increases with age and causes cumulative damage on cell components during the life time. As respect with the fact that AD is seen in elderly individuals, increased oxidative stress in AD patients may be partially attributed to advanced age.

Oxidative stress may also influence hyperphosphorylation and polymerization of tau. It has been suggested that the elevated amounts of fatty acid oxidation that is demonstrated in AD patients, can facilitate the polymerization of tau [35].

Inflammation is a ROS producing state. Neuroinflammation is a contributory factor for AD that will be described below. Production of ROS by glial cells due to inflammation may exacerbate the oxidative stress in AD patients.

2.4. Metal Dyshomeostasis in Alzheimer's Disease

Metal homeostasis is altered in AD brain. Metal ions such as copper (Cu), zinc (Zn) and iron (Fe) are essential for neuronal function. As respect with the fact that metal dyshomeostasis contributes to oxidative stress, disrupted homeostasis of these metal ions in brain are thought to underlie the pathogenesis of AD [36]. Metal dyshomeostasis in AD has attracted the interest of researchers, and reliable results have been obtained. Copper, zinc and iron trigger signaling cascades that mediate cellular physiology by binding to A β . APP and its A β fragments have specific binding sites for Cu²⁺ ions [37]. In vitro studies have demonstrated that A β -Cu complexes catalytically produce hydrogen peroxide and hydroxyl radical and eventually induce oxidative damage [38]. There is growing evidence that Cu can facilitate A β aggregation by binding to A β which results in A β plaque burden [39].

Iron homeostasis is controlled by transferrin and ferritin. Fe, ferritin and transferrin levels have been found to be altered in the hippocampus and cerebral cortex of AD brains [40]. Iron accumulates within senile plaques and neurofibrillary tangles of AD brains. Excess iron leads to increased production of ROS and finally causes oxidative stress. Oxidation of cellular lipids

and proteins results in synaptic dysfunction whereas oxidative damage to DNA induces apoptotic pathways and neuronal death.

Zinc may accelerate the precipitation of A β and produce protease-resistant "non-structured" aggregates [41]. It has been suggested that compounds affecting zinc homeostasis can decrease A β deposition in the brain [42]. Deregulation of neuronal Zn²⁺ homeostasis is believed to be strictly connected to oxidative stress. Zn²⁺ can trigger ROS production by interfering with the activity of the electron transport chain (inhibits complex I and III) [43, 44].

On the other hand, zinc and iron promote tau phosphorylation and aggregation by binding to tau [45]. In consistent with these data, treatment with metal chelators such as clioquinol and desferrioxamine have provided some success in altering the progression of AD [46, 47]. The development of therapeutic interventions targeted to restore metal balance is an active area of current investigations in the field of AD.

2.5. Mitochondrial Dysfunction in Alzheimer's Disease

Mitochondrial dysfunction plays important role in both brain aging and AD. Structural and metabolic changes have been observed in mitochondria of neurons in AD brain. Mitochondrial dysfunction is an early event in AD. Mitochondrial abnormalities, increased mitochondrial degradation by autophagy, and decreased cytochrome oxidase (complex IV) activity that is associated with increased ROS production have been determined in AD brains [48]. Deficient energy production and impaired metabolic activity have been shown in neurons of AD brain and this situation has been explained by reduction in the number and size of mitochondria [49]. The mitochondria is the major source of ROS in the central nervous system, and is also particularly vulnerable to oxidative stress. Oxidation of mitochondrial DNA renders it vulnerable to somatic mutations which leads to initiation of erroneous APP processing [50]. Moreover, A β deposits can contribute to more mitochondrial damage by disrupting lipid polarity and protein mobility in mitochondrial membranes, and inhibiting key enzymes of the mitochondrial respiratory chain [51, 52].

2.6. Neuroinflammation in Alzheimer's Disease

Neuroinflammation plays a prominent and early role in AD pathogenesis. Neuroinflammation is a response to an inflammatory stimuli comprising over activation of microglia and subsequent over expression of proinflammatory cytokines. Although inflammatory response in AD is useful for restoration of tissue integrity, it becomes harmful when chronically induced. Production of pro-inflammatory cytokines, prostoglandins and ROS increases in chronic inflammatory state which in turn leads to aggravation of A β deposition and induction of neuronal dysfunction. The another link between neuroinflammation and A β deposition is BACE1 that is involved in amyloidogenic processing of APP. BACE1 gene has been shown to have four putative NF κ B binding elements in its promoter region, thus its expression is upregulated upon inflammatory stimuli leading to increased production of A β [53]. Furthermore, the accumulation of A β in plaques may produce sequential inflammatory events and excitotoxicity which in turn cause neurodegeneration and cognitive impairment. Microglia activation has been suggested to occur in mild and early stage of AD pathogenesis.

Vom Berg et al. [54] have demonstrated that suppression of expression and signaling of interleukin-12 and interleukin-23 leads to decreases in microglial activity, A β level and A β plaque burden in a mouse model of AD.

It has been shown that suppression of neuroinflammation is concordant with reduction in A β plaques and hyperphosphorylated tau in brain, and is associated with improvements in cognitive and behavioral deficits in AD mouse models [55, 56]. Piro et al. [57] have reported that genetic inactivation of monoacyl-glycerol lipase, an enzyme that hydrolyzes endocannabinoids to generate the primary arachidonic acid pool for neuroinflammatory prostaglandins, controls arachidonic acid release, reduces neuroinflammation and lowers A β levels and plaques in an AD mouse model of A β deposition.

Due to the role of neuroinflammatory process in AD pathogenesis, reducing neuroinflammation by inflammation-based therapeutic strategies can be used in the prevention and treatment of AD. Nonsteroidal anti-inflammatory drugs (NSAIDs) and peroxisome proliferator-activated receptor- γ (PPAR- γ) agonists have been shown to reduce the expression and activity of β -secretase and lower A β secretion [58]. NSAIDs inhibit the enzymatic activity of cyclooxygenase-1 (COX-1) and inducible COX-2 which catalyze the first committed step in the synthesis of prostaglandins. COX inhibitors can decrease levels of highly amyloidogenic A β 1-42 peptide. Suppression of neuroinflammation with NSAIDs has been suggested to rescue memory and cognitive decline. However, retrospective epidemiological studies have demonstrated that prolonged treatment with NSAIDs delays onset of AD when initiated early stage or before disease initiation [59], but its efficiency has not been shown in neither mild nor moderate forms of AD [60]. On the other hand, treatment with antitumor necrosis factor- α (anti-TNF- α) or interleukin-1 β (IL-1 β) antibodies is efficient in reducing the pathology in animal models [61, 62].

2.7. Mutations and Polymorphisms in Alzheimer's Disease

Mutations in several disease-associated genes such as APP, PS1 and PS2 all of which appear to be involved in APP processing have been shown to cause early-onset autosomal dominant AD [63]. Although it seems unlikely to be significant in late-onset AD susceptibility [64] some investigators have suggested the possibility that the rare variants in these three genes may contribute to the risk for late-onset AD [65]. Heterozygous missense mutations in or near A β -coding exons 16 and 17 are the most frequently detected mutations in APP [66]. Whole-gene duplications [67, 68], rare recessive small deletions [69] and recessive missense mutations [70] have also been identified in early onset AD. These mutations result in altered A β production, changed A β ₄₂/A β ₄₀ ratio and increased fibril formation. Mutations in PS-1 and PS-2 also cause to increased production of more aggregative A β ₄₂ by impairing the γ -secretase-mediated cleavage of APP [71, 72]. Recently large-scale genome-wide association studies have changed the face of the complex genetics of AD, and these data have disclosed the novel findings about the candidate risk genes for the development of AD. According to findings of these studies, single nucleotide polymorphisms in or near complement receptor 1 (CR1), phosphatidylinositol binding clathrin assembly protein (PICALM), bridging integrator 1 (BIN1), CD2-associated protein (CD2AP), CD33, EPH receptor A 1 (EPHA1) and ATP-binding cassette subfamily A, member 7 (ABCA7) may increase the risk for AD by affecting APP processing, A β pathway, synaptic cell functioning, and immune system. [73-75].

Apolipoprotein E (APOE) is a polymorphic glycoprotein which contains 299-amino acid. It is highly expressed in the brain and mainly secreted by astrocytes. APOE exists in three isoforms (E2, E3, and E4) in humans encoded by the APOE epsilon (ϵ)2, ϵ 3 and ϵ 4 related genes [76]. Aside from its well known role as a lipid-transporting entity, APOE is involved in neurogenesis, synaptic integrity, plasticity and repair [77]. Although the APO ϵ 4 allele is represented in approximately 25% of population, it has been postulated to be a risk factor for sporadic AD in older adulthood [78]. The influence of APO ϵ 4-allele on development, progression, clinical presentation of AD and rate of cognitive decline have not been fully described. Cross-sectional studies in patients with AD have shown that the APO ϵ 4 allele is associated with impaired memory function, more severe atrophic changes and hypometabolism in the medial temporal lobe structures [79-82]. Autopsy studies on AD have indicated the APOE ϵ 4 allele to be associated with greater densities of A β deposition, senile plaques and neurofibrillary tangles, especially in the hippocampus [83, 84]. Besides this, the ϵ 2 allele may have a protective effect against AD, and carriers of ϵ 2 allele are associated with less amyloid- β plaques and neurofibrillary tangles than both ϵ 4 carriers and ϵ 3 homozygotes [85, 86].

Although presence of APO ϵ 4 allele has been suggested to be associated with faster cognitive decline by some studies, there are also opposing data. Moreover, some studies have reported a converse effect of APO ϵ 4 on the rate of cognitive decline [87-92]. These contradictory data for the influence of APO ϵ 4 allele on AD development might provide clear support for an antagonistic pleiotrophy hypothesis that the APO ϵ 4 allele confers some cognitive advantage in early life despite adverse consequences in old age [93].

3. EPIGENETIC CHANGES IN ALZHEIMER'S DISEASE

Epigenetic modifications regulate gene expression without altering the DNA sequence. Epigenetic changes are dynamic and unlike genetic mutations, they can be reversed by environmental factors and therapeutics. These modifications that can occur at specific gene loci in specific cells result in specific cellular phenotypes. Environmental factors can induce epigenetic changes in individuals in each period of life and these factors are becoming increasingly relevant with respect to the shift from healthy to diseased state [94]. Physical and behavioral factors, nutrition, lifestyle, physical and mental exercise, environmental pollutants, pesticides, chemicals and certain metals affect the epigenetic mechanisms. Especially, nutrition is one of the main epigenetic regulators. It readily modifies epigenetic pathways. Dietary methyl donors involved in one-carbon metabolism such as folate, vitamin B12, betain, choline and methionine play a pivotal role in DNA methylation as well in maintenance of genomic stability. In addition to their antioxidant properties, dietary polyphenols have been shown to modify the epigenome. There is an overwhelming scientific consensus that epigenetic changes contribute to aging and aging-associated diseases. Recently, the diet containing nutrients that are able to modify epigenetic mechanisms has been suggested as a preventive approach for AD [95].

Studies have indicated that genetic mutations are involved in the pathogenesis of early-onset (familial) AD (around 3% to 5% of cases) but epigenetic mechanisms are likely to be important in the etiology of sporadic form of AD. Epigenetic modulations mediated by DNA methylation, histone modification and non-coding RNAs may explain the complex pathogenic

mechanisms in AD. These mechanisms mediate genome-environment interactions by transforming variable combinations of genetic and environmental factors into long-term adaptive changes in gene expression.

In a recent study, expression of *BACE1* gene has been examined in both triple transgenic animal model of AD (3xTg-AD mice) and peripheral blood mononuclear cells (PBMCs) from patients with AD or Mild Cognitive Impairment (MCI). According to findings of this study; *BACE1* mRNA level is increased in brain of aged 3xTg-AD mice, both in the cortex and hippocampus, the promoter region of the *BACE1* is significantly more acetylated without significant alterations in the promoter accessibility, *BACE1* mRNA level is significantly increased in AD patients whereas it is slightly elevated in MCI patients in comparison to controls subjects. Furthermore, as consistent with a higher transcription rate of the gene, *BACE1* promoter is significantly more accessible in AD patients but chromatin in the *BACE1* promoter is closed in MCI patients [96].

γ -Secretase is a multiprotein complex that is an intramembrane-cleaving protease involved in APP processing. Nicastrin (NCT) is one of the integral components of γ -Secretase and is essential for stabilization of the enzyme [97]. Overexpression of NCT is related with increased A β production [98]. Downregulation of NCT gene has been reported in the brain of aged transgenic mice probably due to nucleosome remodelling which may lead to a decreased accessibility of transcription machinery or hypermethylation of CpG island of the gene promotor [96].

3.1. DNA Methylation

DNA methylation is the first epigenetic modification identified on DNA that regulates gene expression in all known vertebrates. DNA methyltransferases (DNMTs), DNMT1, DNMT2, DNMT3a, DNMT3b, catalyze the transfer of a methyl group to the carbon atom in position 5 of a cytosine moiety of single-stranded DNA in the presence of the methyl donor S-adenosyl methionine (SAM) [99, 100]. DNMT3a and 3b fall in the group of de novo methyltransferases (they methylate previously unmethylated CpG sequences), while DNMT1 copies the preexisting methylation marks onto the new strand during replication [101]. SAM is derived from one-carbon metabolism which is regulated by availability of folate, vitamin B12 and vitamin B6. In promoter regions, DNA methylation occurs on the 5' position of cytosine residue in CpG rich regions called CpG islands. Methylation at this site disrupts the binding of transcription factors and recruitment of methyl-CpG-binding domain proteins that are associated with chromatin compaction and gene silencing. On the contrary, methylation within the body of the gene usually marks active transcription [102]. Recently it has been discovered that methylated cytosine (5mC) can be further hydroxylated, predominantly in brain tissue. This reaction is mediated by the TET proteins family. 5-hydroxymethylcytosine (5hmC) formation is suggested as a potential mechanism leading to DNA demethylation [103].

Studies with monozygotic twin siblings who had spent a long period of their lives apart have demonstrated that extrinsic factors derived from environmental exposure affect DNA methylation patterns [104]. DNA methylation may be affected by aging, deficiencies of vitamin B12, B6 and folic acid, oxidative stress, pesticides and heavy metal exposure. Some of these factors such as arsenic, cadmium, lead and pesticides interfere with the regulation of DNMTs

[105]. Others such as deficiencies of vitamin B12, B6 and folic acid influence DNA methylation patterns altering SAM availability via metabolites of the SAM cycle (Figure 2).

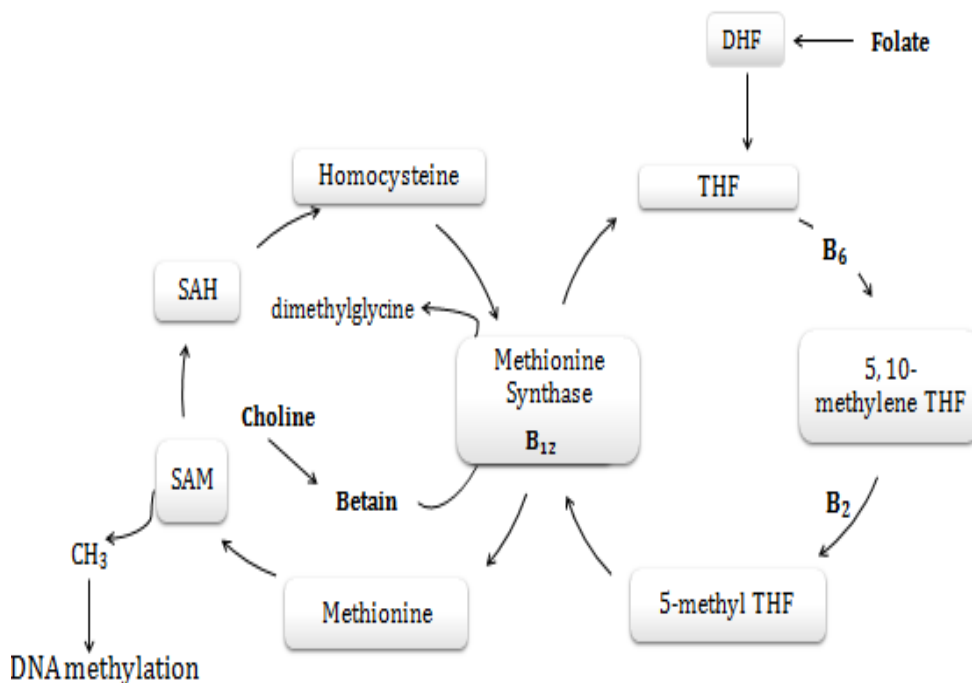


Figure 2. SAM cycle.

5-methyltetrahydrofolate donates its methyl group to homocysteine (HCY) and converts it to methionine in a reaction which requires vitamin B12. Then methionine is converted to SAM. SAM is thereby converted to S-adenosylhomocysteine (SAH), which in turn is converted to HCY in a reversible reaction. SAM level has been demonstrated to be decreased in post mortem AD brain, and increased HCY has been suggested as a risk factor for AD [106]. The lower availability of SAM can be related to the altered expression of genes involved in APP metabolism and A β accumulation.

Normal aging involves the gradual decline in memory and cognitive functions that are correlated with a profound changes (global DNA hypomethylation and hypermethylation of CpG islands) in DNA methylation [107]. DNA methylation modifications can be divided into two groups in AD: global DNA methylation modifications and gene-specific DNA methylation modifications [108]. Genomic and sequence-specific DNA methylation patterns are distinct between different brain areas. In 2010 Mastroeni et al. [109] have demonstrated that global DNA and RNA methylation status in entorhinal cortex layer II neurons are significantly diminished in AD. In accordance with this data, 5mC and 5hmC levels have been found to be decreased in hippocampus of AD patients [110]. Recently, DNA hypomethylation has been demonstrated to be correlated with a greater amyloid plaque burden, enhanced APP production, and increased activities of BACE-1 and PS1 [110, 111]. On the other hand, DNA hypermethylation is also present in AD. A postmortem human frontal cortex study has shown the presence of DNA hypermethylation in AD cases [112]. Evidence from genetic and immunohistochemical studies on AD cases supports the notion that aberrant DNA methylation modification of a number of specific genes, which are believed to be related to AD, can

contribute to onset and progression of the disease. In this context, APP gene has been investigated as a primary target.

APP Gene

In an early study, Toghi et al. (1999) [113] have reported the hypomethylation in APP promoter in aged brain, which would be consistent with an increase in APP expression with aging. Afterwards, methylation status of different regions of APP promoter has been examined in AD by Barrachina and Ferrer, (2009) [114], and no significant difference has been found between control and disease samples obtained from frontal cortex and hippocampi of postmortem brains in various stages of AD. They have concluded that small changes in methylation of DNA promoters in vulnerable cells might have not been detected in total homogenates [114].

BACE, PS1 and DNMT Genes

According to findings of in vitro studies, BACE and PS1 genes are regulated by methylation. The reduction of folate and vitamin B12 in culture medium can cause a decrease in SAM concentration with subsequent increase in PS1 and BACE levels and an increase in A β formation [115-117]. Most recently, the methylation status of PS-1, BACE1, DNMT1, DNMT3A and DNMT3B promoters in blood DNA obtained from AD patients and controls have been analysed. It has been found that none of the studied regions is differently methylated between AD and controls [118].

Neprilysin (NEP)

NEP is one of the enzymes involved in A β degradation. Its expression is decreased in both AD and aged brains. Promoter region for NEP has been found to be hypermethylated in murine cerebral endothelial cells model. It has been concluded that hypermethylation of NEP promoter suppresses NEP's expression at mRNA and protein levels, which in turn causes to increased A β accumulation [119].

Some other genes that are thought to be involved in AD pathogenesis have also been investigated in respect with the promoter methylation status. The promoters for *cAMP-responsive element* that is associated with synaptic plasticity have been found to be hypermethylated in AD [112]. *Glycogen synthase kinase 3 β* (GSK3 β) is the major kinase that phosphorylates tau protein in brain. It has been reported that GSK3 β activity and expression levels are increased in AD brains which may be due to hypomethylation of its promoter region [120]. Promoter methylation of *Neuronal protein phosphatase 2A* has been found to be downregulated in affected brain regions of AD patients, causing the accumulation of both phosphorylated tau and APP isoforms, and increased secretion of A β peptides [121, 122]. Promoters of *cyclooxygenase-2*, *brain-derived neurotrophic factor* and *NF- κ B* have been reported to be hypomethylated, whereas promoters of *cAMP response element-binding protein* and *synaptophysin* have been reported to be hypermethylated in frontal cortex of postmortem AD brain [112].

3.2. Histone Modifications

Reversible post-translational modifications such as phosphorylation, acetylation, methylation, glycosylation and ubiquitination modulate biological activity of proteins [123]. In recent years, protein acetylation at the lysine ϵ -amino group has gained considerable interest due to its ability to regulate important cellular processes. Modifications in histone proteins may alter the access of the transcriptional machinery to genes in DNA. Acetylation of histone tails is the most frequently studied and best defined histone modification. Addition of an acetyl group from acetyl-coenzyme A to lysine residues on the N-terminus that protrudes from the surface of the nucleosome which is named as "histone tail" is accomplished by histone acetyltransferases (HATs). Removal of these acetyl groups from histone tails is catalyzed by histone deacetylases (HDACs) [124]. Acetylation of histone tails neutralizes their positive charge and decreases the interaction between histone tails and negatively charged phosphate groups of DNA backbone. This conformational relaxation enhances accessibility of the gene for transcription machinery. On the contrary, deacetylation of histone tails produces a more condensed chromatin and results in gene silencing. HDACs are classified into four classes as I, II, III and IV according to their sequence homology. Class I, II and IV HDACs are Zn-dependent whereas class III HDACs that are called as sirtuins are nicotinamide adenine dinucleotide (NAD⁺)-dependent.

HDAC2 is a class I HDAC that is suggested to negatively regulate memory and synaptic plasticity in healthy mouse brain. The protein level of HDAC2 in AD mouse models and also in postmortem brain tissues of AD have been found to be increased along with cognitive impairment [125]. After the discovery of the fact that histone acetylation vigorously participates to activity-dependent neural plasticity via regulation of critical gene transcription, inhibition of histone acetylation has been investigated in mouse models of AD. In the study conducted by Francis et al. acetylation level of hippocampal histone 4 (H4) in APP/PS1 double mutant transgenic mice has been found to be 50% lower than in wild-type littermates, and this has been linked with impairment in associative learning that is observed in the transgenic mice. Acute administration of HDAC inhibitor, trichostatin A, has rescued the deficiency in H4 acetylation and increased associative memory formation in APP/PS1 mice [126]. Ricobaraza et al. (2009) [127] have demonstrated that intra-perinatal injection of phenylbutyrate (a HDAC inhibitor) improves spatial memory consolidation in 16 month-old Tg2576 mice. This study has not reported a change in amyloid pathology, but rather determined a reduction in pathological phospho-tau levels. In another study, phenylbutyrate treatment has been shown to rescue decreased hippocampal H4 acetylation and impaired dendritic spine density [127, 128]. Notably, these studies have demonstrated that HDAC inhibitors are effective just after the onset of disease. In the case of advanced stage of disease progression, prolonged treatment with sodium butyrate has shown to improve associative memory in APPPS1-21 mice even when administered at a very advanced stage of pathology. The recovery of memory function has been found to be associated with elevated hippocampal histone acetylation and increased expression of genes implicated in associative learning [129]. Recently, orally administrated MS-275 (Entinostat), a class I HDAC inhibitor, has been shown to improve behavioral impairment, reduce neuroinflammation and amyloid plaque deposition in hippocampus and cortical regions of APPPS1-21 mice [130]. Taken together, HDAC inhibitors reduce amyloid plaque deposition, possibly by regulating expression of enzymes that control A β processing and clearance [131].

There are some evidence for the potential role of A β on histone acetylation. Lithner et al., 2013 [132] have reported that treatment with A β oligomers results in increased levels of acetylated H3K14 and loss of dendritic spines in cortical/hippocampal cultures. Further studies are needed to reveal causative interaction between A β and histone modifications.

HDAC6 is a member of class II HDAC and is primarily localized in the cytoplasm. By modifying acetylation levels of α -tubulin and cortactin, an F-actin-binding protein, HDAC6 regulates microtubule-dependent and actin-dependent cell motility. HDAC6 is suggested as a potential modulator of tau phosphorylation and accumulation. Ding et al. [133] have demonstrated *in vitro* and *in vivo* that HDAC6 co-localizes with tau in hippocampus of AD brain; selective inhibition of tubulin deacetylation activity of HDAC6 by tubacin does not disrupt HDAC6-tau interaction but decreases site-specific tau phosphorylation (at Thr231); protein level of HDAC6 is elevated in cortex and hippocampus of postmortem AD brain as compared to those in young normal brain.

First identified in *saccharomyces cerevisiae*, sirtuins which stands for silent information regulator factor 2 (Sir2), is grouped as class III HDACs. NAD⁺ is required for sirtuin activity [134]. Not only histones but also non-histone proteins such as transcription factors, metabolic and apoptotic modulators are the target for sirtuins. Sirtuins are highly conserved from prokaryotes to humans, the sirtuin family comprises seven homologs (SIRT1-7) in mammals. They differ from each other according to their tissue specificity, cellular localization, expression patterns, enzymatic activity and substrates [135]. SIRT1, SIRT6 and SIRT7 are nuclear proteins. SIRT1 shuttles between nucleus and cytoplasm. SIRT2 is predominantly localized in cytoplasm [136]. SIRT3, SIRT4 and SIRT5 are mitochondrial proteins. Recent findings suggest that localization of SIRT3 changes from the mitochondria to the nucleus, when SIRT3 and SIRT5 are co-expressed [137]. Certain sirtuins have also mono-ADP-ribosyl transferase activity. SIRT4 and SIRT6 have weak deacetylase activity but they are able to transfer ADP-ribosyl group to their substrates [138]. SIRT5 was identified as a lysine demalonylase and desuccinylase [139].

In early 2000s, epidemiological studies have revealed that individuals who maintain a low calorie diet have a reduced risk of developing AD. The influence of calorie restriction (CR) on AD pathology has been examined in AD mouse models, and it has been shown that CR reduces amyloid plaques by activating SIRT1 [140, 141]. Neuroprotective effect of CR has been attributed to increase in NAD⁺/NADH ratio and subsequent activation of SIRT1 during the CR. SIRT1 is the most frequently studied sirtuin in AD. It is highly expressed in certain brain regions including cortex, hippocampus, cerebellum and hypothalamus. Non-histone substrates of SIRT1 are core RNA polymerase I transcriptional machinery, HAT p300/CBP [142], tumor suppressors p53 and p73 [143, 144], nuclear factor κ -light chain enhancer of activated B cells (NF- κ B) [145], forkhead box protein O (FOXO) family of transcription factors [146] and peroxisome proliferator-activated receptor γ (PPAR γ) coactivator-1 α (PGC-1 α) [147]. All of these molecules are involved in regulation of neurogenesis, mitochondrial biogenesis, apoptosis and antioxidant defence in the brain. SIRT1 inhibits p53 and NF- κ B and consequently reduces their pro-apoptotic and pro-inflammatory effects, respectively. SIRT1 may be involved in neuroprotection due to its inhibitor effect on p53-mediated apoptosis and upregulating role in α -secretase production [148]. According to results of studies with cell culture models and transgenic mouse models, SIRT1 overexpression increases α -secretase activity, reduces A β formation, promotes neuronal survival, and is neuroprotective against A β toxicity [141, 149, 150]. SIRT1 deacetylates retinoic acid receptor and activates transcription of ADAM10 which

encodes α -secretase. Since α -secretase is responsible for non-amyloidogenic cleavage of APP, upregulation of α -secretase shifts APP processing to reduce the pathological accumulation of toxic A β peptides [149].

There is some evidence for beneficial effect of SIRT1 activation on taupathy. It has been demonstrated that breakdown of tau is inhibited when tau is acetylated by histone acetyltransferase p300. SIRT1 deacetylates the acetylated tau and promotes its degradation and clearance. SIRT1 downregulation has been shown to inhibit tau polyubiquitination and tau turnover, thus results in accumulation of phospho-tau [151]. Recently Du et al., 2014 [152] have demonstrated that activation of SIRT1 attenuates cerebral ventricular streptozotocin-induced tau hyperphosphorylation and cognitive injuries in rat hippocampi.

Metabolic regulator role of SIRT1 is well documented. SIRT1 controls metabolic processes by activating PGC-1 α , a master regulator of mitochondrial biogenesis. PGC-1 α activation results in increased insulin sensitivity and inhibition of mitochondrial dysfunction. In this context, SIRT1 can mediate neuroprotection [153]. As another neuroprotective mechanism, SIRT1 induces basal rate of autophagy which is an important way for the removal of toxic misfolded protein aggregates such as A β [154].

Julien et al. [155] have reported that cortical SIRT1 expression is decreased in autopsy specimens of AD patients but not in the individuals with MCI; SIRT1 mRNA and protein levels are negatively correlated with the duration of symptoms and the accumulation of tau, but are weakly correlated with the A β 42. In accordance with these data, a significant decline in serum SIRT1 level has been determined in patients with both AD and MCI in comparison to healthy elderly individuals [156].

Taken together, SIRT1 activation seems as a useful approach to prevent onset and progression of AD, thus has gained great interest. Resveratrol is a naturally occurring polyphenol found in red grapes and red wine. It is known with its antioxidant, anti-carcinogenic and anti-inflammatory activities. In addition to these properties it is a powerful activator of SIRT1 [157]. SIRT1 activation by resveratrol has been shown to protect against polyglutamine toxicity [158] and plaque formation in AD models. Also, it may prevent apoptotic death of neurons by deacetylating and subsequently repressing p53 activity [159]. Therapeutic role of resveratrol in AD has been extensively investigated. In transgenic mouse model of AD, resveratrol has been shown to prevent learning impairment by increasing deacetylation of PGC-1 α and p53 [160]. In addition, resveratrol has been demonstrated to inhibit amyloid burden by increasing mitochondrial complex IV protein levels in mouse model of AD [161]. Resveratrol is undergoing evaluation in clinical trials (www.clinicaltrials.gov).

SIRT3 is a mitochondrial protein. SIRT3 decreases mitochondrial membrane potential and production of ROS [162]. SIRT3 has been thought to be involved in neuroprotection. Because deacetylation activity of SIRT3 may protect neurons via decreasing ROS levels or increasing O₂ consumption in mitochondria [163].

In addition to histone acetylation, dysregulation of histone methylation has also been suggested to be related to synaptic plasticity and cognition [164, 165].

3.3. MicroRNAs

MicroRNAs (miRNAs) are small non-coding RNAs (21-25 nucleotides in length) and regulate gene expression post-transcriptionally. miRNAs bind to target mRNAs at their 3'-

untranslated regions and suppress their translation and/or promote their degradation [166]. More than 30,000 miRNAs have been identified in humans. As indicated by a large number of studies in the last decade, miRNAs play an important role in various processes such as proliferation, differentiation and apoptosis in both healthy and pathological states [167]. miRNAs can be used as potential diagnostic tools for many diseases due to their stability and convenience of detection in serum. The finding that specific circulating miRNAs are altered in plasma and serum of cancer patients has opened the door to novel diagnostic and possibly therapeutic tools for AD.

Orchestrating the changes in gene expression and protein production, miRNAs take place in memory consolidation and contribute to pathogenesis of neurodegenerative diseases. miRNAs are highly expressed in different neuronal compartments of the brain, and many of them are brain-specific or brain-enriched [167]. miRNAs have selective distribution within neurons and synapses. They have been found to regulate translation of proteins needed for neurite outgrowth, synaptogenesis and activity-driven synaptic plasticity [168-170]. They may also orchestrate fine-tuning of gene expression required for learning, memory and cognitive performance in general.

A number of miRNAs have been strongly implicated in AD, particularly brain-enriched or brain-specific miRNAs. Identifying deregulated miRNAs, miRNA profiling studies have opened a new area for researches in the field of AD. A number of specific miRNAs have been determined to be dysregulated in patients with AD. These specific miRNAs may affect the AD-related pathology either directly (via APP and BACE1), or indirectly by affecting genes related with neurogenesis and immune response [171-173]. In vitro studies have revealed that miR-106a, -520c [174], members of the miR-20a family, such as -20a, -106a/b and -17 [175], miR-16 and -101 [176, 177] and most recently miR-147, -655, -323-3p and -153 [178] directly regulate APP.

On the other hand, A β is a powerful regulator of miRNA levels. Schonrock et al. [179] have found that A β leads to a severe change in miRNA profiles; incubation of primary hippocampal neurons with A β 42 preparation causes to down-regulation of substantial proportion of miRNAs (miR-9, -181c, -148b, -30c, -20b, -361, -21, 409-3p and Let-71).

The expression of certain miRNAs changes over time, supporting the transient effect on miRNA expression during AD development. Loss of miRNA cluster containing miR-29a, -29b1 and -9 in sporadic forms of AD has been found to correlate with increased BACE1 expression and increased A β levels [172]. In an in vitro study, the level of BACE1 has been found to be down-regulated in mice overexpressing miR-29c and it has been concluded that miR-29c might be an endogenous BACE1 regulator [180]. Animal studies have indicated an inverse correlation between BACE1 protein level and miR-298, -328 [181], -103 and -107 [182].

Previous evidence of post-mortem AD brains and serum samples have demonstrated a dysregulation in a number of miRNAs, including miR-9, miR-29a, miR-29b, miR-101, miR-106, miR-125b, miR-181c, miR-137 and miR-125b that are potentially contribute to increased APP expression and A β production [171, 172, 183-187].

4. FUTURE PERSPECTIVES

There is strong evidence that epigenetic alterations are inextricably linked with several major pathological processes in AD. Pathological alterations finally leading the AD are established very early in life. This can be explained by latent early-life associated regulation (LEARn) model. According to this model various environmental factors including diet, are associated genes in a long-term fashion, beginning at early developmental stages. These modifications that do not have pathological consequences until later in life are mediated by epigenetic alterations in the promoter regions of genes [188]. Regardless a cause or consequence, epigenetic dysregulations are an important factor contributing to development of AD. Unlike DNA mutations, epigenetic disruptions are potentially reversible and therefore provide an important target for pharmacological intervention. The possibility of transgenic animal models of AD has facilitated to search the role of epigenetic mechanisms on disease-related pathologies such as A β and tau. Some epigenetic changes occur long before the appearance of AD symptoms. If such changes are identified and detected early, it could be possible to prevent onset of AD via epigenetic therapeutic approaches. Despite the intense investigation in this field previous findings are not enough for approval of epigenetic drugs in AD. DNA methylation alterations in promoter of AD-related genes should be better understood. Further researches are needed to reveal DNA methylation changes affecting the transcription of AD-related genes, particularly effect of demethylation on hypermethylated gene promoters should be examined in AD models. SIRT1 is frequently studied but not much is known about the role of other sirtuins in AD. Other sirtuins, particularly SIRT2 and SIRT3 also seem to play a role in AD. It would be interesting to see whether other sirtuins have an effect in AD pathogenesis. Recently, SIRT1 activators have become a novel target for AD management and are undergoing evaluation in clinical trials. miRNAs are promising molecular tools to identify main pathways implicated in different aspects of the disease such as accumulation of A β , hyperphosphorylation of tau, neuroinflammation and neuronal cell death. Accumulating evidence suggests that alterations in the miRNA profiles of brain contribute to AD pathogenesis. Thus, understanding of miRNA-mediated gene silencing which contributes to AD pathogenesis provides a new perspective in AD management.

REFERENCES

- [1] Wang, YJ; Zhou, HD; Zhou, XF. Clearance of amyloid-beta in Alzheimer's disease: progress, problems and perspectives. *Drug Discov Today*, 2006, 11, 931-938.
- [2] Anantharaman, M; Tangpong, J; Keller, JN; Murphy, MP; Markesbery, WR; Kinningham, KK. St Clair DK. β -Amyloid mediated nitration of manganese superoxide dismutase: implication for oxidative stress in a APPNLh/NLh X PS-1P264L/ P264L double knock-in mouse model of Alzheimer's disease. *Am J Pathol.*, 2006, 168, 1608–1618.
- [3] Cao, X; Sudhof, TC. A transcriptionally [correction of transcriptively] active complex of APP with Fe65 and histone acetyltransferase Tip60. *Science*, 2001, 293, 115–120
- [4] Cao, X; Sudhof, TC. Dissection of amyloid-beta precursor protein-dependent transcriptional transactivation. *J Biol Chem.*, 2004, 279, 24601–24611.

- [5] Boddapati, S; Levites, Y; Sierks, MR. Inhibiting β -secretase activity in Alzheimer's disease cell models with single-chain antibodies specifically targeting APP. *J Mol Biol.*, 2011, 405, 436–447.
- [6] Zheng, WH; Bastianetto, S; Mennicken, F; Ma, W; Kar, S. Amyloid β peptide induces tau phosphorylation and loss of cholinergic neurons in rat primary septal cultures. *Neuroscience*, 2002, 115, 201–211.
- [7] Castellano, JM; Deane, R; Gottesdiener, AJ; Verghese, PB; Stewart, FR; West, T; Paoletti, AC; Kasper, TR; DeMattos, RB; Zlokovic, BV; Holtzman, DM. Low-density lipoprotein receptor overexpression enhances the rate of brain-to-blood Ab eta clearance in a mouse model of beta-amyloidosis. *Proc Natl Acad Sci U S A*, 2012, 109, 15502–15507.
- [8] Deane, R; Du Yan, S; Submamaryan, RK; LaRue, B; Jovanovic, S; Hogg, E; Welch, D; Manness, L; Lin, C; Yu, J; Zhu, H; Ghiso, J; Frangione, B; Stern, A; Schmidt, AM; Armstrong, DL; Arnold, B; Liliensiek, B; Nawroth, P; Hofman, F; Kindy, M; Stern, D; Zlokovic, B. RAGE mediates amyloid-beta peptide transport across the blood-brain barrier and accumulation in brain. *Nature Medicine*, 2003, 9, 907–913.
- [9] Deane, R; Wu, Z; Zlokovic, BV. RAGE (yin) versus LRP (yang) balance regulates alzheimer amyloid beta-peptide clearance through transport across the blood-brain barrier. *Stroke*, 2004, 35, 2628–2631.
- [10] Holtzman, DM. Role of apoe/Abeta interactions in the pathogenesis of Alzheimer's disease and cerebral amyloid angiopathy. *J Mol Neurosci.*, 2001, 17, 147–155.
- [11] Wang, DS; Dickson, DW; Malter, JS. β -Amyloid Degradation and Alzheimer's Disease. *J Biomed Biotechnol*, 2006, 2006, 1–12.
- [12] Lee, VM; Goedert, M; Trojanowski, JQ. Neurodegenerative tauopathies. *Annu Rev Neurosci.*, 2001, 24, 1121–1159.
- [13] Alonso, AD; Zaidi, T; Grundke-Iqbal, I; Iqbal, K. Role of abnormally phosphorylated tau in the breakdown of microtubules in Alzheimer disease. *Proc Natl Acad Sci USA*, 1994, 91, 5562–5566.
- [14] Sengupta, A; Kabat, J; Novak, M; Wu, Q; Grundke-Iqbal, I; Iqbal, K. Maximal inhibition of tau binding to microtubules requires the phosphorylation of tau at both Thr 231 and Ser 262. *Neurobiol Aging*, 1998, 19S, S124– S524.
- [15] Freir, DB; Fedriani, R; Scully, D; Smith, IM; Selkoe, DJ; Walsh, DM; Regan, CM. Abeta oligomers inhibit synapse remodelling necessary for memory consolidation. *Neurobiol Aging*, 2011, 32, 2211–2218.
- [16] Mufson, EJ; Binder, L; Counts, SE; DeKosky, ST; de Toledo-Morrell, L; Ginsberg, SD; Ikonomic, MD; Perez, SE; Scheff, SW. Mild cognitive impairment: pathology and mechanisms. *Acta Neuropathol.*, 2012, 123, 13–30.
- [17] Scheff, SW; Price, DA; Schmitt, FA; DeKosky, ST; Mufson, EJ. Synaptic alterations in CA1 in mild Alzheimer disease and mild cognitive impairment. *Neurology*, 2007, 68, 1501–1508.
- [18] Berchtold, NC; Coleman, PD; Cribbs, DH; Rogers, J; Gillen, DL; Cotman, CW. Synaptic genes are extensively downregulated across multiple brain regions in normal human aging and Alzheimer's disease. *Neurobiol Aging*, 2013, 34, 1653–1661.
- [19] Chang, RY; Nouwens, AS; Dodd, PR; Etheridge, N. The synaptic proteome in Alzheimer's disease. *Alzheimers Dement.*, 2013, 9, 499–511.

- [20] Krstic, D; Knuesel, I. Deciphering the mechanism underlying late-onset Alzheimer disease. *Nat Rev Neurol.*, 2013, 9, 25–34.
- [21] Praticò, D. Oxidative stress hypothesis in Alzheimer's disease: a reappraisal. *Trends Pharmacol Sci.*, 2008, 29, 609–615.
- [22] Ferreira, E; Baldeiras, I; Ferreira, IL; Costa, RO; Rego, AC; Pereira, CF; Oliveira, CR. Mitochondrial- and endoplasmic reticulum-associated oxidative stress in Alzheimer's disease: from pathogenesis to biomarkers. *Int J Cell Biol.*, 2012, 735206.
- [23] Beal, MF. Oxidatively modified proteins in aging and disease. *Free Radical Bio Med.*, 2002, 32, 797–803.
- [24] Praticò, D; Sung, S. Lipid peroxidation and oxidative imbalance: early functional events in Alzheimer's disease. *J Alzheimers Dis.*, 2004, 6, 171–175.
- [25] Matsuoka, Y; Picciano, M; La Francois, J; Duff, K. Fibrillar β -amyloid evokes oxidative damage in a transgenic mouse model of Alzheimer's disease. *Neuroscience*, 2001, 104, 609–613.
- [26] Kontush, A. Amyloid-beta: an antioxidant that becomes a pro-oxidant and critically contributes to Alzheimer's disease. *Free Radic Biol Med.*, 2001, 31(9), 1120–1131.
- [27] Behl, C; Davis, JB; Lesley, R; Schubert, D. Hydrogen peroxide mediates amyloid β protein toxicity. *Cell*, 1994, 77, 817–827.
- [28] Butterfield, DA. Amyloid b-peptide (1–42)-induced Oxidative Stress and Neurotoxicity: Implications for Neurodegeneration in Alzheimer's Disease Brain. *Free Radic Res.*, 2002, 36, 1307–1313.
- [29] Quiroz-Baez, R; Rojas, E; Arias, C. Oxidative stress promotes JNK-dependent amyloidogenic processing of normally expressed human APP by differential modification of α -, β - and γ -secretase expression. *Neurochem Int.*, 2009, 55, 662–670.
- [30] Li, F; Calingasan, NY; Yu, F; Mauck, WM; Toidze, M; Almeida, CG; Takahashi, RH; Carlson, GA; Flint Beal, M; Lin, MT; Gouras, GK. Increased plaque burden in brains of APP mutant MnSOD heterozygous knockout mice. *J Neurochem.*, 2004, 89, 1308–1312.
- [31] Dumont, M; Wille, E; Stack, C; Calingasan, NY; Beal, MF; Lin, MT. Reduction of oxidative stress, amyloid deposition, and memory deficit by manganese superoxide dismutase overexpression in a transgenic mousemodel of Alzheimer's disease. *The FASEB Journal*, 2009, 23, 2459–2466.
- [32] Murakami, K; Murata, N; Noda, Y; Tahara, S; Kaneko, T; Kinoshita, N; Hatsuta, H; Murayama, S; Barnham, KJ; Irie, K; Shirasawa, T; Shimizu, T. SOD1 (copper/zinc superoxide dismutase) deficiency drives amyloid β protein oligomerization and memory loss in mouse model of Alzheimer disease. *J Biol Chem.*, 2011, 286, 44557–44568.
- [33] Poon, HF; Calabrese, V; Scapagnini, G; Butterfield, DA. Free radicals and brain aging. *Clin Geriatr Med.*, 2004, 20, 329.
- [34] Harman, D. Aging: a theory based on free radical and radiation chemistry. *J Geront.*, 1956, 11, 289
- [35] Gamblin, TC; King, ME; Kuret, J; Berry, RW; Binder, LI. Oxidative regulation of fatty acid-induced tau polymerization. *Biochemistry*, 2000, 39, 14203–14210.
- [36] Bush, AI. The Metal Theory of Alzheimer's Disease. *J Alzheimers Dis.*, 2012, 30, 1–5.
- [37] Kozłowski, H; Łuczkowski, M; Valensin, D; Valensin, G. Metal ion binding properties of proteins related to neurodegeneration. Sigel A, Sigel H and Sigel RKO.Eds., Wiley & Sons. *Metal Ions in Life Sciences*, 2006, 1, 61–87.

- [38] Nunomura, A; Perry, G; Aliev, G; Hirai, K; Takeda, A; Balraj, EK; Jones, PK; Ghanbari, H; Wataya, T; Shimohama, S; Chiba, S; Atwood, CS; Petersen, RB; Smith, MA. Oxidative damage is the earliest event in Alzheimer disease. *J Neuropathol Exp Neurol.*, 2001, 60, 759–767.
- [39] Greenough, MA; Camakaris, J; Bush, AI. Metal dyshomeostasis and oxidative stress in Alzheimer's disease. *Neurochem Int.*, 2013, 62, 540-555.
- [40] Sipe, JC; Lee, P; Beutler, E. Brain iron metabolism and neurodegenerative disorders. *Dev Neurosci.*, 2002, 24, 188–196
- [41] Bush, AI. The metallobiology of Alzheimer's disease. *Trends Neurosci.*, 2003, 26, 207–214
- [42] Cherny, RA; Atwood, CS; Xilinas, ME; Gray, DN; Jones, WD; McLean, CA; Barnham, KJ; Volitakis, I; Fraser, FW; Kim, Y; Huang, X; Goldstein, LE; Moir, RD; Lim, JT; Beyreuther, K; Zheng, H; Tanzi, RE; Masters, CL; Bush, AI. Treatment with a copper-zinc chelator markedly and rapidly inhibits b-amyloid accumulation in Alzheimer's disease transgenic mice. *Neuron*, 2001, 30, 665–676.
- [43] Frazzini, V; Rockabrand, E; Mocchegiani, E; Sensi, SL. Oxidative stress and brain aging, is zinc the link? *Biogerontology*, 2006, 7, 307-314.
- [44] Maret, W; Sandstead, HH. Zinc requirements and the risks and benefits of zinc supplementation. *J Trace Elem Med Biol.*, 2006, 20, 3-18.
- [45] Yamamoto, A; Shin, RW; Hasegawa, K; Naiki, H; Sato, H; Yoshimasu, F; Kitamoto, T. Iron (III) induces aggregation of hyperphosphorylated tau and its reduction to iron (II) reverses the aggregation: implications in the formation of neurofibrillary tangles of Alzheimer's disease. *J Neurochem.*, 2002, 82, 1137-1147.
- [46] Smith, DG; Cappai, R; Barnham, KJ. The redox chemistry of the Alzheimer's disease amyloid beta peptide. *Biochim Biophys Acta*, 2007, 1768, 1976-1990.
- [47] Lannfelt, L; Blennow, K; Zetterberg, H; Batsman, S; Ames, D; Harrison, J; Masters, CL; Targum, S; Bush, AI; Murdoch, R; Wilson, J; Ritchie, CW. PBT2-201-EURO study group. Safety; efficacy; and biomarker findings of PBT2 in targeting Abeta as a modifying therapy for Alzheimer's disease: a phase IIa, double-blind, randomized, placebo-controlled trial. *Lancet Neurol.*, 2008, 7, 779-786.
- [48] Zhu, X; Perry, G; Moreira, PI; Aliev, G; Cash, AD; Hirai, K; Smith, MA. Mitochondrial abnormalities and oxidative imbalance in Alzheimer disease, *J Alzheimers Dis.*, 2006, 9, 147–153.
- [49] Webster, R. *Neurotransmitters, Drugs and Brain Function*. John Wiley & Sons, 2003.
- [50] Swerdlow, RH. Brain aging, Alzheimer's disease, and mitochondria. *Biochim Biophys Acta.*, 2011, 1812, 1630-1639.
- [51] Rodrigues, CMP; Sol'a, S; Brito, MA; Brondino, CD; Brites, D; Moura, JJG. Amyloid β -peptide disrupts mitochondrial membrane lipid and protein structure: protective role of tauroursodeoxycholate. *Biochem Biophys Res Commun.*, 2001, 281, 468–474.
- [52] Casley, CS; Canevari, L; Land, JM; Clark, JB; Sharpe, MA. β -Amyloid inhibits integrated mitochondrial respiration and key enzyme activities. *J Neurochem.*, 2002, 80, 91–100.
- [53] Chen, CH; Zhou, W; Liu, S; Deng, Y; Cai, F; Tone, M; Tone, Y; Tong, Y; Song, W. Increased NF- κ B signalling up-regulates BACE1 expression and its therapeutic potential in Alzheimer's disease. *Int J Neuropsychopharmacol.*, 2012, 15(1), 77-90.

- [54] Vom Berg, J; Prokop, S; Miller, KR; Obst, J; Kälin, RE; Lopategui-Cabezas, I; Wegner, A; Mair, F; Schipke, CG; Peters, O; Winter, Y; Becher, B; Heppner, FL. Inhibition of IL-12/ IL-23 signaling reduces Alzheimer's disease-like pathology and cognitive decline. *Nat Med.*, 2012, 18, 1812-9.
- [55] Choi, SH; Bosetti, F. Cyclooxygenase-1 null mice show reduced neuroinflammation in response to beta-amyloid. *Aging* (Albany NY), 2009, 1, 234–244.
- [56] Liu, S; Liu, Y; Hao, W; Wolf, L; Kiliaan, AJ; Penke, B; Rübe, CE; Walter, J; Heneka, MT; Hartmann, T; Menger, MD; Fassbender, K. TLR2 is a primary receptor for Alzheimer's amyloid b peptide to trigger neuroinflammatory activation. *J Immunol.*, 2012, 188, 1098–1107.
- [57] Piro, JR; Benjamin, DI; Duerr, JM; Pi, Y; Gonzales, C; Wood, KM; Schwartz, JW; Nomura, DK; Samad, TA. A Dysregulated Endocannabinoid-Eicosanoid Network Supports Pathogenesis in a Mouse Model of Alzheimer's Disease. *Cell Reports*, 2012, 1, 617–623.
- [58] Sastre, M; Dewachter, I; Landreth, GE; Willson, TM; Klockgether, T; van Leuven, F; Heneka, MT. Nonsteroidal anti-inflammatory drugs and peroxisome proliferator-activated receptor-gamma agonists modulate immunostimulated processing of amyloid precursor protein through regulation of beta-secretase. *J Neurosci.*, 2003, 23, 9796–9804.
- [59] Szekely, CA; Breitner, JCS; Fitzpatrick, AL; Rea, TD; Psaty, BM; Kuller, LH; Zandi, PP. NSAID use and dementia risk in the Cardiovascular Health Study: role of APOE and NSAID type. *Neurology*, 2008, 70, 17–24.
- [60] Imbimbo, BP; Solfrizzi, V; Panza, F. Are NSAIDs useful to treat Alzheimer's disease or mild cognitive impairment? *Front Aging Neurosci.*, 2010, 2, 19.
- [61] Kitazawa, M; Cheng, D; Tsukamoto, MR; Koike, MA; Wes, PD; Vasilevko, V; Cribbs, DH; LaFerla, FM. Blocking IL-1 signaling rescues cognition, attenuates tau pathology, and restores neuronal b-catenin pathway function in an Alzheimer's disease model. *J Immunol.*, 2011, 187, 6539–6549.
- [62] Shi, JQ; Shen, W; Chen, J; Wang, BR; Zhong, LL; Zhu, YW; Zhu, HQ; Zhang, QQ; Zhang, YD; Xu, J. Anti-TNF- α reduces amyloid plaques and tau phosphorylation and induces CD11c-positive dendritic-like cell in the APP/PS1 transgenic mouse brains. *Brain Res.*, 2011, 1368, 239–247.
- [63] Bertram, L; Lill, CM; Tanzi, RE. The genetics of Alzheimer disease: back to the future. *Neuron*, 2010, 68, 270–281.
- [64] Gerrish, A; Russo, G; Richards, A; Moskvina, V; Ivanov, D; Harold, D; et al. The role of variation at AbPP, PSEN1 PSEN2, and MAPT in late onset Alzheimer's disease. *J Alzheimers Dis*, 2012, 28, 377–87.
- [65] Cruchaga, C; Chakraverty, S; Mayo, K; Vallania, FLM; Mitra, RD; Faber, K; et al. Rare variants in APP PSEN1 and PSEN2 increase risk for AD in late-onset Alzheimer's disease families. *PLoS ONE*, 2012, 7(2), e31039.
- [66] Goate, A; Chartier-Harlin, MC; Mullan, M; Brown, J; Crawford, F; Fidani, L; Giuffra, L; Haynes, A; Irving, N; James, L; et al. Segregation of a missense mutation in the amyloid precursor protein gene with familial Alzheimer's disease. *Nature*, 1991, 349, 704–06.
- [67] Sleegers, K; Brouwers, N; Gijselink, I; Theuns, J; Goossens, D; Wauters, J; Del-Favero, J; Cruts, M; van Duijn, CM; Van Broeckhoven, C. APP duplication is sufficient to cause

- early onset Alzheimer's dementia with cerebral amyloid angiopathy. *Brain*, 2006, 129, 2977–83.
- [68] Kasuga, K; Shimohata, T; Nishimura, A; Shiga, A; Mizuguchi, T; Tokunaga, J; Ohno, T; Miyashita, A; Kuwano, R; Matsumoto, N; Onodera, O; Nishizawa, M; Ikeuchi, T. Identification of independent APP locus duplication in Japanese patients with early-onset Alzheimer disease. *J Neurol Neurosurg Psychiatry*, 2009, 80, 1050–1052.
- [69] Rogaev, EI; Sherrington, R; Rogaeva, EA; Levesque, G; Ikeda, M; Liang, Y; Chi, H; Lin, C; Holman, K; Tsuda, T; et al. Familial Alzheimer's disease in kindreds with missense mutations in a gene on chromosome 1 related to the Alzheimer's disease type 3 gene. *Nature*, 1995, 376, 775–778.
- [70] Di Fede, G; Catania, M; Morbin, M; Rossi, G; Suardi, S; Mazzoleni, G; Merlin, M; Giovagnoli, AR; Prioni, S; Erbetta, A; Falcone, C; Gobbi, M; Colombo, L; Bastone, A; Beeg, M; Manzoni, C; Francescucci, B; Spagnoli, A; Cantù, L; Del Favero, E; Levy, E; Salmona, M; Tagliavini, F. A recessive mutation in the APP gene with dominant-negative effect on amyloidogenesis. *Science*, 2009, 323, 1473–1477.
- [71] Bentahir, M; Nyabi, O; Verhamme, J; Tolia, A; Horré, K; Wiltfang, J; Esselmann, H; De Strooper, B. Presenilin clinical mutations can affect gamma-secretase activity by different mechanisms. *J Neurochem.*, 2006, 96, 732–742.
- [72] Kumar-Singh, S; Theuns, J; Van Broeck, B; Pirici, D; Vennekens, K; Corsmit, E; Cruts, M; Dermaut, B; Wang, R; Van Broeckhoven, C. Mean age-of-onset of familial alzheimer disease caused by presenilin mutations correlates with both increased Abeta42 and decreased Abeta40. *Hum Mutat.*, 2006, 27, 686–695.
- [73] Harold, D; Abraham, R; Hollingworth, P; et al. Genome-wide association study identifies variants at CLU and PICALM associated with Alzheimer's disease. *Nat Genet.*, 2009, 41, 1088–1093.
- [74] Hollingworth, P; Harold, D; Sims, R; et al. Common variants at ABCA7, MS4A6A/MS4A4E, EPHA1, CD33 and CD2AP are associated with Alzheimer's disease. *Nat Genet.*, 2011, 43, 429–435.
- [75] Seshadri, S; Fitzpatrick, AL; Ikram, MA; et al. Genome-wide analysis of genetic loci associated with Alzheimer disease. *JAMA*, 2010, 303, 1832–1840.
- [76] Raygani, AV; Zahrai, M; Raygani, AV; Doosti, M; Javadi, E; Rezaei, M; Pourmotabbed, T. Association between apolipoprotein E polymorphism and Alzheimer disease in Tehran, Iran. *Neurosci Lett.*, 2005, 375, 1-6.
- [77] Mahley, RW. Apolipoprotein E: cholesterol transport protein with expanding role in cell biology. *Science*, 1988, 240, 622–630.
- [78] Rocchi, A; Pellegrini, S; Siciliano, G; Murri, L. Causative and susceptibility genes for Alzheimer's disease: a review. *Brain Res Bull.*, 2003, 61, 1–24.
- [79] Lehtovirta, M; Laakso, MP; Soininen, H; Helisalmi, S; Mannermaa, A; Helkala, EL; Partanen, K; Ryyänänen, M; Vainio, P; Hartikainen, P. Volumes of hippocampus, amygdala and frontal lobe in Alzheimer's patients with different apolipoprotein E genotypes. *Neuroscience*, 1995, 67, 65–72.
- [80] Lehtovirta, M; Soininen, H; Helisalmi, S; Mannermaa, A; Helkala, EL; Hartikainen, P; Hänninen, T; Ryyänänen, M; Riekkinen, PJ. Clinical and neuropsychological characteristics in familial and sporadic Alzheimer's disease: relation to apolipoprotein E polymorphism. *Neurology*, 1996, 46, 413–419.

- [81] Bigler, ED; Lowry, CM; Anderson, CV; Johnson, SC; Terry, J; Steed, M. Dementia, quantitative neuroimaging, and apolipoprotein E genotype. *AJNR Am J Neuroradiol.*, 2000, 21, 1857–1868.
- [82] Hashimoto, M; Yasuda, M; Tanimukai, S; Matsui, M; Hirono, N; Kazui, H; Mori, E. Apolipoprotein E e4 and the pattern of regional brain atrophy in Alzheimer's disease. *Neurology*, 2001, 57, 1461–1466.
- [83] Schmechel, DE; Saunders, AM; Strittmatter, WJ; Crain, BJ; Hulette, CM; Joo, SH; Pericak-Vance, MA; Goldgaber, D; Roses, AD. Increased amyloid beta-peptide deposition in cerebral cortex as a consequence of apolipoprotein E genotype in late-onset Alzheimer disease. *Proc Natl Acad Sci USA*, 1993, 90, 9649–9653.
- [84] Berg, L; McKeel, DW; Jr. Miller, JP; Storandt, M; Rubin, EH; Morris, JC; Baty, J; Coats, M; Norton, J; Goate, AM; Price, JL; Gearing, M; Mirra, SS; Saunders, AM. Clinicopathologic studies in cognitively healthy aging and Alzheimer's disease, relation of histologic markers to dementia severity, age, sex, and apolipoprotein E genotype. *Arch Neurol.*, 1998, 55, 326–335.
- [85] Lippa, C; Smith, T; Saunders, A; Hulette, C; Pulaski-Salo, D; Roses, A. Apolipoprotein E-epsilon 2 and Alzheimer's disease: Genotype influences pathologic phenotype. *Neurology*, 1997, 48, 515–519.
- [86] Nagy, Z; Esiri, M; Jobst, K; Johnston, C; Litchfield, S; Sim, E; Smith, A. Influence of the apolipoprotein-e genotype on amyloid deposition and neurofibrillary tangle formation in Alzheimer's disease. *Neuroscience*, 1995, 69, 757–761.
- [87] Frisoni, GB; Govoni, S; Geroldi, C; Bianchetti, A; Calabresi, L; Franceschini, G; Trabucchi, M. Gene dose of the epsilon 4 allele of apolipoprotein E and disease progression in sporadic late-onset Alzheimer's disease. *Ann Neurol.*, 1995, 37, 596–604.
- [88] Hirono, N; Hashimoto, M; Yasuda, M; Kazui, H; Mori, E. Accelerated memory decline in Alzheimer's disease with apolipoprotein epsilon 4 allele. *J Neuropsychiatry Clin Neurosci.*, 2003, 15, 354–358.
- [89] Hoyt, BD; Massman, PJ; Schatschneider, C; Cooke, N; Doody, RS. Individual growth curve analysis of APOE epsilon 4-associated cognitive decline in Alzheimer disease. *Arch Neurol.*, 2005, 62, 454–459.
- [90] Kleiman, T; Zdany, K; Black, B; Rightmer, T; Grey, M; Garman, K; Macavoy, M; Gelernter, J; van Dyck, C. Apolipoprotein E epsilon 4 allele is unrelated to cognitive or functional decline in Alzheimer's disease: retrospective and prospective analysis. *Dement Geriatr Cogn Disord.*, 2006, 22, 73–82.
- [91] Martins, CA; Oulhaj, A; deJager, CA; Williams, JH. APOE alleles predict the rate of cognitive decline in Alzheimer disease: a nonlinear model. *Neurology*, 2005, 65, 1888–1893.
- [92] Tschanz, JT; Corcoran, CD; Schwartz, S; Treiber, K; Green, RC; Norton, MC; Mielke, MM; Piercy, K; Steinberg, M; Rabins, PV; Leoutsakos, JM; Welsh-Bohmer, KA; Breitner, JC; Lyketsos, CG. Progression of cognitive, functional, and neuropsychiatric symptom domains in a population cohort with Alzheimer dementia: the Cache County Dementia Progression study. *Am J Geriatr Psychiatry*, 2011, 19, 532–542.
- [93] Han, SD; Bondi, MW. Revision of the apolipoprotein E compensatory mechanism recruitment hypothesis. *Alzheimers Dement.*, 2008, 4, 251–254.
- [94] Jang, H; Serra, C. Nutrition, epigenetics, and diseases. *Clin Nutr Res.*, 2014, 3, 1–8.

- [95] Sezgin, Z; Dincer, Y. Alzheimer's disease and epigenetic diet. *Neurochem Int.*, 2014, 78, 105-116.
- [96] Marques, SCF; Lemos, R; Ferreira, E; Martins, M; de Mendonça, A; Santana, I; Outeiro, TF; Pereira, CMF. Epigenetic regulation of BACE1 in Alzheimer's disease patients and in transgenic mice. *Neuroscience*, 2012, 220, 256-66.
- [97] Kong, LN; Zuo, PP; Mu, L; Liu, YY; Yang, N. Gene expression profile of amyloid beta protein-injected mouse model for Alzheimer disease. *Acta Pharmacol Sin.*, 2005, 26, 666–672.
- [98] Papassotiropoulos, A; Fountoulakis, M; Dunckley, T; Stephan, DA; Reiman, EM. Genetics; transcriptomics; and proteomics of Alzheimer's disease. *J Clin Psychiatry*, 2006, 67, 652–670.
- [99] Yen, RW; Vertino, PM; Nelkin, BD; Yu, JJ; el-Deiry, W; Cumaraswamy, A; Lennon, GG; Trask, BJ; Celano, P; Baylin, SB. Isolation and characterization of the cDNA encoding human DNA methyltransferase. *Nucleic Acids Res.*, 1992, 20, 2287–2291.
- [100] Okano, M; Bell, DW; Haber, DA; Li, E. DNA methyltransferases Dnmt3a and Dnmt3b are essential for de novo methylation and mammalian development. *Cell*, 1999, 99, 247–257.
- [101] Jeltsch, A. Molecular enzymology of mammalian DNA methyltransferases. *Curr Top Microbiol Immunol.*, 2006, 301, 203–225.
- [102] Suzuki, MM; Bird, A. DNA methylation landscapes: provocative insights from epigenomics. *Nat Rev Genet.*, 2008, 9, 465–476.
- [103] Tan, L; Shi, YG. Tet family proteins and 5-hydroxymethylcytosine in development and disease. *Development*, 2012, 139, 1895-1902.
- [104] Feil, R; Fraga, MF. Epigenetics and the environment: emerging patterns and implications. *Nat Rev.*, 2012, 13, 97–109.
- [105] Fang, M; Chen, D; Yang, CS. Dietary polyphenols may affect DNAmethylation. *J Nutr.*, 2007, 137, 223S–228S.
- [106] Seshadri, S; Beiser, A; Selhub, J; Jacques, PF; Rosenberg, IH; D'Agostino, RB; Wilson, PW; Wolf, PA. Plasma homocysteine as a risk factor for dementia and Alzheimer's disease. *N Engl J Med.*, 2002, 346(7), 476-83.
- [107] Hernandez, DG; Nalls, MA; Gibbs, JR; Arepalli, S; van der Brug, M; Chong, S; Moore, M; Longo, DL; Cookson, MR; Traynor, BJ; Singleton, AB. Distinct DNA methylation changes highly correlated with chronological age in the human brain. *Hum Mol Genet.*, 2011, 20(6), 1164-72
- [108] Lu, H; Liu, X; Deng, Y; Qing, H. DNA methylation, a hand behind neurodegenerative diseases. *Front Aging Neurosci.*, 2013, 5, 85.
- [109] Mastroeni, D; Grover, A; Delvaux, E; Whiteside, C; Coleman, PD; Rogers, J. Epigenetic changes in Alzheimer's disease: decrements in DNA methylation. *Neurobiol Aging*, 2010, 31, 2025–2037.
- [110] Chouliaras, L; Mastroeni, D; Delvaux, E; Grover, A; Kenis, G; Hof, PR; Steinbusch, HW; Coleman, PD; Rutten, BP; van den Hove, DL. Consistent decrease in global DNA methylation and hydroxymethylation in the hippocampus of Alzheimer's disease patients. *Neurobiol Aging*, 2013, 34, 2091–2099.
- [111] Coppieters, N; Dieriks, BV; Lill, C; Faull, RL; Curtis, MA; Dragunow, M. Global changes in DNA methylation and hydroxymethylation in Alzheimer's disease human brain. *Neurobiol Aging*, 2014, 35, 1334-1344.

- [112] Rao, JS; Keleshian, VL; Klein, S; Rapoport, SI. Epigenetic modifications in frontal cortex from Alzheimer's disease and bipolar disorder patients. *Transl Psychiatry*, 2012, 2, e132.
- [113] Tohgi, H; Utsugisawa, K; Nagane, Y; Yoshimura, M; Genda, Y; Ukitsu, M. Reduction with age in methylcytosine in the promoter region -224 approximately -101 of the amyloid precursor protein gene in autopsy human cortex. *Brain Res Mol Brain Res.*, 1999, 70(2), 288–292.
- [114] Barrachina, M; Ferrer, I. DNA methylation of Alzheimer disease and tauopathy-related genes in postmortem brain. *J Neuropathol Exp Neurol.*, 2009, 68, 880–891.
- [115] Scarpa, S; Fuso, A; D'anselmi, F; Cavallaro, RA. Presenilin 1 gene silencing by S-adenosylmethionine: a treatment for Alzheimer disease? *FEBS Lett.*, 2003, 541, 145–148.
- [116] Fuso, A; Seminara, L; Cavallaro, RA; D'Anselmi, F; Scarpa, S. S-adenosylmethionine/homocysteine cycle alterations modify DNA methylation status with consequent deregulation of PS1 and BACE and beta-amyloid production. *Mol Cell Neurosci.*, 2005, 28, 195–204.
- [117] Fuso, A; Nicolìa, V; Cavallaro, RA; Ricceri, L; D'Anselmi, F; Coluccia, P; Calamandrei, G; Scarpa, S. B-vitamin deprivation induces hyperhomocysteinemia and brain S-adenosylhomocysteine, depletes brain S-adenosylmethionine, and enhances PS1 and BACE expression and amyloid-beta deposition in mice. *Mol Cell Neurosci.*, 2008, 37, 731–746.
- [118] Tannorella, P; Stoccoro, A; Tognoni, G; Petrozzi, L; Salluzzo, MG; Ragalmuto, A; Siciliano, G; Haslberger, A; Bosco, P; Bonuccelli, U; Migliore, L; Coppedè, F. Methylation analysis of multiple genes in blood DNA of Alzheimer's disease and healthy individuals. *Neurosci Lett.*, 2015, 600, 143–147.
- [119] Chen, KL; Wang, SS; Yang, YY; Yuan, RY; Chen, RM; Hu, CJ. The epigenetic effects of amyloid-beta (1-40) on global DNA and neprilysin genes in murine cerebral endothelial cells. *Biochem Biophys Res Commun.*, 2009, 378, 57–61.
- [120] Nicolìa, V; Fuso, A; Cavallaro, RA; DiLuzio, A; Scarpa, S. B vitamin deficiency promotes tau phosphorylation through regulation of GSK3beta and PP2A. *J Alzheimers Dis.*, 2010, 19, 895–907.
- [121] Sontag, E; Nunbhakdi-Craig, V; Sontag, JM; Diaz-Arrastia, R; Ogris, E; Dayal, S; Lentz, SR; Arning, E; Bottiglieri, T. Protein phosphatase 2A methyltransferase links homocysteine metabolism with tau and amyloid precursor protein regulation. *J Neurosci.*, 2007, 27, 2751–2759.
- [122] Sontag, JM; Nunbhakdi-Craig, V; Montgomery, L; Arning, E; Bottiglieri, T; Sontag, E. Folate deficiency induces in vitro and mouse brain region-specific downregulation of leucine carboxyl methyltransferase-1 and protein phosphatase 2A B(alpha) subunit expression that correlate with enhanced tau phosphorylation. *J Neurosci.*, 2008, 28, 11477–11487.
- [123] Choudhary, C; Weinert, BT; Nishida, Y; Verdin, E; Mann, M. The growing landscape of lysine acetylation links metabolism and cell signalling. *Nat Rev Mol Cell Biol.*, 2014, 15, 536–550.
- [124] Yang, XJ; Seto, E. HATs and HDACs: from structure, function and regulation to novel strategies for therapy and prevention. *Oncogene*, 2007, 26, 5310–5318.

- [125] Graff, J; Rei, D; Guan, JS; Wang, WY; Seo, J; Hennig, KM; Nieland, TJ; Fass, DM; Kao, PF; Kahn, M; Su, SC; Samiei, A; Joseph, N; Haggarty, SJ; Delalle, I; Tsai, LH. An epigenetic blockade of cognitive functions in the neurodegenerating brain. *Nature*, 2012, 483, 222–226.
- [126] Francis, YI; Fa, M; Ashraf, H; Zhang, H; Staniszewski, A; Latchman, DS; Arancio, O. Dysregulation of histone acetylation in the APP/PS1 mouse model of Alzheimer's disease. *J Alzheimers Dis.*, 2009, 18(1), 131–139.
- [127] Ricobaraza, A; Cuadrado-Tejedor, M; Pérez-Mediavilla, A; Frechilla, D; Del Río, J; García-Osta, A. Phenylbutyrate ameliorates cognitive deficit and reduces tau pathology in an Alzheimer's disease mouse model. *Neuropsychopharmacology*, 2009, 34, 1721–1732.
- [128] Ricobaraza, A; Cuadrado-Tejedor, M; Marco, S; Pérez-Otaño, I; García-Osta, A. Phenylbutyrate rescues dendritic spine loss associated with memory deficits in a mouse model of Alzheimer disease. *Hippocampus*, 2010, 22, 1040–1050.
- [129] Govindarajan, N; Agis-Balboa, C; Walter, J; Sananbenesi, F; Fischer, A. Sodium butyrate improves memory function in an Alzheimer's disease mouse model when administered at an advanced stage of disease progression. *J Alzheimers Dis.*, 2011, 24, 1–11.
- [130] Zhang, Z; Schluesener, YHJ. Oral administration of histone deacetylase inhibitor MS-275 ameliorates neuroinflammation and cerebral amyloidosis and improves behavior in a mouse model. *J Neuropathol Exp Neurol.*, 2013, 72, 178–185.
- [131] Sung, YM; Lee, T; Yoon, H; Di Battista, AM; Song, J; Sohn, Y; Moffat, E; Turner, RS; Jung, M; Kim, J; Hoe, HS. Mercaptoacetamide-based class II HDAC inhibitor lowers Ab levels and improves learning and memory in a mouse model of Alzheimer's disease. *Exp Neurol.*, 2013, 239, 192–201.
- [132] Lithner, CU; Lacor, PN; Zhao, WQ; Mustafiz, T; Klein, WL; Sweatt, JD; Hernandez, CM. Disruption of neocortical histone H3 homeostasis by soluble Abeta: implications for Alzheimer's disease. *Neurobiol Aging*, 2013, 34(9), 2081–2090.
- [133] Ding, H; Dolan, PJ; Johnson, GV. Histone deacetylase 6 interacts with the microtubule-associated protein tau. *J Neurochem.*, 2008, 106, 2119–2130.
- [134] Imai, SI; Armstrong, CM; Kaeberlein, M; Guarente, L. Transcriptional silencing and longevity protein Sir2 is an NAD-dependent histone deacetylase. *Nature*, 2000, 403, 795–799.
- [135] Haigis, MC; Guarente, LP. Mammalian sirtuins—emerging roles in physiology, aging, and calorie restriction. *Genes Dev.*, 2006, 20, 2913–2921.
- [136] Michishita, E; Park, JY; Burneskis, JM; Barrett, JC; Horikawa, I. Evolutionarily conserved and nonconserved cellular localizations and functions of human SIRT proteins. *Mol Biol Cell*, 2005, 16, 4623–4635.
- [137] Nakamura, Y; Ogura, M; Tanaka, D; Inagaki, N. Localization of mouse mitochondrial SIRT proteins: shift of SIRT3 to nucleus by co-expression with SIRT5. *Biochem Biophys Res Commun.*, 2008, 366, 174–179.
- [138] Imai, SI; Guarente, L. Ten years of NAD-dependent SIR2 family deacetylases – implications for metabolic diseases. *Trends Pharm Sci.*, 2010, 31, 212–220.
- [139] Du, J; Zhou, Y; Su, X; Yu, JJ; Khan, S; Jiang, H; Kim, J; Woo, J; Kim, JH; Choi, BH; He, B; Chen, W; Zhang, S; Cerione, RA; Auwerx, J; Hao, Q; Lin, H. Sirt5 Is an NAD-

- Dependent Protein Lysine Demalonylase and Desuccinylase. *Science*, 2011, 334, 806–809.
- [140] Patel, NV; Gordon, MN; Connor, KE; Good, RA; Engelman, RW; Mason, J; Morgan, DG; Morgan, TE; Finch, CE. Caloric restriction attenuates Abeta-deposition in Alzheimer transgenic models. *Neurobiol Aging*, 2005, 26, 995-1000.
- [141] Qin, W; Yang, T; Ho, L; Zhao, Z; Wang, J; Chen, L; Zhao, W; Thiyagarajan, M; MacGrogan, D; Rodgers, JT; Puigserver, P; Sadoshima, J; Deng, H; Pedrini, S; Gandy, S; Sauve, AA; Pasinetti, GM. Neuronal SIRT1 activation as a novel mechanism underlying the prevention of Alzheimer disease amyloid neuropathology by calorie restriction. *J Biol Chem.*, 2006, 281, 21745-21754.
- [142] Bouras, T; Fu, M; Sauve, AA; Wang, F; Quong, AA; Perkins, ND; Hay, RT; Gu, W; Pestell, RG. SIRT1 deacetylation and repression of p300 involves lysine residues 1020/1024 within the cell cycle regulatory domain 1. *J Biol Chem.*, 2005, 280, 10264–10276.
- [143] Luo, J; Nikolaev, AY; Imai, S; Chen, D; Su, F; Shiloh, A; Guarente, L; Gu, W. Negative control of p53 by Sir2alpha promotes cell survival under stress. *Cell*, 2001, 107, 137–148.
- [144] Michan, S; Sinclair, D. Sirtuins in mammals: insights into their biological function. *Biochem J.*, 2007, 404, 1–13.
- [145] Yeung, F; Hoberg, JE; Ramsey, CS; Keller, MD; Jones, DR; Frye, RA; Mayo, MW. Modulation of NF-kappaB-dependent transcription and cell survival by the SIRT1 deacetylase. *EMBO J.*, 2004, 23, 2369–2380.
- [146] Motta, MC; Divecha, N; Lemieux, M; Kamel, C; Chen, D; Gu, W; Bultsma, Y; McBurney, M; Guarente, L. Mammalian SIRT1 represses forkhead transcription factors. *Cell*, 2004, 116, 551–563.
- [147] Nemoto, S; Fergusson, MM; Finkel, T. SIRT1 functionally interacts with the metabolic regulator and transcriptional coactivator PGC-1{alpha}. *J Biol Chem.*, 2005, 280, 16456–16460.
- [148] Cohen, TJ; Guo, JL; Hurtado, DE; Kwong, LK; Mills, IP; Trojanowski, JQ; Lee, VM. The acetylation of tau inhibits its function and promotes pathological tau aggregation. *Nat Commun.*, 2011, 2, 252.
- [149] Donmez, G; Wang, D; Cohen, DE; Guarente, L. SIRT1 suppresses beta-amyloid production by activating the alpha-secretase gene ADAM10. *Cell*, 2010, 142, 320–332.
- [150] Theendakara, V; Patent, A; Peters Libeu, CA; Philpot, B; Flores, S; Descamps, O; Poksay, KS; Zhang, Q; Cailing, G; Hart, M; John, V; Rao, RV; Bredesen, DE. Neuroprotective Sirtuin ratio reversed by ApoE4. *Proc Natl Acad Sci USA*, 2013, 110, 18303-18308.
- [151] Min, SW; Cho, SH; Zhou, Y; Schroeder, S; Haroutunian, V; Seeley, WW; Huang, EJ; Shen, Y; Masliah, E; Mukherjee, C; Meyers, D; Cole, PA; Ott, M; Gan, L. Acetylation of tau inhibits its degradation and contributes to tauopathy. *Neuron*, 2010, 67, 953-966.
- [152] Du, LL; Xie, JZ; Cheng, XS; Li, XH; Kong, FL; Jiang, X; Ma, ZW; Wang, JZ; Chen, C; Zhou, XW. Activation of sirtuin 1 attenuates cerebral ventricular streptozotocin-induced tau hyperphosphorylation and cognitive injuries in rat hippocampi. *Age (Dordr)*, 2014, 36, 613-623.

- [153] Rodgers, JT; Lerin, C; Haas, W; Gygi, SP; Spiegelman, BM; Puigserver, P. Nutrient control of glucose homeostasis through a complex of PGC-1 alpha and SIRT1. *Nature*, 2005, 434, 113–118.
- [154] Lee, IH; Cao, L; Mostoslavsky, R; Lombard, DB; Liu, J; Bruns, NE; Tsokos, M; Alt, FW; Finkel, T. A role for the NAD-dependent deacetylase Sirt1 in the regulation of autophagy. *Proc Natl Acad Sci USA*, 2008, 105(9), 3374–3379.
- [155] Julien, C; Tremblay, C; Émond, V; Lebbadi, M; Salem, N; Bennett, DA; Calon, F. SIRT1 Decrease Parallels the Accumulation of tau in Alzheimer Disease. *J Neuropathol Exp Neurol.*, 2009, 68(1), 48.
- [156] Kumar, R; Chatterjee, P; Sharma, PK; Singh, AK; Gupta, A; Gill, K; Tripathi, M; Dey, AB; Dey, S. Sirtuin1: A Promising Serum Protein Marker for Early Detection of Alzheimer's Disease. *PLoS ONE*, 2013, 8(4), e61560.
- [157] Grubisha, O; Smith, BC; Denu, JM. Small molecule regulation of Sir2 protein deacetylases. *FEBS J.*, 2005, 272, 4607–4616.
- [158] Pallos, J; Bodai, L; Lukacsovich, T; Purcell, JM; Steffan, JS; Thompson, LM; Marsh, JL. Inhibition of specific HDACs and sirtuins suppresses pathogenesis in a Drosophila model of Huntington's disease. *Hum Mol Genet.*, 2008, 17, 3767–3775.
- [159] Anekonda, TS. Resveratrol—a boon for treating Alzheimer's disease? *Brain Res Rev.*, 2006, 52, 316–326.
- [160] Kim, D; Nguyen, MD; Dobbin, MM; Fischer, A; Sananbenesi, F; Rodgers, JT; Delalle, I; Baur, JA; Sui, G; Armour, SM; Puigserver, P; Sinclair, DA; Tsai, LH. SIRT1 deacetylase protects against neurodegeneration in models for Alzheimer's disease and amyotrophic lateral sclerosis. *The EMBO J.*, 2007, 26(13), 3169–3179.
- [161] Porquet, D; Griñán-Ferré, C2; Ferrer, I3; Camins, A1; Sanfeliu, C4; Del Valle, J5; Pallàs, M1. Neuroprotective role of trans-resveratrol in a murine model of familial Alzheimer's disease. *J Alzheimers Dis.*, 2014, 42(4), 1209–1220.
- [162] Shi, T; Wang, F; Stieren, E; Tong, Q. SIRT3, a mitochondrial sirtuin deacetylase, regulates mitochondrial function and thermogenesis in brown adipocytes. *J Biol Chem.*, 2005, 280, 13560–13567.
- [163] Anekonda, TS; Reddy, PH. Neuronal protection by sirtuins in Alzheimer's disease. *J Neurochem.*, 2006, 96, 305–313.
- [164] Millan, MJ. An epigenetic framework for neurodevelopmental disorders: from pathogenesis to potential therapy. *Neuropharmacology.*, 2013, 68, 2–82.
- [165] Jarome, TJ; Lubin, FD. Histone lysine methylation: critical regulator of memory and behavior. *Rev Neurosci.*, 2013, 24, 375–387.
- [166] Filipowicz, W; Jaskiewicz, L; Kolb, FA; Pillai, RS. Post- transcriptional gene silencing by siRNAs and miRNAs. *Curr Opin Struct Biol.*, 2005, 15, 331–341.
- [167] Landgraf, P; Rusu, M; Sheridan, R; Sewer, A; Iovino, N; Aravin, A; Pfeffer, S; Rice, A; Kamphorst, AO; Landthaler, M; Lin, C; Socci, ND; Hermida, L; Fulci, V; Chiaretti, S; Foà, R; Schliwka, J; Fuchs, U; Novosel, A; Müller, RU; Schermer, B; Bissels, U; Inman, J; Phan, Q; Chien, M; Weir, DB; Choksi, R; De Vita, G; Frezzetti, D; Trompeter, HI; Hornung, V; Teng, G; Hartmann, G; Palkovits, M; Di Lauro, R; Wernet, P; Macino, G; Rogler, CE; Nagle, JW; Ju, J; Papavasiliou, FN; Benzing, T; Lichter, P; Tam, W; Brownstein, MJ; Bosio, A; Borkhardt, A; Russo, JJ; Sander, C; Zavolan, M; Tuschl, T. A mammalian microRNA expression atlas based on small RNA library sequencing. *Cell*, 2007, 129, 1401–1414.

- [168] Schratt, GM; Tuebing, F; Nigh, EA; Kane, CG; Sabatini, ME; Kiebler, M; Greenberg, ME. A brain-specific microRNA regulates dendritic spine development. *Nature*, 2006, 439, 283–289.
- [169] Fiore, R; Khudayberdiev, S; Christensen, M; Siegel, G; Flavell, SW; Kim, TK; Greenberg, ME; Schratt, G. Mef2- mediated transcription of the miR379-410 cluster regulates activity-dependent dendritogenesis by fine-tuning Pumilio2 protein levels. *EMBO J.*, 2009, 28, 697–710.
- [170] Gao, J; Wang, WY; Mao, YW; Graff, J; Guan, JS; Pan, L; Mak, G; Kim, D; Su, SC; Tsai, LH. A novel pathway regulates memory and plasticity via SIRT1 and miR-134. *Nature*, 2010, 466, 1105–1109.
- [171] Cogswell, JP; Ward, J; Taylor, IA; Waters, M; Shi, Y; Cannon, B; Kelnar, K; Kemppainen, J; Brown, D; Chen, C; Prinjha, RK; Richardson, JC; Saunders, AM; Roses, AD; Richards, CA. Identification of miRNA changes in Alzheimer's disease brain and CSF yields putative biomarkers and insights into disease pathways. *J Alzheimers Dis.*, 2008, 14, 27–41.
- [172] Hebert, SS; Horre, K; Nicolai, L; Papadopoulou, AS; Mandemakers, W; Silahtaroglu, AN; Kauppinen, S; Delacourte, A; De Strooper, B. Loss of microRNA cluster miR-29a/b-1 in sporadic Alzheimer's disease correlates with increased BACE1/ beta-secretase expression. *Proc Natl Acad Sci. USA.*, 2008, 105, 6415–6420.
- [173] Soreq, H; Wolf, Y. NeurimmiRs: microRNAs in the neuroimmune interface. *Trends Mol Med.*, 2011, 17, 548-555.
- [174] Patel, N; Hoang, D; Miller, N; Ansaloni, S; Huang, Q; Rogers, JT; Lee, JC; Saunders, AJ. MicroRNAs can regulate human APP levels. *Mol Neurodegener.*, 2008, 3, 10.
- [175] Hebert, SS; Horre, K; Nicolai, L; Bergmans, B; Papadopoulou, AS; Delacourte, A; De Strooper, B. MicroRNA regulation of Alzheimer's Amyloid precursor protein expression. *Neurobiol Dis.*, 2009, 33, 422–428.
- [176] Long, JM; Lahiri, DK. MicroRNA-101 downregulates Alzheimer's amyloid-beta precursor protein levels in human cell cultures and is differentially expressed. *Biochem Biophys Res Commun.*, 2011, 404, 889–895.
- [177] Vilardo, E; Barbato, C; Ciotti, M; Cogoni, C; Ruberti, F. MicroRNA-101 regulates amyloid precursor protein expression in hippocampal neurons. *J Biol Chem.*, 2010, 285, 18344–18351.
- [178] Delay, C; Calon, F; Mathews, P; Hebert, SS. Alzheimer-specific variants in the 3' UTR of Amyloid precursor protein affect microRNA function. *Mol Neurodegener.*, 2011, 6, 70.
- [179] Schonrock, N; Ke, YD; Humphreys, D; Staufenbiel, M; Ittner, LM; Preiss, T; Gotz, J. Neuronal microRNA deregulation in response to Alzheimer's disease amyloid-beta. *PLoS One*, 2010, 5, e11070.
- [180] Zong, Y; Wang, H; Dong, W; Quan, X; Zhu, H; Xu, Y; Huang, L; Ma, C; Qin, C. miR-29c regulates BACE1 protein expression. *Brain Res.*, 2011, 1395, 108–115.
- [181] Boissonneault, V; Plante, I; Rivest, S; Provost, P. MicroRNA-298 and microRNA-328 regulate expression of mouse beta-amyloid precursor protein-converting enzyme 1. *J Biol Chem.*, 2009, 284, 1971–1981.
- [182] Faghihi, MA; Modarresi, F; Khalil, AM; Wood, DE; Sahagan, BG; Morgan, TE; Finch, CE; St Laurent III, G; Kenny, PJ; Wahlestedt, C. Expression of a noncoding RNA is

elevated in Alzheimer's disease and drives rapid feed-forward regulation of beta-secretase. *Nat Med.*, 2008, 14, 723–730.

- [183] Hebert, SS; Sergeant, N; Buee, L. MicroRNAs and the regulation of tau metabolism. *Int J Alzheimers Dis.*, 2012, 2012(9), 406561.
- [184] Lukiw, WJ; Alexandrov, PN. Regulation of complement factor H (CFH) by multiple miRNAs in Alzheimer's disease (AD) brain. *Mol Neurobiol.*, 2012, 46, 11–19.
- [185] Nunez-Iglesias, J; Liu, CC; Morgan, TE; Finch, CE; Zhou, XJ. Joint genome-wide profiling of miRNA and mRNA expression in Alzheimer's disease cortex reveals altered miRNA regulation. *PLoS One*, 2010, 5, e8898.
- [186] Geekiyanage, H; Jicha, GA; Nelson, PT; Chan, C. Blood serum miRNA: non-invasive biomarkers for Alzheimer's disease. *Exp Neurol.*, 2012, 235, 491–496.
- [187] Tan, L; Yu, JT; Liu, QY; Tan, MS; Zhang, W; Hu, N; Wang, YL; Sun, L; Jiang, T; Tan, L. Circulating miR-125b as a biomarker of Alzheimer's disease. *J Neurol Sci.*, 2014, 336, 52–56.
- [188] Lahiri, DK; Maloney, B; Zawia, NH. The LEARN model: an epigenetic explanation for idiopathic neurobiological diseases. *Mol Psychiatry*, 2009, 14, 992– 1003.

Chapter 63

EPIGENETIC MECHANISMS IN CARDIOVASCULAR DISEASES

***Mehmet Güven^{1,*}, PhD
and Bahadır Batar^{1,2}, PhD***

¹Department of Medical Biology, Cerrahpasa Medical Faculty,
Istanbul University, Istanbul, Turkey

²Department of Molecular Virology, Immunology and Medical Genetics,
Comprehensive Cancer Center,
Ohio State University Wexner Medical Center,
Columbus, OH, US

ABSTRACT

Cardiovascular disease (CVD) is not a single condition, but an umbrella term used to describe a range of common diseases affecting the heart and the circulatory system. The term commonly includes diseases of the cardiac muscle and of the vascular system supplying the heart, brain, and other vital organs. Many of these conditions can be life-threatening. CVD is the leading cause of death in the world, affecting all populations, irrespective of demographic or socioeconomic differences and is responsible for one third of all deaths.

In the present chapter, we focus on the epigenetic control of embryonic cardiac development and the role of epigenetic mechanisms in CVD observed from results in human, animal and cell culture studies' approaches. We discuss the main epigenetic mechanisms involved in heart development and major CVDs such as coronary heart disease (CHD), heart failure (HF), myocardial infarction (MI), hypertension, stroke, arrhythmias, cardiomyopathy and cardiac hypertrophy (CH). Additionally, this chapter also focuses on the epigenetic modifiers that are involved in the development of CVD, and the potential utility of epigenetics-based therapeutic strategies in CVD.

* Corresponding Author's Email: mgguven@istanbul.edu.tr.

Keywords: epigenetic, cardiovascular disease, coronary heart disease, heart, post-translational modification

ABBREVIATIONS

11 β HSD2	11 β -hydroxysteroid dehydrogenase 2
5-azaC	5-azacytidine
5hmC	5-hydroxy-methylcytosine
5mC	Methylation of cytosine
AMOTL2	Angiomotin like protein-2
ANF	Atrial natriuretic factor
ANRIL	Antisense non-coding RNA in the <i>INK4</i> locus
ARH	
GAP24	Rho-GTPase activating protein-24
AT1bR	Angiotensin type 1b receptor
BAF	BRG1-associated factor
BNP	Brain Natriuretic Peptide
BRG1	Brahma-related gene-1
BRM	Brahma
Bvht	Braveheart
CARD8	Caspase recruitment domain 8
Cbp	Creb-binding protein
CH	Cardiac hypertrophy
CHD	Coronary heart disease
CHD7	Chromodomain helicase DNA-binding 7
COL15A1	Collagen type XV A1
CRB	CREB binding protein
CVD	Cardiovascular disease
CYP1A1	Cytochrome P450A1
DAC	5-aza-2-deoxycytidine
DGCR8	DiGeorge syndrome critical region gene 8
DNMT	DNA methyltransferase
DOT1L	DOT1-Like Histone H3K79 Methyltransferase
DOT1L	Telomeric silencing-1-like
EGFR	Epidermal growth factor receptor
ENaC	Epithelial sodium channel
ER	Estrogen receptor
ESC	Embryonic stem cell
F2RL3	Coagulation factor II (thrombin) receptor-like 3
Fendrr	Foxf1 adjacent non-coding developmental regulatory RNA
Gata4	GATA binding protein 4
GR	Glucocorticoid
H3K4	Histone H3 lysine 4
HAND1	Heart and neural crest derivatives expressed 1

HAT	Histone acetyltransferase
Hcy	Homocysteine
HDAC	Histone deacetylase
HF	Heart failure
HHcy	Hyperhomocysteinemia
IL	Interleukin
lncRNA	Long non-coding RNA
Irx4	Iroquois homeobox 4
IUGR	Intrauterine growth restriction
Jarid2	Jumonji AT rich interactive domain 2
KDM1A	Lysine-specific histone demethylase 1A
LINE-1	Long interspersed nucleotide elements-1
LIPCAR	Long intergenic noncoding RNA predicting cardiac remodeling
LSD-1	Lysine (K)-specific demethylase 1A
MBD	Methyl-CpG binding domain protein
MeCP2	Methyl-CpG-binding protein 2
MED13	Mediator complex subunit 13
MEF2C	Myocyte enhancer factor 2C
MHC	Myosin heavy chain
MI	Myocardial infarction
MIAT	Myocardial infarction-associated lncRNA transcript
MiRNA	MicroRNA
MLC2V	Myosin light chain 2 ventricular
Msx2	Msh homeobox 2
MyoD	Myogenic differentiation
NAD	Nicotinamide adenine dinucleotide
NF- κ B	Nuclear factor- κ B
NKCC1	Na ⁺ /K ⁺ /2Cl ⁻ cotransporter 1
PBAF	Polybromo-associated BAF
PBMC	Peripheral blood mononuclear cell
PPAR	Peroxisome proliferator-activated receptor
PRC2	Polycomb silencing complex 2
RNAPII	RNA polymerase II
ROS	Reactive oxygen species
RXR α	Retinoid X receptor-alpha
sACE	Somatic angiotensin-converting enzyme
SAM	S-adenosylmethionine
SMC	Smooth muscle cell
SMYD1	SET and mynd domain containing 1
SRF	Serum response factor
SWI/SNF	SWItch/Sucrose non-fermentable
TET	Ten-eleven translocation
TNF	Tumor necrosis factor
VANGL2	Van Gogh-like 2
VDR	Vitamin D Receptor
VSMC	Vascular smooth muscle cell

1. EPIGENETIC MECHANISMS

Epigenetics refers to genetic factors that change the phenotype of an organism or the biological functions without changing the DNA sequence. DNA base modifications, post-translational modifications of histone proteins, and RNA-based mechanisms are major processes involved in epigenetics.

Among the epigenetic factors, methylation of cytosine (5mC) bases within the DNA is the essential pathway by which DNA is modified that is relevant for epigenetic changes. DNA methylation usually occurs within the context of CpG dinucleotide sequences (CpG-sites) [1]. The main function of DNA methylation is to modulate the expression of the genetic information by modifying the accessibility of DNA to the transcription binding factors. DNA methylation plays a key role in X-chromosome inactivation, genomic imprinting, and tissue-restricted gene expression during development and differentiation [2]. The process of DNA methylation is conducted by enzymes called DNA methyltransferases (DNMTs) [1]. Methylation in DNA can be actively removed through oxidative demethylation by the Ten-eleven translocation (TET) family (TET1, TET2 and TET3) [3]. TET protein family is involved in oxidizing of 5mC to 5-hydroxymethylcytosine (5hmC). 5hmC is another important epigenetic modification on DNA in mammalian cells.

Although histones are the proteins that make up nucleosomes, the post-translational modifications of these proteins are another epigenetic mechanism affecting gene expression. “Histone Code Hypothesis” proposes that various combinations of histone modifications may control chromatin structure and transcriptional status. Among histone modifications defined by the histone code, the acetylation, methylation and phosphorylation are important mechanisms of epigenetic regulation [4]. Histone acetylation is catalyzed by proteins known as histone acetyltransferases (HATs) and histone deacetylation is carried out by histone deacetylases (HDACs) [5]. Similarly, histone methylation is regulated by histone methyltransferases and histone demethylases. Histone phosphorylation is also catalyzed by several distinct kinases that are mostly specific for individual histone residues [6].

RNA-based mechanisms are the most recently defined of the three major types of epigenetic regulation. RNA-based mechanisms exclusively focus on microRNAs (miRNAs), short non-coding RNAs, which modulates post-transcriptional repression of epigenetic regulators of gene expression. MiRNAs are endogenously produced RNA molecules that are approximately 22 nucleotides in length. At least 30% of human genes are thought to be regulated by miRNAs [7]. MiRNAs are transcribed by miRNA genes in a primary transcript (pri-miRNA). Drosha-DGCR8 complex processes pri-miRNA as precursor-miRNAs (pre-miRNAs) in the nucleus. Pre-miRNAs are transported into the cytosol where Dicer cleaves the pre-miRNAs to generate a short double-stranded miRNA-miRNA duplex [8, 9]. Mature miRNA targets to complementary mRNA sequence to downregulate the translation of mRNA. Long non-coding RNAs (lncRNAs) which are transcripts 200 nucleotides in length could promote gene activation and inhibition by interacting with epigenetic regulators. lncRNAs are transcribed by Polymerase II enzyme, capped of 5' end, alternatively spliced and polyadenylated.

2. EPIGENETIC MECHANISMS IN HEART DEVELOPMENT

Heart being one of the first organs to develop during embryogenesis has a complex development process which involves coordinated cellular proliferation, migration, differentiation, programmed cell death and structural remodeling including looping and septation. Nucleosome remodeling, DNA methylation, and histone modification are processes controlled by enzymes that are dependent on the correct activation of a gene program which is necessary for the accurate growth of the mammalian heart during embryonic period and the terminal differentiation in the adult. Defects in the aforementioned processes result in deleterious phenotypes like septal defects, atrioventricular malformations, or developmental absence of right or the left heart and may end with mortality anytime during embryogenesis [10]. Epigenetic mechanisms have been implicated during the development of the heart in both directly by regulation of gene expression patterns as well as indirectly by modulating the expression of transcription factors.

Chromatin remodeling complexes catalyze the relocation of nucleosomes to modulate the accessibility of transcription factors and other regulatory proteins to DNA. There are multiple distinct families of chromatin-remodeling complexes [11]. For example, the SWItch/Sucrose Non-Fermentable (SWI/SNF) complex is an evolutionarily conserved multisubunit chromatin remodeling complex that are critical for differentiation and proliferation. SWI/SNF family exists in several subgroups and each subgroup contains one of two related ATPases, *Brahma-related gene-1* (BRG1) or *Brahma* (BRM), and BRG1-associated factors (BAFs) per complex [12]. BAF and PBAF (Polybromo-associated BAF), SWI/SNF members, are structurally related but functionally distinct. As shown by multiple studies, when a cell differentiates from a pluripotent stem cell to a multi-potent progenitor leading ultimately into a terminally differentiated state the configuration of SWI/SNF complexes are changed. Brg1 null mouse embryos die shortly after E11.5 day post-conception when cardiomyocyte expansion and maturation begins [13]. This fact proves the importance of Brg1 during cardiac development. Thinning of the compact myocardium and failure to form an inter-ventricular septum are the cardiac defects seen in these embryos. These defects are harmonious with a loss of cardiac factor Bmp10 expression and up-regulation of cell cycle inhibitor p57 (Kip2) thus portraying a decrease in cardiac myocyte proliferation [14].

The BAF180 subunit of the PBAF chromatin remodeling complex facilitates the ligand dependent transcription of nuclear receptor genes such as Retinoid X receptor-alpha (RXRa), Vitamin D Receptor (VDR), and peroxisome proliferator-activated receptors (PPARs) in vitro [15]. BAF180 knockout mice are prone to be a mortality at around the 15th post-conceptual day. They display severe hypoplastic ventricular development. Ventricular wall thinning with ventricular septal defect is observed histologically in the hearts of BAF180 knockout mice [16]. BAF180 is indispensable for coronary vessel formation. During the embryogenesis of the heart BAF180 is expressed sufficiently especially in the pro-epicardium and epicardium [16].

There is a relationship between cardiac-specific transcription factors and histone modifiers that constitutes a mechanism for target specificity. During the early stages of cardiac development a key transcription factor called Nkx2.5 regulates the development of myocardial precursors [17]. Precise regulation of cardiac gene expression is possible with multiple layers of interactions between Nkx2.5 and epigenetic factors such as histone demethylase/methyltransferases. Defects in either Nkx2.5 or epigenetic factors that interact with Nkx2.5

have the potential to end up with congenital heart disease. Modulation of the expression of *Jumonji*, *AT Rich Interactive Domain 2* (*Jarid2/Jmj*), and a component of a protein complex that contains histone demethylase and methyltransferases activities, in the second heart field by *Nkx2.5* itself might contribute to the epigenetic regulation [18]. *Nkx2.5* mutations have been linked with various septation and atrioventricular conduction defects in humans. Patients with congenital heart disease often showed decreased DNA binding and transcriptional activation by mutant *Nkx2.5* when compared with wild type. With loss of *Nkx2.5*, the expression of *Atrial natriuretic factor* (*ANF*), brain natriuretic peptide, myosin light chain 2 ventricular (*MLC2V*), Myocyte Enhancer Factor 2C (*MEF2C*), *Heart And Neural Crest Derivatives Expressed 1* (*HAND1*), *Msh Homeobox 2* (*Msx2*), and *iroquois homeobox 4* (*Irx4*) was significantly lost [19].

Congenital heart defects are among the most common of all medically significant birth defects. Several large population-based studies estimate the prevalence of congenital heart defects to range from 3 to 6 cases per 1000 live births [20]. Some syndromes, including Wolf–Hirschhorn syndrome, Holt–Oram syndrome, CHARGE syndrome, DiGeorge syndrome are related to congenital heart defects caused by the defects in epigenetic regulation. For example, CHARGE syndrome is defined by coloboma of the iris or retina, heart defects, atresia of the choanae, retardation of growth and/or development, genital, and ear abnormalities. Most of the children (71%) diagnosed with CHARGE have genetic mutations in the chromodomain helicase DNA-binding 7 (*CHD7*) gene. *CHD7* gene is a member of the chromo domain helicase DNA binding family of ATP-dependent chromatin modifiers [21]. *CHD7* is associated with all three forms of methylated histone H3 lysine 4 (*H3K4*), but most robustly with the mono-methylated (*H3K4me1*) and di-methylated (*H3K4me2*) sites as shown by analysis [19]. *CHD7* binding sites were found to be located further away from the transcriptional start sites, overlapping with DNase hypersensitivity sites. This reveals that *CHD7* may bind to and regulate putative enhancer regions or insulator elements for the deposition of epigenetic markers that define lineage specification [22].

2.1. DNA Base Modifications

Gametogenesis, hematopoiesis and stem cell differentiation, as well as various disease processes are several developmental processes in which DNA methylation patterns and dynamics have been described. The role of DNA methylation in regulation of heart development is under-studied. The role of DNA methylation in outflow tract septation through promoter regulation of the Van Gogh-like (*VANGL2*) gene has been suggested by recent scientific data [23]. *VANGL2* is essential for the polarization of cells at the outflow tract. The homozygous mutation of the gene in mice results in a phenotype resembling Tetralogy of Fallot (overriding aorta, double-outlet right ventricle and ventricular septal defects). *VANGL2* promoter methylation has been found to be elevated in patients with Tetralogy of Fallot when compared with healthy controls [24]. Genomic location within cells, cell types within the developing heart, as well as across various stages of development are factors which vary the role of DNA methylation in heart development.

Studies in embryonic stem cell (ESC) or animal models are the sources that provide our knowledge on DNA methylation. Inhibition of DNA methylation in ESCs by 5-azacytidine (5-azaC) has been shown to induce cardiomyocyte differentiation [25]. However, recent studies

on murine embryonic heart development have shown that heart development is associated with increases and decreases (greater increases than decreases) in DNA methylation, especially in cardiac-specific genes [26]. Studies with mice lacking key enzymes involved in the regulation of DNA methylation and demethylation were the first to suggest the gravity of DNA methylation in the development of the heart. The balance between DNA methylation and demethylation is imperative for proper mammalian development as portrayed by these results. The absence of DNMT1 led to global hypomethylation and early embryonic death as shown in these studies [27, 28]. DNMT3A-null mice survive until birth but die around 4 weeks of age, whereas DNMT3B-null mice die during gestation [29]. These findings are consistent with studies of postnatal cardiac development in rats, showing increases in DNA methylation in cardiomyocytes, as well as increased DNMT1, Methyl-CpG binding domain protein 1-3 (MBD1-3) and methyl-CpG-binding protein 2 (MeCP2), over the weeks and months following birth [30].

2.2. Histone Code

Epigenetic histone modifications have important roles in both normal and abnormal development as suggested by recent evidence. Factors that affect cardiac development are histone methylation and demethylation, acetylation and deacetylation status.

The status of histone methylation affects cardiac development. Being a histone methyltransferase, SET And MYND Domain Containing 1 (SMYD1) has been proposed to play a role in the development process of the heart. Mice lacking SMYD1 exhibit hypoplastic right ventricles and embryonic lethality [31]. In addition, Hand2 and Irx4 transcription factors, which are essential for normal cardiomyocyte development are found to be reduced in these mice [31, 32]. These changes suggest that SMYD1-mediated histone methylation is prerequisite for the expression of these essential cardiac transcription factors. Histone demethylase Jumonji family members have been proposed to have a role in septation. In the case of their deletion, ventricular septal defects and outflow tract defects are observed [33, 34]. Lysine (K)-specific demethylase 1A / Lysine-specific histone demethylase 1A (LSD-1/KDM1A) is another lysine demethylase that plays a role in the formation of ventricular septal defects [35].

The Creb-binding protein (Cbp) and p300 are two HATs and play critical roles in physiological and pathological growth of cardiac myocytes [36]. Cardiovascular defects and embryonic lethality are observed in p300 knockout mice, which die between days 9 and 11.5 of gestation and show reduced expression of muscle structural proteins such as β -MHC and α -actinin, as well as cardiac structural defects and reduced trabeculation [37, 38]. The potential connection between p300 regulatory complex and the actions of cardiac signals has led to the identification of transcription factors such as GATA binding protein 4 (Gata4), Mef2d, and myogenic differentiation (MyoD), as well as Mef2c involvement with transcriptional programs that result from signaling cues that control pathological gene expression in the heart [39].

Effects of histone deacetylation have been investigated in the development process of the heart. HDAC1 and 2 seem to have prominent roles in cardiac development. Double knockout mice show cardiac defects. Mice lacking HDAC1 and 2 are lost in the early periods of neonatal life. Their mortality is due to cardiac arrhythmias and dilated cardiomyopathy [40]. Being associated especially with cardiomyocyte metabolism and proliferation, HDAC3 has indeed

been thought to have a role in heart development. HDAC3's loss leads to impairment of glucose metabolism, thus provides an inclination towards an increase in lipid metabolism [41]. Overexpression of HDAC3 promotes proliferation of cardiomyocytes by the reduction of cell cycle inhibitors. As a result of these data it can be concluded that histone modifications lead to major alterations in heart development.

2.3. MiRNAs and lncRNAs

Loss and gain-of function studies demonstrated that miR1 and miR133 families are an essential miRNAs in the cardiac development [42]. In mice, mostly expressed miRNAs in the heart are miR1 (miR1-1, miR1-2, and miR206) and miR133 (miR133a-1, miR133a-2, miR133b) [43]. MiR133 targets cyclin D2 and serum response factor (SRF) and negatively controls cardiomyocyte proliferation [44]. In addition, miR208 and miR499 are highly expressed in the heart [45]. Also, it is found decrease levels of miR133 in the infarcted areas of the heart [46].

MiR1 promotes cardiac specification [47]. It has been shown that increased miR1 expression was associated with decreased ventricular cardiomyocyte proliferation [42]. It is suggested that mir208a regulates mediator complex subunit 13 (MED13) expression and controls systemic energy homeostasis [48]. MiR499 enhances myocyte differentiation [49]. In addition, miRNAs could control cardiovascular differentiation [50].

Recent studies suggested the importance of lncRNAs as regulator molecules of transcriptional regulation in cardiovascular development. Braveheart (Bvht), one of the lncRNAs, is expressed in the early stages of embryonic heart development [51]. Bvht interacts with the epigenetic silencing complex, polycomb silencing complex 2 (PRC2), and affect the expression of critical genes during early cardiac development [52, 51]. PRC2 contains trimethylates histone H3 at Lys 27 that promotes tissue-specific differentiation [53, 54]. PRC2 has an important function during early and late cardiac development and homeostasis [55].

Another lncRNA, Foxf1 adjacent non-coding developmental regulatory RNA (Fendrr), is exclusively expressed in mesoderm and plays a role in the development of mouse heart [43]. Fendrr modulates the chromatin signatures of genes which forms and differentiates the lateral mesoderm [52]. Fendrr may also control the activating of HE3K4me3 mark on promoters. Several studies suggested that Fendrr inactivation causes to defects in vascularization [55].

Discovery of Bvht and Fendrr functions support the role of lncRNAs in cardiac development.

3. Epigenetic Mechanisms in CVDs

Our current scientific knowledge cannot fully describe the complex pathophysiology of CVD yet. Epigenetics provides a different path to approach human disease. The dynamic nature and response to environmental cues of epigenetic processes constitute a plan for apprehending the constraints to CVD research which focused solely on the static DNA code.

3.1. DNA Base Modifications

Atherosclerosis is characterized by chronic inflammation of the arterial wall. When atherosclerosis affects the arteries of the heart, it's called CHD. CHD is the most common form of CVD and also called coronary artery disease. Differentially methylated CpGs association with the onset of atherosclerosis onset together with endothelial and smooth muscle functionality was shown in a DNA methylation map of human atherosclerosis. In the atherosclerotic lesions, a gain of DNA methylation was observed. This process strikes the opportunity to develop DNA demethylating agents for therapeutic aims. The global DNA methylation profile of peripheral blood leukocytes may produce a suitable biomarker for increased CHD risk [56].

Homocysteine (Hcy) is one of the known risk factors for atherosclerosis and CVD. Its action may be caused by an epigenetic mechanism involving methylation. Hcy is biochemically linked to the principal epigenetic tag found in the DNA. High levels of Hcy in circulation are a risk factor for CVD. Nonetheless, in recent clinical trials where folate and/or other vitamin B therapies were used to lower Hcy; cardiovascular event rates did not decrease, making Hcy's direct causative role in vascular disease controversial [57]. In a lot of *in vivo* studies it was suggested that Hcy levels may modulate global DNA methylation. For example, in healthy human subjects, increased levels of plasma Hcy were linked with both increased S-adenosylhomocysteine concentrations and DNA hypomethylation in lymphocytes [58]. Vascular complications associated with high levels of circulating Hcy was partially caused by DNA hypomethylation in several studies [59]. Increased levels of plasma Hcy was observed with decreased DNA methylation in CVD patients. This situation supports the role for hyperhomocysteinemia (HHcy) in the modification of epigenetic mechanisms. In a study done by Ingrosso and colleagues increased levels of Hcy together and reduced global lymphocyte DNA methylation pattern was reported [60]. After the reduction of the plasma Hcy levels with folate administration both global DNA methylation patterns and allelic gene expression were normalized. Data suggests that HHcy is linked with reduction in global levels of CpG methylation and gene-specific methylation of promoters [61, 62]. Changes in the expression of specific genes are seen with these alterations. The increased expression of the nuclear factor- κ B (NF- κ B) transcription factor, interacts with the IRE1 and c-Jun NH2-terminal kinase family of proteins [63]. This increased expression may cause inflammatory-mediated vascular damage [64]. With the finding of recent data it was suggested that that alterations in lipid metabolism may play a role in vascular pathology associated with HHcy [64, 65]. Many studies propose that epigenetics may play a role in these processes. In mice with Hhcy, it was suggested that changes in DNA methylation have might be the potential mechanism for altered *apoA-I* and *apoA-IV* gene expression [66, 67]. Hcy is significantly and inversely correlated with HDL-bound cholesterol and apoA-I in both human and murine models of Hhcy [68].

In vascular smooth muscle cell (VSMC) phenotypic modulation in human atherosclerosis, certain atheroprotective genes like *ER α* and *ER β* (*ESR1* and *ESR2*, respectively) which encode estrogen receptors (ERs) are hypermethylated [69]. ERs are found in the coronary arterial walls on both the smooth muscle cells (SMCs) and the endothelial cells (ECs). Their functions protect against atherosclerosis, especially in CHD. Deficiencies in *ER α* give way to accelerated atherosclerosis in humans [70]. Hypermethylation of *ESR1* was also seen in *in vitro* senescing ECs and SMCs, together with *ESR2* hypermethylation. With the light of these findings epigenetic changes in both *ESR1* and 2 can influence vascular aging and atherosclerosis.

Demethylation using 5-aza-2-deoxycytidine (DAC) could enhance the expression of silenced genes. These treatments increased the expression of ER α and ER β in normal SMCs and ECs. DAC and/or trichostatin A treatments showed low toxicity in these cells [71]. Silencing of ERs in women, with epigenetic changes, might explain the failure of estrogen therapy to exert a cardioprotective effect. The combined use of epigenetic therapy and hormone replacement therapy could be suitable for the prevention of CVDs [71, 72].

Zhu and colleagues portrayed that methylation of the monocarboxylate transporter gene suppresses its transcription leading to the passage of smooth muscle cells and the progression of atherosclerosis in cultured tissues of aorta and coronary arteries with varying degrees of atherosclerosis [73]. The risk of vascular damage may increase with the loss of the atheroprotective factors expressed by these genes.

In patients with atherosclerosis, global DNA methylation decreases in the peripheral blood cells [74]. The era of DNA methylation in atherosclerosis was uncovered by two different groups [75, 76]. They portrayed coexistence of hypomethylation with atherosclerotic lesions in mice and rabbits. It was shown in other studies that hypomethylation could be detected at the early stages of atherosclerosis, before the appearance of the anatomical manifestations of the disease [77]. Baccarelli and colleagues revealed an association between hypomethylation of long interspersed nucleotide elements-1 (LINE-1) and higher incidence of ischemic heart disease [56]. Friso and colleagues showed that hypomethylation of the promoter of coagulation factor VII was linked with coronary artery disease in patients, who are angiographically proven to have coronary lesions [78].

Heart failure is a condition caused by the heart failing to pump enough blood around the body at the right pressure. The etiology of HF is quite complex depending on genetic predisposition and multiple environmental factors. Development of HF depends on several molecular and cellular mechanisms. These include reprogramming of the expression of certain critical cardiac genes. Downregulation of the alpha-myosin heavy chain (alpha-MHC) gene and sarcoplasmic reticulum Ca⁺ ATPase genes and reactivation of specific fetal cardiac genes such as atrial natriuretic factor and brain natriuretic peptide are examples of these cardiac gene alterations [79].

Changes in patterns of methylation and extent of regulatory regions of genes is known to modify their expression which correlates with status of HF. Without respect to the etiology of HF there are three angiogenesis-related genes that are found to have differential methylation. Hypermethylation of the 5'-regulatory region of platelet endothelial cell adhesion molecule-1 and hypomethylation of the angiomin like protein-2 (AMOTL2) are noted in HF [80]. These are closely linked with reduced expression of these genes. Hypermethylation within the Rho-GTPase activating protein-24 gene (ARHGAP24) in failing hearts is associated with elevated expression of this gene [80]. Movassagh and colleagues showed that three angiogenesis-related genes underwent differential methylation, irrespective of the etiology in patients with HF [80]. *AMOTL2* gene was hypomethylated whereas the 5' promoter region in the platelet/endothelial cell adhesion molecule gene and the gene body of *ARHGAP24* were hypermethylated [80]. The modified epigenomics of the three genes in endstage HF may reflect common epigenetic pathways in heart remodeling and vasculature. Dilated cardiomyopathy is seen in 33% of cases of HF. Nguyen and colleagues in a mouse model proved that cardiac specific knockout of telomeric silencing-1-like (DOT1L), the gene encoding for the HKMT Dot1L which catalyzes H3K79 methylation in mammals that gave way to the appearance of a phenotype similar to that seen in dilated cardiomyopathy [81].

Hypertension is a multifactorial disease whose pathogenesis is underlined by several mechanisms due to the complex interplay of genetic and environmental factors. Epigenetic mechanisms play a role in the pathogenesis of hypertension. Global 5mC levels of peripheral blood cells in patients with essential hypertension were found to be lower than healthy controls. The levels of 5mC correlates with the the stage of hypertension in these patients [82]. Valinluck et al. reported that 5hmC at CpG sites was discovered to suppress the binding of MeCP2 [83]. MeCP2 is a transcriptional repressor which probably has a potential role of 5hmC in gene expression regulation [83]. 5hmC may have a role in mediation of fetal heart and lung growth [84].

The role of gene-specific DNA methylation in the pathogenesis of hypertension has been recently proposed. 5hmC has a role in determining tissue-specific gene expression and regulating the expression of genes involved in cellular differentiation and pluripotency [84, 85]. In patients with essential hypertension, promoter methylation of 11 β HSD2, the gene encoding 11 β -hydroxysteroid dehydrogenase 2 (11 β HSD2), has been shown to be increased in the peripheral blood mononuclear cells [82]. 11 β HSD2 breaks down cortisol to its biologically active form, cortisone. Reduction in the activity of 11 β HSD2 promotes cortisol-mediated activation of mineralocorticoid receptors thus end up in increased blood pressure [82]. An increase in the risk of development of adult type hypertension and an increase in promoter methylation at hydroxysteroid (11-beta) dehydrogenase 2 (HSD11B2) site were observed in an intrauterine growth restriction (IUGR) rat model of hypertension [86]. Methylation levels at HSD11B2 promoter sites were significantly higher in IUGR newborns when compared with controls. This was related to lower HSD11B2 expression [87]. The pattern of methylation at different CpG sites were in conjunction with birth weight and ponderal index which are measures of fetal growth [87]. Methylation of the promoter at HSD11B2 site may associated with development of IUGR and the possible risk of future development of hypertension [87]. Another human study suggests that promoter methylation of HSD11B2 gene is closely linked to hypertension [88]. Methylation at HSD11B2 promoter in DNA of peripheral blood mononuclear cell (PBMCs) of hypertensive patients is inversely related to the function of the enzyme [88].

Somatic angiotensin-converting enzyme (sACE) is another major mediator of blood pressure regulation which is influenced by DNA methylation in human cell lines [89]. With the combination of knowledge from these it can be suggested that aberrant DNA methylation could lead to the formation of essential hypertension in a gene-specific manner. Overall reduction in global methylation is indicative of the progression of disease.

The role of membrane transporter genes for the known role of alterations in ion flux mechanisms in the pathogenesis of hypertension have been also evaluated. Promoter methylation of the Na⁺/K⁺/2Cl⁻ cotransporter 1 (NKCC1) was studied by Lee et al. [90]. NKCC1 is a gene that encodes a solute carrier which is responsible for the transportation of sodium, potassium, and chloride through the cellular membrane. In spontaneously hypertensive rats, the methylation status of NKCC1 promoter were measured in the aorta and the heart [90]. In the study it was concluded that in a spontaneously hypertensive rodent model NKCC1 expression is upregulated by a mechanism induced by gene promoter hypomethylation [90].

Cardiomyopathies are defined as diseases of the myocardium associated with cardiac dysfunction. There are two types of cardiomyopathy: primary (idiopathic) or secondary to ischemia or myocardial infarction. They can also be defined on the type of remodeling. Ventricular dilation and hypertrophy take place as the patient develops HF.

DNA methylation patterns in the hearts of patients undergoing cardiac transplantation for primary or secondary cardiomyopathy were investigated in two current studies [91, 92]. Genomic profiling of DNA methylation in end-stage HF patients who have primary or secondary cardiomyopathy showed significant hypomethylation at the promoter regions and hypermethylation at the gene body regions when compared with healthy controls with normal hearts. There was no statistically significant difference in intergenic regions, including enhancer regions [92]. Haas et al. studied patients with primary cardiomyopathy and found that there was statistically significant difference in differential methylation in pathways related to cardiac disease, as well as differential methylation of several key genes when compared with healthy controls with normal hearts [91].

3.2. Histone Code

NOS3 is the well-characterized endothelial gene that has cardiovascular implications and it is controlled by histone modification. eNOS is a protein which is coded by the NOS3 gene. It catalyzes the formation of NO from L-arginine in blood vessels [93]. NO is a factor responsible for vasodilation. It has a major part in the regulation of vascular tone and protection against atherosclerosis [93]. The NOS3 gene is transcriptionally active in ECs and it is transcriptionally suppressed in VSMCs under normal conditions [94]. Acetyl H3K9, acetyl H4K12, and methyl H3K4 at the NOS3 proximal promoter site in ECs are activating histone modifications. When they are present they allow the recruitment of RNA polymerase II (RNAPII) [94]. In VSMCs, eNOS expression is suppressed by the absence of activating histone modifications at the NOS3 proximal promoter. It is further suppressed by the presence of cytosine methylation and MeCP2 recruitment [94]. The expression of eNOS is controlled by environmental stimuli. If short term hypoxia is experienced by the ECs, a reduction in the expression of eNOS is observed [95]. Significant reduction in NOS3 transcription observed in the ECs when faced with hypoxia is considered to happen via histone eviction, including processes that are in relation with transcriptional activation (acetyl H3K9, methyl H3K4, acetyl H4K12), at the NOS3 proximal promoter site [95]. H3 and H4 activating acetylation marks are activated by laminar shear stress. This is partially accomplished by the p300 HAT complex, at a shear stress reporter element in the human eNOS promoter [96]. The fusion of the activating acetylation marks via laminar shear stress provides for the chromatin at the eNOS shear stress reporter element site to stay in an open, transcriptionally-accessible state, permitting eNOS expression.

The activity of HDAC seems to have a major part in the prediction of the degree of myocardial ischemia and reperfusion damage especially after an MI [97]. There is proof that myogenesis and angiogenesis are stimulated in embryonic stem cell cultures with the inhibition of HDACs [98]. In vivo studies in mouse portrayed that by prevention of myocardial remodeling and a reduction in myocardial and serum tumour necrosis factor alpha, inhibition of HDAC improved functional myocardial recovery after MI [98]. Increases in the formation of myocytes and microvessels in the heart were seen after the inhibition of HDAC. After MI there is loss of myocardial performance due to stimulation of angiogenesis [98]. With the light of these findings it can be said that HDAC inhibition can make a reduction in the loss in myocardial performance by inhibiting angiogenesis [97]. The administration of the pharmacologic HDAC inhibitor TSA to atherosclerosis-prone Ldlr (-/-) mice exacerbated

neointimal lesions [99]. With better understanding of epigenetic pathways with the help of animal models of atherosclerosis there is more to be discovered.

Curcumin (diferuloylmethane) is a polyphenol present in a curry spice. It is p300 HAT inhibitor which has a wide variety of molecular targets including transcription factors, growth factors and their receptors, cytokines, enzymes, and genes regulating cell proliferation and apoptosis. Curcumin has cardioprotective function [100]. In studies done in patients with atherosclerosis and healthy controls, its administration decreases LDL levels, increases high-density lipoprotein (HDL) levels [101, 102]. Moreover, both in a rat model of HF and primary cultured rat cardiac myocytes and fibroblasts, curcumin inhibits p300 HAT activity thus preventing ventricular hypertrophy, and preserving systolic function in rat model of HF and primary cultured rat cardiac myocytes, fibroblasts [103]. In the rat cardiomyocytes, curcumin is thought to act by inhibition of histone acetylation and hypertrophy-responsive transcription factors (e.g., GATA4), and by disruption of p300/GATA4 complex [104].

In the experimental animal CH models, CH was found to be closely associated with histone acetylation. HATs and HDACs are both playing major roles in this process. It has been shown that hypertrophic growth of cardiomyocytes could be induced by the depending on their HAT activity both overexpression of transcriptional co-activators CREB binding protein (CRB) or p300 itself may induce hypertrophy in cardiomyocytes [105]. Phenylephrine-induced hypertrophic cell growth may be suppressed by the suppression of these co-activators. Catalytic activity of p300 or overexpression of CBP may give rise to hypertrophy. On the contrary mutant forms of these without HAT activity do not end up with the same effects [105]. However, the activity of HDACs' was found to be involve both prohypertrophic and antihypertrophic pathways leading to controversial data. Class IIa type HDACs were discovered to inhibit CH [106]. However, in other animal in vivo studies HDAC inhibitors were found to have a protective role against hypertrophy. In the study by Liu and colleagues done on cardiac hypertrophy model in transgenic mice it was portrayed that injection of a specific HDAC inhibitor reverses atrial fibrosis and diminishes atrial fibrillation vulnerability following an electrical stimulation [107]. Deletion of both HDAC-1 and HDAC-2 from the myocardium lead to early rodent mortality from fetal arrhythmia in other experimental models [40]. Also, Chen and colleagues revealed that inhibition of HDAC suppresses myocardial remodeling [108].

New data uncovers the relationship between histone acetylation and vascular inflammation. Interleukin (IL)-1 β , IL-6 and tumor necrosis factor (TNF)- α are some of the proinflammatory cytokine that are decreased by HDAC inhibitors as shown by multiple in vivo studies [109, 110]. The function of histone modifications in VSMC phenotypic modulation and proliferation in response to these cytokines have been studied in multiple other studies [111].

In a histone epigenetic modification study done by Wierda et al. it was shown that hypomethylation was linked with pathophysiological status [112]. Immunohistological examination of the samples taken from perirenal aortic tissues from clinical organ transplant cases, where morphological evidence of progression of atherosclerosis was more evident, showed that fewer nuclei exhibited trimethylation status at lysine 27 of histone H3 (H3K27Me3) [112].

Histone modifications play their role in HF. Kaneda and colleagues focused on genome-wide histone methylation of heart tissues [113]. They reported that tri-methylated histone H3H4 and H3K9 were altered in HF.

In animal models, especially H3K9 and H3K4 type, histone methylations are related with CH. Hypermethylation associated to hydroxymethylation of epidermal growth factor receptor

(EGFR) gene was found to be linked with aortic valve calcification and subsequent ventricular hypertrophy in another study that utilized experimental mice model [114].

Many histone modifications are discovered to have a role in the pathophysiology of hypertension. Following pressure overload in the heart, in more than half of the differentially expressed genes methylation or acetylation alterations are seen [115]. Hypermethylation of H3 by disruptor of DOT1-Like Histone H3K79 Methyltransferase (DOT1L), a methyltransferase, enhances the expression of hypertension-promoting telomeric genes [116]. DOT1L also has a role as an antihypertensive. The antihypertensive role is played by the suppression of amiloride-sensitive renal epithelial sodium channel (ENaC) expression [117]. DOT1L is suppressed by aldosterone. As a result of this ENaC expression increases leading to an increase in blood pressure. In several animal models it was found that lack of LSD1 caused hypermethylation of H3 which in turn lead to increased inclination to hypertension [118]. When mice fed a high-salt diet, LSD1 deficiency was closely linked with arterial hypertension. LSD1 induces demethylation of H3K4 or H3K9 changing gene transcription [118]. In the high-salt diet group, LSD1 deficiency lead to hypertension, enhanced vascular contraction, and reduced relaxation via nitric oxide–cyclic guanosine monophosphate pathway [118]. This finding portrays that LSD1-mediated histone demethylation has a role in modifying NOS/guanylate cyclase gene expression and blood pressure.

Status of histone acetylation is involved in the control of blood pressure. Epigenetic modulation has a function in the beta2 adrenergic receptor–GR–WNK4 pathway. Data from murine studies suggest that this pathway is linked with the activation of HDAC8 [119]. HDAC8 reduces the binding of the acetylated histones 3 and 4 with the promoter of the WNK4 gene. This reduces the transcriptional activity of WNK4 through the recruitment of glucocorticoid receptors to a negative glucocorticoid response element [120]. WNK4 is a member of serine/threonine kinase family which down regulates expression of sodium chloride transporters and epithelial sodium channels [121].

Hyperglycemia with prolonged exposure induces specific chromatin alterations and transcriptional responses. These illustrate the role of epigenetic events controlling signaling pathways relevant to diabetic CVD [122]. Genome-wide analysis of human vascular cells demonstrates that histone modifications and DNA methylation take place at the same time under hyperglycemic conditions allowing for prediction of gene expression [123]. In the study done by Mutshtov and colleagues it was proved that insulin gene in human islet-derived precursor cells showed increased levels of histone modifications (H4 hyperacetylation and dimethylation of H3 lysine 4) which is indicative that a gene is active [124].

Alterations in epigenomics can happen in the inflammatory pathways with hyperglycemia. In the study performed by Villeneuve and colleagues, it was discovered that there was a rise in the expression of inflammatory genes in vascular cells cultured in high glucose. On the other hand, H3K9me3 expression was reduced at its promoter [125]. H3K9me3 is a histone H3 which prevents against the biochemical state of diabetic inflammation. El-Osta and colleagues portrayed that the slightest transient exposure to high levels of glucose induced sustained epigenomic alterations in the p65 subunit promoter of the NF- κ B [126]. This change altered the epigenomics of cultured aortic endothelial cells. This could be in part the reason of the irreversibility of cardiovascular complications of long-standing uncontrolled diabetes which is usually the case in the routine clinical setting.

3.3. MiRNAs and lncRNAs

Some of the miRNA and lncRNA molecules are under study to define the association between epigenetic control and the development of CVDs. We focus on the role of several miRNAs and lncRNAs in many CVDs.

Several miRNAs are known to be associated with atherosclerosis. It has been reported that atherosclerosis patients had reduced levels of miR-126, miR-17, miR92a (in endothelial cells), miR-145 (in smooth muscle cells), and miR-155 (in inflammatory response-associated cells) in plasma and serum [127]. Conversely, miR-133 and miR-92-a (in cardiac muscle) were significantly increased in plasma of atherosclerosis patients [127]. In addition to these findings, it is compared many miRNA levels in atherosclerosis patients with stable and unstable angina. And this research group showed that unstable angina patients had significantly increased levels of miR-134, miR-370, and miR-198 vs stable angina patients [128]. The other research group compared the platelet miRNA levels between control and patients with atherosclerosis. They revealed significantly increased miR-340 and miR-624 levels in atherosclerosis patients [129]. It has been recently shown that miR146a-miR146b are implicated in the pathogenesis of atherosclerosis [130].

Recently, it has been shown that increased expression of an antisense lncRNA (ANRIL, antisense non-coding RNA in the *INK4* locus) is associated with the severity of atherosclerosis [130]. Also, another study found that SNP rs10757278 in the *INK4* locus is related with loss of ANRIL expression and increased risk of atherosclerosis [131]. Cunningham and colleagues showed an association between SNP rs1333045 and susceptibility of atherosclerosis [132].

Several studies have been performed to define the association of miRNAs with MI. In a human study, Bostjancic and colleagues found an increased levels of miR-208 in MI patients compared with healthy controls [46]. MiR-1 is one of the most studied miRNA. It has been shown that plasma miR-1 levels were significantly increased in MI patients vs healthy controls [133, 134]. Furthermore, there was a significantly positive correlation between miR-1 expression and infarct size in rat tissues [135].

D'Alessandra and colleagues analyzed the pattern of miRNAs plasma in patients with acute MI and healthy controls. Onset of MI symptoms showed significantly increased miR-1, miR-133a/b, and miR-499-5p in plasma [136]. In consistent with these results, another research group revealed a markedly increased of plasma miR-499 in MI patients [137].

Little is known about the association between lncRNAs and MI. Myocardial infarction-associated lncRNA transcript (MIAT) is a one of the risk factor for MI. Vausort and colleagues found higher MIAT levels in MI patients than healthy controls [138]. The same research group also showed an increased ANRIL expression in MI patients [138].

Many studies have determined a significant role of miRNAs in the pathogenesis of HF. The levels of several miRNAs showed alterations in animal models of HF and in patients with human cardiac [139]. While transgenic miR-195 mice had dilated cardiomyopathy, increased expression of miR-23a, miR-23b, miR-24, miR-195, and miR-214 induced hypertrophy in human cardiomyocytes [139]. In addition, overexpression of miR-133 decreased protein synthesis and inhibited hypertrophic growth of mouse cardiac myocytes [140].

Furthermore, Fukushima and colleagues has studied plasma miRNAs profiling in HF patients. They found that patients with HF had reduced miR-126 plasma concentration compared with healthy controls [141]. Also, the profiling of miRNAs has been analyzed by

Corsten and colleagues, only miR-499 levels were significantly increased in HF patients compared to controls [142].

Recently, it has been defined markedly expression changes of lncRNAs in various mice and human models of HF. It was found that Pdk1 knockout mouse models had severe HF [143]. Also, one report revealed elevated levels of long intergenic noncoding RNA predicting cardiac remodeling (LIPCAR) in patients with HF [55]. In addition, many research studies suggested that it can be an association of two lncRNAs, Fendrr and Bvht, in HF [52, 51].

MiRNAs were characterized as important regulatory elements in differentiation and maintenance of nervous system [144]. In rat studies, miR-290, miR-292-5p, and miR-497 were upregulated in blood and brain, on the other hand miR-210, miR-215, miR-324-3P, miR-422b, miR-451, and miR-154 were found to be increased only in brain [145].

Another study demonstrated an increased expression of miR-10a, miR-182, miR-200b, and miR-298 in animal models with induced stroke [146]. In converse, the same research group found that miR-155, miR-362-3p, miR-223, and miR-210 were significantly reduced expression in ischemic brain [146]. MiRNA expression has been also investigated in a human study. It has been shown overexpression of five miRNAs in symptomatic, stroke-related, and atherosclerotic plaques [147].

It is considered that there is an association of ANRIL with stroke. ANRIL may increase the risk of stroke via regulation of caspase recruitment domain 8 (CARD8). ANRIL regulates the expression of its downstream gene CARD8 [148].

In conclusion, definition of the miRNA and lncRNA profiles could have an important impact on understanding the pathogenesis of CVDs. Identification of the biological functions of miRNAs and lncRNAs in the cardiovascular system will be essential for CVD prevention, diagnosis, and therapy. Also, these molecules may have the potential to be used as prognostic biomarkers.

4. CARDIOVASCULAR RISK FACTORS AND EPIGENETICS

The majority of CVDs is caused by risk factors that can be controlled, treated or modified, such as diet and nutrition, tobacco use, obesity, lack of physical activity, high blood pressure, cholesterol and diabetes. However, there are also some major CVD risk factors (e.g., age, gender, family history) that cannot be controlled. When those risk factors are altered there could be a modification in epigenetics leading to a decline in the incidence of CVD.

Current evidence shows that epigenetic mechanisms can regulate nutritional effects. They can be the reason for the development of common complex or chronic diseases [149]. Nutritional factors are of major importance in CVD disease modification and prevention. Hypertension, atherosclerosis, many metabolic disorders, obesity, and weight-loss outcomes are affected by alterations in the epigenetic patterns. Histone modifications, DNA methylation, non-coding ribonucleic acid (ncRNA) expression, and chromatin remodeling resulting in a dynamic regulation of gene expression leading to control of the cellular phenotype are among the many different factors that show complex interaction with food components [150]. Changes in epigenetics seem to be more significantly linked with gene-nutrition and gene-environment interactions. The changes in lipid metabolism, inflammation, and other metabolic imbalances which resulted in CVD disease and obesity.

Obesity is often seen together with the metabolic syndrome. They both have major function in the establishment of CVD disease. Epigenetics have effects on both of them concerning the occurrence or prevention of CVD disease. In a study performed in pregnant mice, a high fat diet during pregnancy was found to be linked with epigenetic changes in the expression of adipocytokine genes [151]. Higher blood pressure, worse glucose tolerance, higher triglyceride levels, higher leptin levels and significantly lower adiponectin expression in white adipose tissue were seen in the offspring of the previously mentioned mice. The aforementioned relevance was with lower acetylation and higher methylation levels of histone H3 at lysine 9 of the adipose-tissue promoter of adiponectin [151].

Gene expression can be changed by epigenetic mechanisms imposed by nutritional factors. Recent data suggests that dietary lipids and their derivatives are dynamic modulators of pro- or anti-inflammatory gene expression pathways via nuclear receptors. These nuclear receptors take part in the regulation of various biological functions. Lipid metabolism, inflammatory mediator production, and vascular homeostasis are some examples of these biological functions. Selected dietary components may induce an atherosclerotic cellular phenotype. This could be partially done by producing epigenetic marks that shift differential gene activation and repression [74]. The PPAR receptor family is induced by small lipophilic ligands. These ligands are transcription factors involved in the regulation of genes in the cellular processes such as inflammation, beta-oxidation, and glucose homeostasis. They may be thought to be important targets for pharmacologic intervention in chronic diseases [74]. PPARs are major parts of a transcriptional network that imposes epigenetic marks. PPARs are also considered to be appropriate transcriptional response elements to specific ligands and molecules as factors that cross-talk with PPAR signaling. Fatty acids, intracellular inflammatory mediators, polyphenols, and multiple dietary factors that control the transcription of genes enrolled in the metabolic processes linked with inflammation and fatty acid metabolism are members of these factors.

Nutritional epigenetics may clear the way to understand the pathways of how nutrition can influence health with its effects on the genome. It is known that with administration of folate, homocysteine levels will be lowered. Nonetheless no reduction in the risk of CVD disease risk could be established. Some authors claim that a combination of vitamin B, folic acid, vitamin B6 and vitamin 12 could have harmful effects [152]. Especially during crucial times like fetal and postnatal development, nutritional effects on epigenetic programming could cause major effects to CVD disease [153]. An important therapeutic concept is that nutrients and medication could reciprocate epigenetic patterns. During special development milestones, epigenetic programming, or reprogramming could be effective mechanisms which may result in improvement in the health status of the patient. This could be achieved by dietary regulation [154].

Among the epigenetic modifications, DNA methylation is the most comprehensively investigated. Nutrition and other environmental stimuli can affect methylation. S-adenosylmethionine (SAM) is the prime methyl donor for DNA methylation. With the availability of specific dietary micronutrients such as folate, choline, betaine (trimethylglycine), and various B vitamins, this metabolism can be catalyzed by several enzymes. The relationship between nutrition and DNA methylation is endorsed by findings from animal studies. Epidemiological evidence from human studies is quite limited.

Folate affects the formation of SAM. SAM is a methyl donor for methylation of cytosines in DNA [155]. This kind of methylation is closely linked with gene silencing. Another part in

gene silencing is played by the covalent attachment of biotin to histones. This reaction is also involved in any cellular response to DNA damage. Tryptophan and niacin are converted to nicotinamide adenine dinucleotide (NAD). NAD is a substrate for polyADP-ribosylation (PAR) of histones and other DNA-binding proteins. PAR is involved in DNA repair and apoptosis. Epigenetic changes seem to be a mechanism to mediate the effects of environmental exposures in the early periods of life and gene-environment interactions on the development of adult type disease [156]. Maternal dietary consumption of fat, folate, protein, and the amount of total energy intake may change the epigenetic regulation of specific genes in the offspring. This could end up in functional changes in the tissues [157].

Raising the amount of vitamin D consumption does not seem to have major disadvantages. Current evidence shows that vitamin D could have important effects on epigenetics during the intrauterine period leading to reduction in chronic disease later in life. Vitamin D also has many potential benefits to wellbeing including CHD [158].

Flavonoids are potent inhibitors of DNA methyltransferases *in vitro*. They have the power to reverse hypermethylation and reactivate DNA repair genes. Folate is found in high concentration in green leafy vegetables. It mediates DNA methylation by its generation of SAM. People who have diets with low levels of folate or who have low levels of folate in their blood, have a significantly increased risk of developing CHD disease [159]. Nonetheless there is controversy about the possible lack of value, or even harm, from direct supplementation with folate combined with vitamins B6 and B12 [152, 160]. Folate is important in preventing certain birth defects. Supplementation of folate with myo-inositol in the early periods of pregnancy to a potential mother could alter the epigenetics aiding to prevent environmentally induced birth defects which are closely linked with congenital heart disease [161].

Epigenetic modifications that are affected by the intrauterine environment are investigated with the current animal studies. In a study, offspring of low protein fed pregnant mothers was found to have hypomethylated angiotensin type 1b receptor (AT1bR) gene promoters in addition to increased adrenal expression of AT1bR [162]. From this finding it can be inferred that specific hypomethylation could have a role in the regulation of elevated blood pressure. Overexpression of the hepatic glucocorticoid (GR) and PPAR α receptors in offspring were observed in the other studies that used a model of low protein diet receiving pregnant rats [163, 164]. A lot of different metabolic pathways could be affected by the key transcription factors GR and PPAR α receptors. They have been considered to have a role in the pathogenesis of several disease conditions like obesity, diabetes, and atherosclerosis [165, 166].

Risk of adult CHD disease, osteoporosis, and type 2 DM are increased in people who were small for gestational age at birth [167]. They also have poor growth rates in infancy. The increased risk more apparent when the early growth is restricted and it is followed by increased childhood weight gain [167]. The previously mentioned aspects are linked with epigenetic processes altering the phenotype of the offspring, reflecting the developmental responses of the fetus and/or infant based on environment.

Barker and colleagues hypothesized that environmental factors in crucial periods of early life (during fetal development, for instance) can influence risks for cardiovascular and metabolic diseases later in life [168]. This concept is supported by a number of studies that have associated low birth weight in human populations with increased risk of cardiovascular disease. For example, individuals prenatally exposed to famine during the Dutch Hunger Winter (1944–45) experienced higher prevalence of obesity and coronary heart disease as adults, when compared to adults born before or conceived after that period [169]. In Barker's

seminal work, it was found that low birth-weight babies who survived infancy had an increased risk of coronary heart disease later in life, and that increasing birth weight was associated with a graded decrease in risk [170].

In experimental animal models during the early stages of post-natal life exposure to different behavioral patterns was discovered to affect epigenetic modifications [171]. It could be hypothesized from this finding that the lifelong alterations are at the least partially affected by the epigenetic changes in the gene expression during the early periods of life [172]. When applied to the human population, social and environmental stresses during development that change epigenetic processes could be the reason of adult race-based US health differences seen in hypertension, diabetes, stroke, and coronary heart disease [173].

Wheatley et al. studied the hypothesis of modulation of adipose gene expression with the reversal of obese status by restriction of calories and exercise in 48 female mice [174]. 209 genes were found to be responsive to both calorie restriction and exercise in the gene expression microarray analysis of visceral white adipose tissue. 496 genes were affected by calorie restriction alone while just 20 genes were altered by exercise. 17 genes are related to carbohydrate metabolism and glucose transport from the genes which were responsive to calorie restriction. Glucose transporter 4 was among the genes affected by calorie restriction. Calorie restriction significantly increased acetylation of histone 4. This finding was in parallel with differential changes in adipose transcription with calorie restriction [174].

Long-term repetitive strenuous exercise is considered by many as having a positive effect on health, a reduction in aging, and leading to decreased incidence of disease. Modification of epigenetics can be achieved by physical exercise. Prevention from CVD disease by exercise can be explained by the exercise's power to modify epigenetics [175]. Possible benefits on health are caused by adaptations in skeletal muscle adaptations. These adaptations happen partly due to alterations in gene expression in the skeletal muscle. Chromatin remodeling by epigenetic histone modification seems to be the key regulatory mechanism modifying gene expression. Class IIa HDACs enzymes are closely linked with the adaptation in exercise. These enzymes inhibit histone acetylation. In their study McGee et al. portrayed that global histone 3 acetylation was increased at lysine 36, a site that was related with transcriptional elongation. They used cycling for 60 minutes as the method of exercise in their study [176]. They also discovered that that HDAC₄ and HDAC₅ were exported from the nucleus during exercise. The suppressive role of HDAC₄ and HDAC₅ was expelled.

In spite of the data that supports the positive influence of physical exercise on epigenetic mechanisms and an improvement in health, there is not a clear connection between exercise and epigenetics. Excessive and persistent physical exercise could in fact may have a negative influence on health. Gene expression may be affected by modified epigenetics that were changed by the body's physical adaptation to environmental conditions [175].

The functional genome in cardiac and peripheral vascular beds can be altered by exercise or physical activity. Reactive oxygen species (ROS) are formed during exercise but it also enhances anti-oxidative capacity. A deviation in the balance of cellular oxidative stress ensues. ROS have important regulatory function in muscle contraction, antioxidant protection, and repair of oxidative damage. DNA methylation and histone modification could have role in exercise-related ROS production thus leading to heritable conditions which are overseen by epigenetics. Many signal cascades controlling physiological and pathophysiological cardiac and peripheral vascular adaptations are modified [177]. The result is the modification in the

molecular composition of the extracellular matrix. As a result several signaling cascades are changed.

There are many epigenetic signatures of environmental exposure pertaining to CVD as discovered from different studies (population cohorts for global and gene specific alterations). Oxidative stress, aging, and atherosclerosis are CVD disease outcome measures [178]. Exposure to traffic-related air pollution is closely linked with an increase in hospitalization rates and mortality from the perspective of CVD disease outcome measures. Metals affect health in many perspectives [179]. Under the light of new data, epigenetics could be a crucial pathway where metals produce their effects on health. Metallomics studies concerning environmental toxins are using epigenetics as a brand new major area to make their investigations. DNA methylation, modification of histone proteins, and RNA interference are three factors which are involved in CVD disease and other inherited conditions that are being studied by epigenetics [179]. Tobacco is a well-known CVD disease risk factor. It exerts its effects through epigenetics [180]. DNA methylation, histone modification, and microRNA alterations are environmental epigenetic mechanisms produced by tobacco. Tobacco consumption by the mother has been linked with changes in methylation of placental cytochrome P450A1 (CYP1A1) gene restriction [181]. This change leads to fetal growth restriction. Recently there has been a new finding between the prognosis for patients with stable CVD and smoking-related methylation patterns in the Coagulation factor II (thrombin) receptor-like 3 (F2RL3) gene [182].

Recent evidences have also shown the role of modulating or modifying miRNAs in established several risk factors for CVDs. Dyslipidemia is a common increased risk factor for developing of CVDs. Decreased levels of atherogenic lipoproteins are related with reduced CVD risk [183].

Numerous studies are performed for analyzing the biological functions of miRNAs on lipid metabolism. MiR122 that is an emerging miR in animal model hepatocytes targets many genes required in lipid metabolism [184]. Two studies demonstrated that miR122 inhibition reduced the levels of total serum cholesterol by increasing the fatty acid oxidation and hepatic lipid synthesis [185, 186]. Furthermore, it has been shown that miR122 reduced triglycerides and decreased the synthesis of the LDL receptor [187]. Rayner and colleagues found that another miR, miR33, regulates the cholesterol homeostasis by modulating either HDL biogenesis in the liver or cellular cholesterol efflux [188]. In African green monkeys, suppression of miR33 reduced serum triglycerides and increased HDL [188]. A genome-wide association study defined that presence of rare rs13702 C allele within the miR410 leads to lower triglycerides and higher HDL [189].

Diabetes and obesity which are another emerging risk factors for CVDs modified by miRNAs. It has been found an increased miR103 and miR107 in obese and diabetic mice liver [190, 191]. The treatment with Anti-miR103/miR107 extends glucose homeostasis and suppresses the size of adipocytes and total amount of visceral fat in these mice [191].

One research group demonstrated the association between antisense inhibition of miR208 and markedly repression of weight gain in mice which were bleeding with high-fat diet [192].

Tobacco smoke has lots of different chemical compounds that are CVD disease-causing compounds. Nicotine is one of the compound of cigarette smoke. It has been demonstrated that miR133 and miR590 progressed the remodeling of atrial in the present of nicotine treatment by modifying the TGF- β 1 and TGF- β receptor II [193].

In addition, miR-197, miR-23a, and miR-509-5p were shown being correlated with body mass index, the risk factor for CVDs, in human study [194].

5. EPIGENETICS-BASED THERAPEUTIC STRATEGIES IN CVDs

Pharmacological therapy centered on the epigenome is a novel therapeutic concept. Pharmacoepigenomics uses epigenetic modification by medications as its basis of approach [195]. Pharmacoepigenomics might be able to more personally tailored treatment options. These kind of pharmaceutical agents are called epi-drugs. DNA methyltransferases, HDACs, HATs, histone methyltransferases, and histone demethylases are type of epi-drugs. Pharmacoepigenomics' major use was thought to be pharmacotherapy related to cancer [196]. However, currently new evidence shows that personally tailored oral antidiabetic treatment modalities could be more beneficial than standard treatments. In the future, pharmacoepigenomics could alter different individual response to oral antidiabetic treatment. Dysfunction of smooth muscle cells in diabetes could be the reason of increased incidence of macrovascular complications. Epigenetic modification of smooth muscle cell function may be the advantage of this treatment option as CHD risk in diabetic patients. Because, this treatment strategy does not require the tight glucose control in these patients [197].

CVD has a complex pathophysiological process and pathogenic pathways of CVDs remain unclear. Genetic approaches including epigenetics and personally tailored medical treatment established a novel approach in the treatment of CVDs. Chromatin structure has been chosen as a treatment target by the majority of epigenetic challenges. Most of the isoform-selective HDAC-inhibitors could have an importance in the treatment of cancer, Huntington's disease, sickle cell disease, or CVDs [198, 199].

Epigenetic tags can be affected by modifiers of methylation. DNA methylation status alter the function of target atheroprotective genes, such as ER α and ER β genes in atherosclerosis [76]. Human coronary atherosclerotic tissues contain hypermethylated ER α and ER β genes. This hypermethylation status could be demethylated by using epigenetic inhibitors such as DAC. DAC could upregulate the expression of silenced ER α and ER β genes [75]. Therefore, therapeutic options could be used in the treatment of atherosclerosis by changing the epigenetic modifications.

DAC is a DNA methylation inhibitor. It has worked as a successful treatment option in hematological diseases and malignancies. As portrayed by cardiovascular research, DNA methylation inhibitor promotes the differentiation of cardiomyocytes from mesenchymal stem cells at low frequencies [200, 201]. DAC was given prime importance after this finding. Unluckily, myeloid suppression and induction of cardiomyocyte cell death are the toxic affects imposed by DAC on behalf of its potential benefits. Zebularine is another cytidine analog. It produces less toxicity than DAC [202]. This property of zebularine promotes its use in studies on cardiomyocyte differentiation from ESCs. In the recent study, Horrillo et al. demonstrated that zebularine promotes cardiomyocyte differentiation from ESCs with increased expression of cardiac-specific genes [203]. However, using of this knowledge in the development of stem-cell-based therapy for degenerative cardiac diseases is still elusive.

In addition, collagen type XV A1 (COL15A1) is hypomethylated in the process of aortic SMCs proliferation. DAC treatment induces aortic proliferation by inhibiting DNMTs and demethylating COL15A1 [204].

There are many documents which revealed that epigenetic mechanisms could be limited by HDAC and HAT inhibitors in CVDs. Trichostatin A and valproic acid, HDAC inhibitors, upregulated CYP7A1 expression and reduced plasma cholesterol levels in LDL-knockout mice [205]. Furthermore, it has been defined that Trichostatin A increased the ER α and ER β levels in SMCs and ECs [76]. The other study demonstrated that reduction of HAT activity by inhibition of the co-activators nuclear cap-binding protein subunit 1 and p300 HAT reduced cardiac hypertrophy [105]. Curcumin is one of the inhibitor of HAT activity [206]. It has been found that Curcumin prevents ventricular hypertrophy and preserved systolic pressure by inhibiting p300 HAT activity in an animal model of HF [104].

MiRNA biological network in the cardiovascular system plays an important role for CVD prevention, diagnosis, and therapy [207]. MiRNAs seems have critical importance in the understanding of the diagnosis and therapy of CVDs, especially diagnostic rather than the therapeutic approach [208].

A number of miRNAs are supposed as biomarkers for the diagnosis of MI. Many reports showed that miR208a-b was upregulated in blood and heart tissue [209, 210]. Furthermore, miR1 and miR499 were demonstrated as biomarkers for MI [134, 210].

Fukushima and colleagues profiled plasma miRNA levels in patients with HF. They showed decreased plasma concentration of miR126 in patients vs healthy subjects, it is also negatively correlated with Brain Natriuretic Peptide (BNP) plasma levels. BNP is a biomarker of HF and so increased miR126 levels can be a biomarker for HF [141]. In addition, miR508-5p could be defined as novel diagnostic and therapeutic miRNAs for HF [211].

MiRNAs have been recently studying as likely therapeutic targets. It is supposed that miRNAs could be novel clinical therapeutic approaches for CVDs. Preclinical studies are being done on miRNAs and CVD treatment. Despite markedly advances, in vivo studies just remain scientific and therapeutic challenge [130].

Several animal models have been generated for therapeutic approaches in CVDs. It has been shown that treatment of antagomir-92a protects against endothelial activation and dysfunction and CAD in mice [212]. MiR155 induces the cardiac hypertrophy. So, miR155 inhibition might provide clinical therapeutic strategies for cardiac hypertrophy and HF [213]. Also, it has been shown negative correlation between silencing of miR329, miR487b, miR494, and miR495 and in arteriogenesis [214]. Recently, it is found that miR146a/b could be modulated by combined therapy [215]. Furthermore, pharmacological compounds can be used for modulation of miRNAs as an effective therapy [130]. Montgomery and colleagues silenced the miR208a by Anti-miR and improved heart function and surviving in HF progression in rats [216]. In addition, it has been shown that downregulation of miR1 could be therapeutic target against acute MI [217].

More human and animal studies are still few, so it is needed further investigations of miRNAs to improve therapeutic advance in CVDs.

At present lncRNAs have been studying to understand the function of these molecules in heart development and CVDs. So, future studies should reveal diagnostic and specific therapeutic benefits in CVDs.

Since 2012 when it was reported at the American Heart Association Scientific Sessions regenerative medicine is now a prominent actor of CVD research [218]. Genome wide

association studies and epigenetics are major actors of both basic and clinical CVD research. There are ongoing trials and basic science research projects to establish the part of epigenetics in regenerative medicine. There is massive potential of induced pluripotent stem cells (IPSCs) in altering the way we approach CVD. Pluripotent stem cells could have a role in regenerative medicine and individualized CVD therapies [218].

In order to alter the programming of somatic cells in the production of IPSCs, Sun et al. proposed the use of methods using microRNA. Sustaining pluripotency, specific cell lineage and modification of chromatin by epigenetics requires microRNA [219].

The other advanced application of regenerative medicine is utilization of a distinct biologic framework as the foundation of an epigenetically-modified, surgically insertable graft of different tissues. Jungebluth et al. portrayed this in their case of human tracheobronchial transplantation. The utilization of this method for vascular or other CV graft may effortlessly be visualized [220].

Integration and solution of these kind of problems fits perfectly for the application of the principles of epigenetics in regenerative medicine.

CONCLUSION

CVDs have a complex pathophysiological basis. Multiple genetic and environmental risk factors are associated with CVDs. Modification of epigenetic marks are one of the risk factors of CVDs. Many epigenetic alterations may play key roles in initiation and progression of CVDs. Definition of the epigenetic mechanisms required to contribute to CVDs will identify strong significant contributions for novel and targeted therapeutic approaches in CVDs. Success of these clinical approaches could have an important impact on a large fraction of the human CVDs. Future studies will also clarify the detailed epigenetic modifications linked to CVDs and use this knowledge to prevent of CVDs.

REFERENCES

- [1] Bird AP. CpG-rich islands and the function of DNA methylation. *Nature* 1986;321:209-213.
- [2] Portela A, Esteller M. Epigenetic modifications and human disease. *Nat Biotechnol* 2010;28:1057-1068.
- [3] Tahiliani M, Koh KP, Shen Y, Pastor WA, Bandukwala H, Brudno Y, Agarwal S, Iyer LM, Liu DR, Aravind L, Rao A. Conversion of 5-Methylcytosine to 5-Hydroxymethylcytosine in Mammalian DNA by MLL Partner TET1. *Science* 2009;324:930-935.
- [4] Lennartsson A, Ekwall K. Histone modification patterns and epigenetic codes. *Biochim Biophys Acta* 2009;1790:863-868.
- [5] Chen HP, Zhao YT, Zhao TC. Histone Deacetylases and Mechanisms of Regulation of Gene Expression. *Crit Rev Oncog* 2015;20:35-47.
- [6] Bannister AJ, Kouzarides T. Regulation of chromatin by histone modifications. *Cell Research* 2011;21:381-395.

- [7] Lewis BP, Burge CB, Bartel DP. Conserved seed pairing, often flanked by adenosines, indicates that thousands of human genes are microRNA targets. *Cell* 2005;120:15-20.
- [8] Jinek M, Doudna JA. A three-dimensional view of the molecular machinery of RNA interference. *Nature* 2009;457:405-412.
- [9] Liang M, Liu Y, Mladinov D, Cowley AW, Trivedi H, Fang Y, Xu X, Ding X, Tian Z. MicroRNA: a new frontier in kidney and blood pressure research. *Am J Physiol Renal Physiol* 2009;297:553-558.
- [10] Seema R. Lalani, John W. Belmont. Genetic basis of congenital cardiovascular malformations. *Eur J Med Genet* 2015;57:402-413.
- [11] Paolo Piatti, Anette Zeilner, Alexandra Lusser. ATP-Dependent Chromatin Remodeling Factors and Their Roles in Affecting Nucleosome Fiber Composition. *Int J Mol Sci* 2011;12:6544-6565.
- [12] Wang W, Cote J, Xue Y, Zhou S, Khavari PA, Biggar SR. Purification and biochemical heterogeneity of the mammalian SWI-SNF complex. *The EMBO journal* 1996;15:5370-5382.
- [13] Hang CT, Yang J, Han P, Cheng HL, Shang C, Ashley E, Zhou B. Chromatin regulation by Brg1 underlies heart muscle development and disease. *Nature* 2010;466:62-67.
- [14] Workman JL. Nucleosome displacement in transcription. *Genes Dev* 2006;20:2009-2017.
- [15] Wang Z, Zhai W, Richardson JA, Olson EN, Meneses JJ, Firpo MT, Kang C, Skarnes WC, Tjian R. Polybromo protein BAF180 functions in mammalian cardiac chamber maturation. *Genes Dev* 2004;18:3106-3116.
- [16] Huang X, Gao X, Diaz-Trelles R, Ruiz-Lozano P, Wang Z. Coronary development is regulated by ATP-dependent SWI/SNF chromatin remodeling component BAF180. *Dev Biol* 2008;319:258-266.
- [17] Shirai M, Osugi T, Koga H, Kaji Y, Takimoto E, Komuro I, Hara J, Miwa T Yamauchi-Takahara K, Takihara Y. The Polycomb-group gene *Rae28* sustains *Nkx2.5/Csx* expression and is essential for cardiac morphogenesis. *J Clin Invest* 2002;110:177-184.
- [18] Barth JL, Clark CD, Fresco VM, Knoll EP, Lee B, Argraves WS, Lee KH. *Jarid2* is among a set of genes differentially regulated by *Nkx2.5* during outflow tract morphogenesis. *Dev Dyn* 2010;239:2024-2033.
- [19] Bruneau BG, Bao ZZ, Tanaka M, Schott JJ, Izumo S, Cepko CL, Seidman JG, Seidman CE. Cardiac expression of the ventricle-specific homeobox gene *Irx4* is modulated by *Nkx2-5* and *dHand*. *Dev Biol* 2000;217:266-277.
- [20] Hoffman JI, Kaplan S, Liberthson RR: Prevalence of congenital heart disease. *Am Heart J* 2004;147:425-439.
- [21] Aramaki M, Udaka T, Kosaki R, Makita Y, Okamoto N, Yoshihashi H, Oki H. Phenotypic spectrum of CHARGE syndrome with *CHD7* mutations. *J Pediatr* 2006;148:410-414.
- [22] Schnetz MP, Bartels CF, Shastri K, Balasubramanian D, Zentner GE, Balaji R, Zhang X. Genomic distribution of *CHD7* on chromatin tracks H3K4 methylation patterns. *Genome Res* 2009;19:590-601.
- [23] Ramsbottom SA, Sharma V, Rhee HJ, Eley L, Phillips HM, Rigby HF, Dean C, Chaudhry B, Henderson DJ. *Vangl2*-regulated polarisation of second heart field-derived cells is required for outflow tract lengthening during cardiac development. *PLoS Genet* 2014;10:e1004871.

- [24] Yuan Y, Gao Y, Wang H, Ma X, Ma D, Huang G. Promoter methylation and expression of the *VANGL2* gene in the myocardium of pediatric patients with Tetralogy of Fallot. *Birth Defects Res A Clin Mol Teratol* 2014;100:973-984.
- [25] Abbey D, Seshagiri PB. Aza-induced cardiomyocyte differentiation of P19 EC-cells by epigenetic co-regulation and ERK signaling. *Gene* 2013;526:364-373.
- [26] Chamberlain AA., Lin M, Lister RL, Maslov AA, Wang Y, Suzuki M, Wu B, Grealley JM, Zheng D, Zhou B. DNA methylation is developmentally regulated for genes essential for cardiogenesis. *J Am Heart Assoc* 2014;3:e000976.
- [27] Hirasawa R, Chiba H, Kaneda M, Tajima S, Li E, Jaenisch R, Sasaki H. Maternal and zygotic Dnmt1 are necessary and sufficient for the maintenance of DNA methylation imprints during preimplantation development. *Genes Dev* 2008;22:1607-1616.
- [28] Kurihara Y, Kawamura Y, Uchijima Y, Amamo T, Kobayashi H, Asano T, Kurihara H. Maintenance of genomic methylation patterns during preimplantation development requires the somatic form of DNA methyltransferase 1. *Dev Biol* 2008;313:335-346.
- [29] Okano M, Bell DW, Haber DA, Li E. DNA methyltransferases Dnmt3a and Dnmt3b are essential for de novo methylation and mammalian development. *Cell* 1999;99:247-257.
- [30] Kou CY, Lau SL, Au KW, Leung PY, Chim SS, Fung KP, Waye MM, Tsui SK. Epigenetic regulation of neonatal cardiomyocytes differentiation. *Biochem Biophys Res Commun* 2010;400:278-83.
- [31] Gottlieb PD, Pierce SA, Sims RJ, Yamagishi H, Weihe EK, Harriss JV, Maika SD, Kuziel WA, King HL, Olson EN, Nakagawa O, Srivastava D. Bop encodes a muscle-restricted protein containing MYND and SET domains and is essential for cardiac differentiation and morphogenesis. *Nat Genet* 2002;31:25-32.
- [32] Srivastava D, Cserjesi P, Olson EN. A subclass of bHLH proteins required for cardiac morphogenesis. *Science* 1995;270:1995-1999.
- [33] Lee Y, Song AJ, Baker R, Micales B, Conway SJ, Lyons GE. Jumonji, a nuclear protein that is necessary for normal heart development. *Circ Res* 2000;86:932-938.
- [34] Jung J, Kim TG, Lyons GE, Kim HR, Lee Y. Jumonji regulates cardiomyocyte proliferation via interaction with retinoblastoma protein. *J Biol Chem* 2005;280:30916-30923.
- [35] Wang J, Scully K, Zhu X, Cai L, Zhang J, Prefontaine GG, Krones A, Ohgi KA, Zhu P, Garcia-Bassets I, Liu F, Taylor H, Lozach J, Jayes FL, Korach KS, Glass CK, Fu XD, Rosenfeld MG, Wang J. Opposing LSD1 complexes function in developmental gene activation and repression programmes. *Nature* 2007;446:882-887.
- [36] Mathiyalagan P, Chang L, Du XJ, El-Osta A. Cardiac ventricular chambers are epigenetically distinguishable. *Cell Cycle* 2010;9:612-617.
- [37] Yao TP, Oh SP, Fuchs M, Zhou ND, Ch'ng LE, Newsome D, Bronson RT, Li E, Livingston DM, Eckner R. Gene dosage-dependent embryonic development and proliferation defects in mice lacking the transcriptional integrator p300. *Cell* 1998;93:361-372.
- [38] Shikama N, Lutz W, Kretzschmar R, Sauter N, Roth JF, Marino S, Wittwer J, Scheidweiler A, Eckner R. Essential function of p300 acetyltransferase activity in heart, lung and small intestine formation. *EMBO J* 2003;22:5175-5185.
- [39] Mathiyalagan P, Keating ST, Du XJ, El-Osta A. Chromatin modifications remodel cardiac gene expression. *Cardiovasc Res* 2014;103:7-16.

- [40] Montgomery RL, Davis CA, Potthoff MJ, Haberland M, Fielitz J, Qi X, Hill JA, Richardson JA, Olson EN. Histone deacetylases 1 and 2 redundantly regulate cardiac morphogenesis, growth, and contractility. *Genes Dev* 2007;21:1790-1802.
- [41] Montgomery RL, Potthoff MJ, Haberland M, Qi X, Matsuzaki S, Humphries KM, Richardson JA, Bassel-Duby R, Olson EN. Maintenance of cardiac energy metabolism by histone deacetylase 3 in mice. *J Clin Invest* 2008;118:3588-3597.
- [42] Zhao Y, Samal E, Srivastava D. Serum response factor regulates a muscle-specific microRNA that targets Hand2 during cardiogenesis. *Nature* 2005;436:214-220.
- [43] Philippen LE, Dirkx E, da Costa-Martins PA, De Windt LJ. Non-coding RNA in control of gene regulatory programs in cardiac development and disease. *J Mol Cell Cardiol* 2015 (in print).
- [44] Cordes KR, Srivastava D, Ivey KN. MicroRNAs in cardiac development. *Pediatr Cardiol* 2010;31:349-356.
- [45] Yang J, Xu WW, Hu SJ. Heart failure: advanced development in genetics and epigenetics. *Biomed Res Int* 2015 (in print).
- [46] Bostjanjic E, Zidar N, Stajer D, Glavač D. MicroRNAs miR-1, miR-133a, miR-133b and miR-208 are dysregulated in human myocardial infarction. *Cardiology* 2010;115: 163–169.
- [47] Chen JF, Mandel EM, Thomson JM, Wu Q, Callis TE, Hammond SM, Conlon FL, Wang DZ. The role of microRNA-1 and microRNA-133 in skeletal muscle proliferation and differentiation. *Nat Genet* 2006;38:228-233.
- [48] Grueter CE, van Rooij E, Johnson BA, DeLeon SM, Sutherland LB, Qi X, Gautron L, Elmquist JK, Bassel-Duby R, Olson EN. A cardiac MicroRNA governs systemic energy homeostasis by regulation of MED13. *Cell* 2012;149:671-683.
- [49] Hosoda T, Zheng H, Cabral-da-Silva M, Sanada F, Ide-Iwata N, Ogórek B, Ferreira-Martins J, Arranto C, D'Amario D, del Monte F, Urbanek K, D'Alessandro DA, Michler RE, Anversa P, Rota M, Kajstura J, Leri A. Human cardiac stem cell differentiation is regulated by a mircrine mechanism. *Circulation* 2011;123:1287-1296.
- [50] Ohtani K, Dimmeler S. Control of cardiovascular differentiation by microRNAs. *Basic Res Cardiol* 2011;106:5-11.
- [51] Klattenhoff CA, Scheuermann JC, Surface LE, Bradley RK, Fields PA, Steinhauser ML, Ding H, Butty VL, Torrey L, Haas S, Abo R, Tabebordbar M, Lee RT, Burge CB, Boyer LA. Braveheart, a long noncoding RNA required for cardiovascular lineage commitment. *Cell* 2013;152:570-583.
- [52] Grote P, Wittler L, Hendrix D, Koch F, Währisch S, Beisaw A, Macura K, Bläss G, Kellis M, Werber M, Herrmann BG. The tissue-specific lncRNA Fendrr is an essential regulator of heart and body wall development in the mouse. *Dev Cell* 2013;24:206-214.
- [53] Margueron R, Reinberg D. The Polycomb complex PRC2 and its mark in life. *Nature* 2011;469:343-349.
- [54] Simon JA, Kingston RE. Mechanisms of polycomb gene silencing: knowns and unknowns. *Nat Rev Mol Cell Biol* 2009;10:697-708.
- [55] Rizki G, Boyer LA. Lncing Epigenetic Control of Transcription to Cardiovascular Development and Disease. *Circ Res* 2015;117:192-206.
- [56] Baccarelli A, Wright R, Bollati V, Litonjua A, Zanobetti A, Tarantini L, Sparrow D, Vokonas P, Schwartz J. Ischemic heart disease and stroke in relation to blood DNA methylation. *Epidemiology* 2010;21:819-828.

- [57] Antoniadou C, Antonopoulos AS, Tousoulis D, Marinou K, Stefanadis C. Homocysteine and coronary atherosclerosis: from folate fortification to the recent clinical trials. *Eur Heart J* 2009;30:6-15.
- [58] Yi P, Melnyk S, Pogribna M, Pogribny IP, Hine RJ, James SJ. Increase in plasma homocysteine associated with parallel increases in plasma S-adenosylhomocysteine and lymphocyte DNA hypomethylation. *J Biol Chem* 2000;275:29318-29323.
- [59] Pogribny IP, Beland FA. DNA hypomethylation in the origin and pathogenesis of human diseases. *Cell Mol Life Sci* 2009;66:2249-2261.
- [60] Ingrosso D, Cimmino A, Perna AF, Masella L, De Santo NG, De Bonis ML, Vacca M, D'Esposito M, D'Urso M, Galletti P, Zappia V. Folate treatment and unbalanced methylation and changes of allelic expression induced by hyperhomocysteinemia in patients with uraemia. *Lancet* 2003;361:1693-1699.
- [61] Jiang X, Yang F, Tan H, Liao D, Bryan RM Jr, Randhawa JK, Rumbaut RE, Durante W, Schafer AI, Yang X, Wang H. Hyperhomocysteinemia impairs endothelial function and eNOS activity via PKC activation. *Arterioscler Thromb Vasc Biol* 2005;25:2515-2521.
- [62] De Smet C, Lurquin C, Lethé B, Martelange V, Boon T. DNA methylation is the primary silencing mechanism for a set of germ line- and tumor-specific genes with a CpG-rich promoter. *Mol Cell Biol* 1999;19:7327-7335.
- [63] Ron D. Hyperhomocysteinemia and function of the endoplasmic reticulum. *J Clin Invest* 2001;107:1221-1222.
- [64] Jacobson TA, Miller M, Schaefer EJ. Hypertriglyceridemia and cardiovascular risk reduction. *Clin Ther* 2007;29:763-777.
- [65] Mikael LG, Wang XL, Wu Q, Jiang H, Maclean KN, Rozen R. Hyperhomocysteinemia is associated with hypertriglyceridemia in mice with methylenetetrahydrofolate reductase deficiency. *Mol Genet Metab* 2009;98:187-194.
- [66] Devlin AM, Lentz SR. ApoA-I: a missing link between homocysteine and lipid metabolism? *Circ Res* 2006;98:431-433.
- [67] Mikael LG, Genest J Jr, Rozen R. Elevated homocysteine reduces apolipoprotein A-I expression in hyperhomocysteinemic mice and in males with coronary artery disease. *Circ Res* 2006;98:564-571.
- [68] Liao D, Tan H, Hui R, Li Z, Jiang X, Gaubatz J, Yang F, Durante W, Chan L, Schafer AI, Pownall HJ, Yang X, Wang H. Hyperhomocysteinemia decreases circulating high-density lipoprotein by inhibiting apolipoprotein A-I protein synthesis and enhancing HDL cholesterol clearance. *Circ Res* 2006;99:598-606.
- [69] Kim GH, Ryan JJ, Archer SL. The role of redox signaling in epigenetics and cardiovascular disease. *Antioxid Redox Signal* 2013;18:1920-1936.
- [70] Kim J, Kim JY, Song KS, Lee YH, Seo JS, Jelinek J, Goldschmidt-Clermont PJ, Issa JP. Epigenetic changes in estrogen receptor beta gene in atherosclerotic cardiovascular tissues and in vitro vascular senescence. *Biochim Biophys Acta* 2007;1772:72-80.
- [71] Lund G, Zaina S. Atherosclerosis: an epigenetic balancing act that goes wrong. *Curr Atheroscler Rep* 2011;13:208-214.
- [72] Castro R, Rivera I, Struys EA, Jansen EE, Ravasco P, Camilo ME, Blom HJ, Jakobs C, Tavares de Almeida I. Increased homocysteine and S-adenosyl homocysteine concentrations and DNA hypomethylation in vascular disease. *Clin Chem* 2003;49:1292-1296.

- [73] Zhu S, Goldschmidt-Clermont PJ, Dong C. Inactivation of monocarboxylate transporter MCT3 by DNA methylation in atherosclerosis. *Circulation* 2005;112:1353-1361.
- [74] Zaina S, Heyn H, Carmona FJ, Varol N, Sayols S, Condom E, Ramírez-Ruz J, Gomez A, Gonçalves I, Moran S, Esteller M. A DNA methylation map of human atherosclerosis. *Circ Cardiovasc Genet* 2014;7:692-700.
- [75] Laukkanen M, Mannermaa S, Hiltunen M, Aittomäki S, Airenne K, Jänne. Local hypomethylation in atherosclerosis found in rabbit ec-sod gene. *Arterioscler Thromb Vasc Biol* 1999;19:2171-2178.
- [76] Lund G, Andersson L, Lauria M, Lindholm M, Fraga M, Villar-Garea A. DNA methylation polymorphisms precede any histological sign of atherosclerosis in mice lacking apolipoprotein E. *J Biol Chem* 2004;279:29147-29154.
- [77] Hiltunen M, Turunen M, Häkkinen T, Rutanen J, Hedman M, Mäkinen K. DNA hypomethylation and methyltransferase expression in atherosclerotic lesions. *Vasc Med* 2002;7:5-11.
- [78] Friso S, Lotto V, Choi SW, Girelli D, Pinotti M, Guarini P, Udali S, Pattini P, Pizzolo F, Martinelli N, Corrocher R, Bernardi F, Olivieri O. Promoter methylation in coagulation F7 gene influences plasma FVII concentrations and relates to coronary artery disease. *J Med Genet* 2012;49:192-199.
- [79] Backs J, Olson EN. Control of cardiac growth by histone acetylation/deacetylation. *Circ Res* 2006;98:15-24.
- [80] Movassagh M, Choy MK, Goddard M, Bennett MR, Down TA, Foo RS. Differential DNA methylation correlates with differential expression of angiogenic factors in human heart failure. *PLoS One* 2010;5:e8564.
- [81] Nguyen AT, Zhang Y. The diverse functions of Dot1 and H3K79 methylation. *Genes Dev* 2011;25:1345-1358.
- [82] Smolarek I, Wyszko E, Barciszewska AM, Nowak S, Gawronska I, Jablecka A, Barciszewska MZ. Global DNA methylation changes in blood of patients with essential hypertension. *Med Sci Monit* 2010;16:149-155.
- [83] Valinluck V, Tsai HH, Rogstad DK, Burdzy A, Bird A, Sowers LC. Oxidative damage to methyl-CpG sequences inhibits the binding of the methyl-CpG binding domain (MBD) of methyl-CpG binding protein 2 (MeCP2). *Nucleic Acids Res* 2004;32: 4100-4108.
- [84] Kinney SM, Chin HG, Vaisvila R, Bitinaite J, Zheng Y, Estève PO, Feng S, Stroud H, Jacobsen SE, Pradhan S. Tissue-specific distribution and dynamic changes of 5-hydroxymethylcytosine in mammalian genomes. *J Biol Chem* 2011;286:24685-24693.
- [85] Wu H, Zhang Y. Tet1 and 5-hydroxymethylation: a genome-wide view in mouse embryonic stem cells. *Cell Cycle* 2011;10:2428-2436.
- [86] Baserga M, Kaur R, Hale MA, Bares A, Yu X, Callaway CW, McKnight RA, Lane RH. Fetal growth restriction alters transcription factor binding and epigenetic mechanisms of renal 11beta-hydroxysteroid dehydrogenase type 2 in a sex-specific manner. *Am J Physiol Regul Integr Comp Physiol* 2010;299:334-342.
- [87] Zhao Y, Gong X, Chen L, Li L, Liang Y, Chen S, Zhang Y. Site-specific methylation of placental HSD11B2 gene promoter is related to intrauterine growth restriction. *Eur J Hum Genet* 2013;22:734-740.
- [88] Friso S, Carvajal CA, Fardella CE, Olivieri O. Epigenetics and arterial hypertension: the challenge of emerging evidence. *Transl Res* 2015;165:154-165.

- [89] Riviere G, Lienhard D, Andrieu T, Vieau D, Frey BM, Frey FJ. Epigenetic regulation of somatic angiotensin-converting enzyme by DNA methylation and histone acetylation. *Epigenetics* 2011;6:478-489.
- [90] Lee HA, Baek I, Seok YM, Yang E, Cho HM, Lee DY, Hong SH, Kim IK. Promoter hypomethylation upregulates Na⁺-K⁺-2Cl⁻ cotransporter 1 in spontaneously hypertensive rats. *Biochem Biophys Res Commun* 2010;396:252-257.
- [91] Haas J, Frese KS, Park YJ, Keller A, Vogel B, Lindroth AM, Weichenhan D, Franke J. Alterations in cardiac DNA methylation in human dilated cardiomyopathy. *EMBO Mol Med* 2013;5:413-429.
- [92] Movassagh M, Choy MK, Knowles DA, Cordeddu L, Haider S, Down T, Siggins L, Vujic A, Simeoni I, Penkett C, Goddard M, Lio P, Bennett MR, Foo RS. Distinct epigenomic features in end-stage failing human hearts. *Circulation* 2011;124:2411-2422.
- [93] Sessa WC. eNOS at a glance. *J Cell Sci* 2004;117:2427-2429.
- [94] Yan MS, Matouk CC, Marsden PA. Epigenetics of the vascular endothelium. *J Appl Physiol* 2010;109:916-912.
- [95] Ho J, Man H, Marsden P. Nitric oxide signaling in hypoxia. *J Mol Med (Berl)* 2012;90:217-231.
- [96] Chen W, Bacanamwo M, Harrison DG. Activation of p300 histone acetyltransferase activity is an early endothelial response to laminar shear stress and is essential for stimulation of endothelial nitric-oxide synthase mRNA transcription. *J Biol Chem* 2008;283:16293-16298.
- [97] Granger A, Abdullah I, Huebner F, Stout A, Wang T, Huebner T, Epstein JA, Gruber PJ. Histone deacetylase inhibition reduces myocardial ischemia-reperfusion injury in mice. *FASEB J* 2008;22:3549-3560.
- [98] Zhang L, Qin X, Zhao Y, Fast L, Zhuang S, Liu P, Cheng G, Zhao TC. Inhibition of histone deacetylases preserves myocardial performance and prevents cardiac remodeling through stimulation of endogenous angiomyogenesis. *J Pharmacol Exp Ther* 2012;341:285-293.
- [99] Choi JH, Nam KH, Kim J, Baek MW, Park JE, Park HY, Kwon HJ, Kwon OS, Kim DY, Oh GT. Trichostatin A exacerbates atherosclerosis in low density lipoprotein receptor-deficient mice. *Arterioscler Thromb Vasc Biol* 2005;25:2404-2409.
- [100] Wongcharoen W, Phrommintikul A. The protective role of curcumin in cardiovascular diseases. *Int J Cardiol* 2009;133:145-151.
- [101] Soni KB, Kuttan R. Effect of oral curcumin administration on serum peroxides and cholesterol levels in human volunteers. *Indian J Physiol Pharmacol* 1992;36:273-275.
- [102] Ramírez-Boscá A, Soler A, Carrión MA, Díaz-Alperi J, Bernd A, Quintanilla C, Quintanilla Almagro E, Miquel J. An hydroalcoholic extract of *Curcuma longa* lowers the abnormally high values of human-plasma fibrinogen. *Mech Ageing Dev* 2000;114:207-210.
- [103] Pan MH, Lai CS, Wu JC, Ho CT. Epigenetic and disease targets by polyphenols. *Curr Pharm Des* 2013;19:6156-6185.
- [104] Morimoto T, Sunagawa Y, Kawamura T, Takaya T, Wada H, Nagasawa A, Komeda M, Fujita M, Shimatsu A, Kita T, Hasegawa K. The dietary compound curcumin inhibits p300 histone acetyltransferase activity and prevents heart failure in rats. *J Clin Invest* 2008;118:868-878.

- [105] Gusterson RJ, Jazrawi E, Adcock IM, Latchman DS. The transcriptional co-activators CREB-binding protein (CBP) and p300 play a critical role in cardiac hypertrophy that is dependent on their histone acetyltransferase activity. *J Biol Chem* 2003;278:6838-6847.
- [106] Cavasin MA, Demos-Davies K, Horn TR, Walker LA, Lemon DD, Birdsey N, Weiser-Evans MC, Harral J, Irwin DC, Anwar A, Yeager ME, Li M, Watson PA, Nemenoff RA. Selective class I histone deacetylase inhibition suppresses hypoxia-induced cardiopulmonary remodeling through an antiproliferative mechanism. *Circ Res* 2012;110:739-748.
- [107] Liu F, Levin MD, Petrenko NB, Lu MM, Wang T, Yuan LJ, Stout AL, Epstein JA, Patel VV. Histone-deacetylase inhibition reverses atrial arrhythmia inducibility and fibrosis in cardiac hypertrophy independent of angiotensin. *J Mol Cell Cardiol* 2008;45:715-723.
- [108] Chen Y, Du J, Zhao YT, Zhang L, Lv G, Zhuang S, Qin G, Zhao TC. Histone deacetylase (HDAC) inhibition improves myocardial function and prevents cardiac remodeling in diabetic mice. *Cardiovasc Diabetol* 2015;14:99.
- [109] Fass DM, Shah R, Ghosh B, Hennig K, Norton S, Zhao WN, Reis SA, Klein PS, Mazitschek R, Maglathlin RL, Lewis TA, Haggarty SJ. Effect of Inhibiting Histone Deacetylase with Short-Chain Carboxylic Acids and Their Hydroxamic Acid Analogs on Vertebrate Development and Neuronal Chromatin. *ACS Med Chem Lett* 2010;2:39-42.
- [110] Cardinale JP, Sriramula S, Pariaut R, Guggilam A, Mariappan N, Elks CM, Francis J. HDAC inhibition attenuates inflammatory, hypertrophic, and hypertensive responses in spontaneously hypertensive rats. *Hypertension* 2010;56:437-444.
- [111] Findeisen HM, Kahles FK, Bruemmer D. Epigenetic Regulation of Vascular Smooth Muscle Cell Function in Atherosclerosis. *Curr Atheroscler Rep* 2013;15:319.
- [112] Wierda RJ, Rietveld IM, van Eggermond MCJA, Belien JAM, van Zwet EW, Lindeman JHN. Global histone H3 lysine 27 triple methylation levels are reduced in vessels with advanced atherosclerotic plaques. *Life Sci* 2015;129:3-9.
- [113] Kaneda R, Takada S, Yamashita Y, Choi YL, Nonaka-Sarukawa M, Soda M, Misawa Y, Isomura T, Shimada K, Mano H. Genome-wide histone methylation profile for heart failure. *Genes Cells* 2009;14:69-77.
- [114] Barrick CJ, Roberts RB, Rojas M, Rajamannan NM, Suitt CB, O'Brien KD, Smyth SS, Threadgill DW. Reduced EGFR causes abnormal valvular differentiation leading to calcific aortic stenosis and left ventricular hypertrophy in C57BL/6J but not 129S1/SvImJ mice. *Am J Physiol Heart Circ Physiol* 2009;297:65-75.
- [115] Papait R, Cattaneo P, Kunderfranco P, Greco C, Carullo P, Guffanti A, Viganò V, Stirparo GG, Latronico MV, Hasenfuss G, Chen J, Condorelli G. Genome-wide analysis of histone marks identifying an epigenetic signature of promoters and enhancers underlying cardiac hypertrophy. *Proc Natl Acad Sci U S A* 2013;110: 20164-20169.
- [116] Rodriguez-Iturbe B. Arteriolar remodeling in essential hypertension: are connective tissue growth factor and transforming growth factor involved? *Kidney Int* 2006;69:1104-1105.
- [117] Zhang D, Yu ZY, Cruz P, Kong Q, Li S, Kone BC. Epigenetics and the control of epithelial sodium channel expression in collecting duct. *Kidney Int* 2009;75:260-267.
- [118] Pojoga LH, Williams JS, Yao TM, Kumar A, Raffetto JD, do Nascimento GR, Reslan OM, Adler GK, Williams GH, Shi Y, Khalil RA. Histone demethylase LSD1 deficiency during high-salt diet is associated with enhanced vascular contraction, altered NO-cGMP

- relaxation pathway, and hypertension. *Am J Physiol Heart Circ Physiol* 2011;301:1862-1871.
- [119] Lee H, Rezai-Zadeh N, Seto E. Negative regulation of histone deacetylase 8 activity by cyclic AMP-dependent protein kinase A. *Mol Cell Biol* 2004;24:765-773.
- [120] Li C, Li Y, Li Y, Liu H, Sun Z, Lu J, Zhao Y. Glucocorticoid repression of human with-no-lysine (K) kinase-4 gene expression is mediated by the negative response elements in the promoter. *J Mol Endocrinol* 2008;40:3-12.
- [121] Mu S, Shimosawa T, Ogura S, Wang H, Uetake Y, Kawakami-Mori F, Marumo T, Yatomi Y, Geller DS, Tanaka H, Fujita T. Epigenetic modulation of the renal beta-adrenergic-WNK4 pathway in salt-sensitive hypertension. *Nat Med* 2011;17:573-580.
- [122] El-Osta A. Glycemic memory. *Curr Opin Lipidol* 2012;23:24-29.
- [123] Pirola L, Balcerczyk A, Tothill RW, Haviv I, Kaspi A, Lunke S, Ziemann M, Karagiannis T, Tonna S, Kowalczyk A, Beresford-Smith B, Macintyre G, Kelong M, Hongyu Z, Zhu J, El-Osta A. Genome-wide analysis distinguishes hyperglycemia regulated epigenetic signatures of primary vascular cells. *Genome Res* 2011;21:1601-1615.
- [124] Mutskov V., Raaka B., Felsenfeld G., Gershengorn M. The human insulin gene displays transcriptionally active epigenetic marks in islet-derived mesenchymal precursor cells in the absence of insulin expression. *Stem Cells* 2007;25:3223-3233.
- [125] Villeneuve L., Reddy M., Lanting L., Wang M., Meng L., Natarajan R. Epigenetic histone H3 lysine 9 methylation in metabolic memory and inflammatory phenotype of vascular smooth muscle cells in diabetes. *Proc Natl Acad Sci U S A* 2008;105:9047-9052.
- [126] El-Osta A, Brasacchio D, Yao D, Pocai A, Jones PL, Roeder RG, Cooper ME, Brownlee M. Transient high glucose causes persistent epigenetic changes and altered gene expression during subsequent normoglycemia. *J Exp Med* 2008;205:2409-2417.
- [127] Fichtlscherer S, De Rosa S, Fox H, Schwietz T, Fischer A, Liebetrau C, Weber M, Hamm CW, Röxe T, Müller-Ardogan M, Bonauer A, Zeiher AM, Dimmeler S. Circulating microRNAs in patients with coronary artery disease. *Circ Res* 2010;107: 677-684.
- [128] Hoekstra M, van der Lans CA, Halvorsen B, Gullestad L, Kuiper J, Aukrust P, van Berkel TJ, Biessen EA. The peripheral blood mononuclear cell microRNA signature of coronary artery disease. *Biochem Biophys Res Commun* 2010;394:792-797.
- [129] Sondermeijer BM, Bakker A, Halliani A, de Ronde MW, Marquart AA, Tijssen AJ, Mulders TA, Kok MG, Battjes S, Maiwald S, Sivapalaratnam S, Trip MD, Moerland PD, Meijers JC, Creemers EE, Pinto-Sietsma SJ. Platelets in patients with premature coronary artery disease exhibit upregulation of miRNA340* and miRNA624*. *PLoS One* 2011;6:e25946.
- [130] Schiano C, Vietri MT, Grimaldi V, Picascia A, Pascale MR, Napoli C. Epigenetic-related therapeutic challenges in cardiovascular disease. *Trends Pharmacol Sci* 2015;36:226-235.
- [131] Liu Y, Sanoff HK, Cho H, Burd CE, Torrice C, Mohlke KL, Ibrahim JG, Thomas NE, Sharpless NE. INK4/ARF transcript expression is associated with chromosome 9p21 variants linked to atherosclerosis. *PLoS ONE* 2009;4:e5027.
- [132] Cunnington MS, Santibanez Koref M, Mayosi BM, Burn J, Keavney B. Chromosome 9p21 SNPs associated with multiple disease phenotypes correlate with ANRIL expression. *PLoS Genet* 2010;6:e1000899.

- [133] Long G, Wang F, Duan Q, Chen F, Yang S, Gong W, Wang Y, Chen C, Wang DW. Human Circulating MicroRNA-1 and MicroRNA-126 as Potential Novel Indicators for Acute Myocardial Infarction. *Int J Biol Sci* 2012;8:811-818.
- [134] Ai J, Zhang R, Li Y, Pu J, Lu Y, Jiao J, Li K, Yu B, Li Z, Wang R, Wang L, Li Q, Wang N, Shan H, Li Z, Yang B. Circulating microRNA-1 as a potential novel biomarker for acute myocardial infarction. *Biochem Biophys Res Commun* 2010;391:73-77.
- [135] Cheng Y, Tan N, Yang J, Liu X, Cao X, He P, Dong X, Qin S, Zhang C. A translational study of circulating cell-free microRNA-1 in acute myocardial infarction. *Clin Sci (Lond)* 2010;119:87-95.
- [136] D'Alessandra Y, Devanna P, Limana F, Straino S, Di Carlo A, Brambilla PG, Rubino M, Carena MC, Spazzafumo L, De Simone M, Micheli B, Biglioli P, Achilli F, Martelli F, Maggolini S, Marenzi G, Pompilio G, Capogrossi MC. Circulating microRNAs are new and sensitive biomarkers of myocardial infarction. *Eur Heart J* 2010;31:2765-2773.
- [137] Adachi T, Nakanishi M, Otsuka Y, Nishimura K, Hirokawa G, Goto Y, Nonogi H, Iwai N. Plasma microRNA 499 as a biomarker of acute myocardial infarction. *Clin Chem* 2010;56:1183-1185.
- [138] Vausort M, Wagner DR, Devaux Y. Long noncoding RNAs in patients with acute myocardial infarction. *Circ Res* 2014;115:668-677.
- [139] Latronico MV, Condorelli G. MicroRNAs and cardiac pathology. *Nat Rev Cardiol* 2009;6:419-229.
- [140] Soifer HS, Rossi JJ, Saetrom P. MicroRNAs in disease and potential therapeutic applications. *Mol Ther* 2007;15:2070-2079.
- [141] Fukushima Y, Nakanishi M, Nonogi H, Goto Y, Iwai N. Assessment of plasma miRNAs in congestive heart failure. *Circ J* 2011;75:336-340.
- [142] Corsten MF, Dennert R, Jochems S, Kuznetsova T, Devaux Y, Hofstra L, Wagner DR, Staessen JA, Heymans S, Schroen B. Circulating MicroRNA-208b and MicroRNA-499 reflect myocardial damage in cardiovascular disease. *Circ Cardiovasc Genet* 2010;3:499-506.
- [143] Mora A, Davies AM, Bertrand L, Sharif I, Budas GR, Jovanović S, Mouton V, Kahn CR, Lucocq JM, Gray GA, Jovanović A, Alessi DR. Deficiency of PDK1 in cardiac muscle results in heart failure and increased sensitivity to hypoxia. *EMBO J* 2003;22:4666-4676.
- [144] Qureshi IA, Mehler MF. The emerging role of epigenetics in stroke: II. RNA regulatory circuitry. *Arch Neurol* 2010;67:1435-1441.
- [145] Jeyaseelan K, Lim KY, Armugam A. MicroRNA expression in the blood and brain of rats subjected to transient focal ischemia by middle cerebral artery occlusion. *Stroke* 2008;39:959-966.
- [146] Liu DZ, Tian Y, Ander BP, Xu H, Stamova BS, Zhan X, Turner RJ, Jickling G, Sharp FR. Brain and blood microRNA expression profiling of ischemic stroke, intracerebral hemorrhage, and kainate seizures. *J Cereb Blood Flow Metab* 2010;30: 92-101.
- [147] Cipollone F, Felicioni L, Sarzani R, Uchino S, Spigonardo F, Mandolini C, Malatesta S, Bucci M, Mammarella C, Santovito D, de Lutiis F, Marchetti A, Mezzetti A, Buttitta F. A unique microRNA signature associated with plaque instability in humans. *Stroke* 2011;42:2556-2563.
- [148] Bai Y, Nie S, Jiang G, Zhou Y, Zhou M, Zhao Y, Li S, Wang F, Lv Q, Huang Y, Yang Q, Li Q, Li Y, Xia Y, Liu Y, Liu J, Qian J, Li B, Wu G, Wu Y, Wang B, Cheng X, Yang

- Y, Ke T, Li H, Ren X, Ma X, Liao Y, Xu C, Tu X, Wang QK. Regulation of CARD8 expression by ANRIL and association of CARD8 single nucleotide polymorphism rs2043211 (p.C10X) with ischemic stroke. *Stroke* 2014;45: 383-388.
- [149] McKay JA, Mathers JC. Diet induced epigenetic changes and their implications for health. *Acta Physiol (Oxf)* 2011;202:103-118.
- [150] Milagro FI, Mansego ML, De Miguel C, Martinez JA. Dietary factors, epigenetic modifications and obesity outcomes: progresses and perspectives. *Mol Aspects Med* 2012;34:782-812.
- [151] Masuyama H, Hiramatsu Y. Effects of a high-fat diet exposure in utero on the metabolic syndrome-like phenomenon in mouse offspring through epigenetic changes in adipocytokine gene expression. *Endocrinology* 2012;153:2823–2830.
- [152] Bonna KH, Njolstad I, Ueland PM, Schirmer H, Tverdal A, Steigen T, Wang H, Nordrehaug JE, Arnesen E, Rasmussen K. Homocysteine lowering and cardiovascular events after acute myocardial infarction. *N Engl J Med* 2006;354:1578–1588.
- [153] Junien C, Gallou-Kabani C, Vige A, Gross MS. [Nutritional epigenomics: consequences of unbalanced diets on epigenetics processes of programming during lifespan and between generations]. *Ann Endocrinol (Paris)* 2005;66:2S19-28.
- [154] Hanley B, Dijane J, Fewtrell M, Grynberg A, Hummel S, Junien C, Koletzko B, Lewis S, Renz H, Symonds M, Gros M, Harthoorn L, Mace K, Samuels F, van Der Beek EM. Metabolic imprinting, programming and epigenetics—a review of present priorities and future opportunities. *Br J Nutr* 2010;104:1-25.
- [155] Oommen AM, Griffin JB, Sarath G, Zempleni J. Roles for nutrients in epigenetic events. *J Nutr Biochem* 2005;16:74-77.
- [156] Hong X, Wang X. Early life precursors, epigenetics, and the development of food allergy. *Semin Immunopathol* 2012;34:655-669.
- [157] Burdge GC, Hoile SP, Lillycrop KA. Epigenetics: are there implications for personalised nutrition? *Curr Opin Clin Nutr Metab Care* 2012;15:442-447.
- [158] Hossein-Nezhad A, Holick MF. Optimize dietary intake of vitamin D: an epigenetic perspective. *Curr Opin Clin Nutr Metab Care* 2012;15:567-579.
- [159] Duthie SJ. Epigenetic modifications and human pathologies: cancer and CVD. *Proc Nutr Soc* 2011;70:47-56.
- [160] Whayne TH. Epigenetics in the development, modification, and prevention of cardiovascular disease. *Mol Biol Rep* 2015;42:765-776.
- [161] Linask KK, Huhta J. Folate protection from congenital heart defects linked with canonical Wnt signaling and epigenetics. *Curr Opin Pediatr* 2010;22:561-566.
- [162] Bogdarina I, Welham S, King PJ, Burns SP, Clark AJ. Epigenetic modification of the renin-angiotensin system in the fetal programming of hypertension. *Circ Res* 2007;100:520-526.
- [163] Lillycrop KA, Phillips ES, Jackson AA, Hanson MA, Burdge GC. Dietary protein restriction of pregnant rats induces and folic acid supplementation prevents epigenetic modification of hepatic gene expression in the offspring. *J Nutr* 2005;135: 1382-1386.
- [164] Lillycrop KA, Slater-Jefferies JL, Hanson MA, Godfrey KM, Jackson AA, Burdge GC. Induction of altered epigenetic regulation of the hepatic glucocorticoid receptor in the offspring of rats fed a protein-restricted diet during pregnancy suggests that reduced DNA methyltransferase-1 expression is involved in impaired DNA methylation and changes in histone modifications. *Br J Nutr* 2007;97:1064-1073.

- [165] Das SK, Chakrabarti R. Role of PPAR in cardiovascular diseases. *Recent Pat Cardiovasc Drug Discov* 2006;1:193–209.
- [166] Walker BR. Glucocorticoids and cardiovascular disease. *Eur J Endocrinol* 2007;157:545–559.
- [167] Godfrey KM, Inskip HM, Hanson MA. The long-term effects of prenatal development on growth and metabolism. *Semin Reprod Med* 2011;29:257–265.
- [168] Barker DJ. The developmental origins of well-being. *Philos Trans R Soc Lond B Biol Sci* 2004;359:1359–1366.
- [169] Painter RC, Roseboom TJ, Bleker OP. Prenatal exposure to the Dutch famine and disease in later life: an overview. *Reprod Toxicol* 2005;20:345–352.
- [170] Godfrey KM, Barker DJ. Fetal programming and adult health. *Public Health Nutr* 2001;4:611–624.
- [171] Weaver IC, Cervoni N, Champagne FA, D'Alessio AC, Sharma S, Seckl JR, Dymov S, Szyf M, Meaney MJ. Epigenetic programming by maternal behavior. *Nat Neurosci* 2004;7:847–854.
- [172] Gluckman PD, Hanson MA, Buklijas T, Low FM, Beedle AS. Epigenetic mechanisms that underpin metabolic and cardiovascular diseases. *Nat Rev Endocrinol* 2009;5:401–408.
- [173] Kuzawa CW, Sweet E. Epigenetics and the embodiment of race: developmental origins of US racial disparities in cardiovascular health. *Am J Hum Biol* 2009;21:2–15.
- [174] Wheatley KE, Nogueira LM, Perkins SN, Hursting SD. Differential effects of calorie restriction and exercise on the adipose transcriptome in diet-induced obese mice. *J Obes* 2011;26:5417.
- [175] Sanchis-Gomar F, Garcia-Gimenez JL, Perez-Quilis C, Gomez-Cabrera MC, Pallardo FV, Lippi G. Physical exercise as an epigenetic modulator. Eustress, the “positive stress” as an effector of gene expression. *J Strength Cond Res* 2012;26:3469–3472.
- [176] McGee SL, Fairlie E, Garnham AP, Hargreaves M. Exercise-induced histone modifications in human skeletal muscle. *J Physiol* 2009;587:5951–5958.
- [177] Bloch W, Suhr F, Zimmer P. Molecular mechanisms of exercise-induced cardiovascular adaptations. Influence of epigenetics, mechanotransduction and free radicals. *Herz* 2012;37:508–515.
- [178] Baccarelli A, Ghosh S. Environmental exposures, epigenetics and cardiovascular disease. *Curr Opin Clin Nutr Metab Care* 2012;15:323–329.
- [179] Wrobel K, Caruso JA. Epigenetics: an important challenge for ICP-MS in metallomics studies. *Anal Bioanal Chem* 2009;393:481–486.
- [180] Breitling LP. Current genetics and epigenetics of smoking/tobacco-related cardiovascular disease. *Arterioscler Thromb Vasc Biol* 2011;33:1468–1472.
- [181] Ordovas JM, Smith CE. Epigenetics and cardiovascular disease. *Nat Rev Cardiol* 2010;7:510–519.
- [182] Suter M, Ma J, Harris A, Patterson L, Brown KA, Shope C, Showalter L, Abramovici A, Aagaard-Tillery KM. Maternal tobacco use modestly alters correlated epigenome-wide placental DNA methylation and gene expression. *Epigenetics* 2011;6:1284–1294.
- [183] NCEP. Third Report of the National Cholesterol Education Program (NCEP) Expert Panel on Detection, Evaluation, and Treatment of High Blood Cholesterol in Adults (Adult Treatment Panel III) final report. *Circulation* 2002;106:3143–3421.

- [184] Flowers E, Aouizerat BE. MicroRNA associated with dyslipidemia and coronary disease in humans. *Physiol Genomics* 2013;45:1199-1205.
- [185] Esau C, Davis S, Murray SF, Yu XX, Pandey SK, Pear M, Watts L, Booten SL, Graham M, McKay R, Subramaniam A, Propp S, Lollo BA, Freier S, Bennett CF, Bhanot S, Monia BP. miR-122 regulation of lipid metabolism revealed by in vivo antisense targeting. *Cell Metab* 2006;3:87-98.
- [186] Krützfeldt J, Rajewsky N, Braich R, Rajeev KG, Tuschl T, Manoharan M, Stoffel M. Silencing of microRNAs in vivo with 'antagomirs.' *Nature* 2005;438:685-689.
- [187] Flowers E, Singh K, Molina C, Mathur A, Aouizerat BE. MicroRNA associated with atherogenic dyslipidemia in South Asian men. *Int J Cardiol* 2013;168:4884-4885.
- [188] Rayner KJ, Suárez Y, Dávalos A, Parathath S, Fitzgerald ML, Tamehiro N, Fisher EA, Moore KJ, Fernández-Hernando C. MiR-33 contributes to the regulation of cholesterol homeostasis. *Science* 2010;328:1570-1573.
- [189] Richardson K, Nettleton JA, Rotllan N, Tanaka T, Smith CE, Lai CQ, Parnell LD, Lee YC, Lahti J, Lemaitre RN, Manichaikul A, Keller M, Mikkila V, Ngwa J, van Rooij FJ, Ballentyne CM, Borecki IB, Cupples LA, Garcia M, Hofman A, Ferrucci L, Mozaffarian D, Peralta MM, Raitakari O, Tracy RP, Arnett DK, Bandinelli S, Boerwinkle E, Eriksson JG, Franco OH, Kahonen M, Nalls M, Siscovick DS, Houston DK, Psaty BM, Viikari J, Witteman JC, Goodarzi MO, Lehtimäki T, Liu Y, Zillikens MC, Chen YD, Uitterlinden AG, Rotter JJ, Fernandez-Hernando C, Ordovas JM. Gain-of-function lipoprotein lipase variant rs13702 modulates lipid traits through disruption of a microRNA-410 seed site. *Am J Hum Genet* 2013;92:5-14.
- [190] Herrera BM, Lockstone HE, Taylor JM, Ria M, Barrett A, Collins S, Kaisaki P, Argoud K, Fernandez C, Travers ME, Grew JP, Randall JC, Gloyn AL, Gauguier D, McCarthy MI, Lindgren CM. Global microRNA expression profiles in insulin target tissues in a spontaneous rat model of type 2 diabetes. *Diabetologia* 2010;53:1099-1109.
- [191] Trajkovski M, Hausser J, Soutschek J, Bhat B, Akin A, Zavolan M, Heim MH, Stoffel M. MicroRNAs 103 and 107 regulate insulin sensitivity. *Nature* 2011;474:649-653.
- [192] van Rooij E, Quiat D, Johnson BA, Sutherland LB, Qi X, Richardson JA, Kelm RJ Jr, Olson EN. A family of microRNAs encoded by myosin genes governs myosin expression and muscle performance. *Dev Cell* 2009;17:662-673.
- [193] Shan H, Zhang Y, Lu Y, Zhang Y, Pan Z, Cai B, Wang N, Li X, Feng T, Hong Y, Yang B. Downregulation of miR-133 and miR-590 contributes to nicotine-induced atrial remodelling in canines. *Cardiovasc Res* 2009;83:465-472.
- [194] Karolina DS, Tavintharan S, Armugam A, Sepramaniam S, Pek SL, Wong MT, Lim SC, Sum CF, Jeyaseelan K. Circulating miRNA profiles in patients with metabolic syndrome. *J Clin Endocrinol Metab* 2012;97:2271-2276.
- [195] Manolopoulos VG, Ragia G, Tavridou A. Pharmacogenomics of oral antidiabetic medications: current data and pharmacoeconomic perspective. *Pharmacogenomics* 2011;12:1161-1191.
- [196] Mai A, Altucci L. Epi-drugs to fight cancer: from chemistry to cancer treatment, the road ahead. *Int J Biochem Cell Biol* 2009;41:199-213.
- [197] Porter KE, Riches K. The vascular smooth muscle cell: a therapeutic target in Type 2 diabetes? *Clin Sci (Lond)* 2013;125:167-182.
- [198] Bieliauskas AV, Pflum MK. Isoform-selective histone deacetylase inhibitors. *Chem Soc Rev* 2008;37:1402-1413.

- [199] Pons D, de Vries FR, van den Elsen PJ, Heijmans BT, Quax PH, Jukema JW. Epigenetic histone acetylation modifiers in vascular remodelling: new targets for therapy in cardiovascular disease. *Eur Heart J* 2009;30:266-277.
- [200] Naeem N, Haneef K, Kabir N, Iqbal H, Jamall S, Salim A. DNA methylation inhibitors, 5-azacytidine and zebularine potentiate the transdifferentiation of rat bone marrow mesenchymal stem cells into cardiomyocytes. *Cardiovasc Ther* 2013;31:201-209.
- [201] Antonitsis P, Ioannidou-Papagiannaki E, Kaidoglou A, Papakonstantinou C. In vitro cardiomyogenic differentiation of adult human bone marrow mesenchymal stem cells. The role of 5-azacytidine. *Interact Cardiovasc Thorac Surg* 2007;6:593-597.
- [202] Zhou L, Cheng X, Connolly BA, Dickman MJ, Hurd PJ, Hornby DP. Zebularine: a novel DNA methylation inhibitor that forms a covalent complex with DNA methyltransferases. *J Mol Biol* 2002;321:591-599.
- [203] Horrillo A, Pezzolla D, Fraga MF, Aguilera Y, Salguero-Aranda C, Tejedo JR, Martin F, Bedoya FJ, Soria B, Hmadcha A. Zebularine regulates early stages of mESC differentiation: effect on cardiac commitment. *Cell Death Dis* 2013;4:e570.
- [204] Connelly JJ, Cherepanova OA, Doss JF, Karaoli T, Lillard TS, Markunas CA, Nelson S, Wang T, Ellis PD, Langford CF, Haynes C, Seo DM, Goldschmidt-Clermont PJ, Shah SH, Kraus WE, Hauser ER, Gregory SG. Epigenetic regulation of COL15A1 in smooth muscle cell replicative aging and atherosclerosis. *Hum Mol Genet* 2013;22:5107-5120.
- [205] Mitro N, Godio C, De Fabiani E, Scotti E, Galmozzi A, Gilardi F, Caruso D, Vigil Chacon AB, Crestani M. Insights in the regulation of cholesterol 7 α -hydroxylase gene reveal a target for modulating bile acid synthesis. *Hepatology* 2007;46:885-897.
- [206] Balasubramanyam K, Varier RA, Altaf M, Swaminathan V, Siddappa NB, Ranga U, Kundu TK. Curcumin, a novel p300/CREB-binding protein-specific inhibitor of acetyltransferase, represses the acetylation of histone/nonhistone proteins and histone acetyltransferase-dependent chromatin transcription. *J Biol Chem* 2004;279:51163-51171.
- [207] Bauersachs J, Thum T. Biogenesis and regulation of cardiovascular microRNAs. *Circ Res* 2011;109:334-347.
- [208] Udali S, Guarini P, Moruzzi S, Choi SW, Friso S. Cardiovascular epigenetics: from DNA methylation to microRNAs. *Mol Aspects Med* 2013;34:883-901.
- [209] Widera C, Gupta SK, Lorenzen JM, Bang C, Bauersachs J, Bethmann K, Kempf T, Wollert KC, Thum T. Diagnostic and prognostic impact of six circulating microRNAs in acute coronary syndrome. *J Mol Cell Cardiol* 2011;51:872-875.
- [210] Zile MR, Mehurg SM, Arroyo JE, Stroud RE, DeSantis SM, Spinale FG. Relationship between the temporal profile of plasma microRNA and left ventricular remodeling in patients after myocardial infarction. *Circ Cardiovasc Genet* 2011;4:614-619.
- [211] Qiang L, Hong L, Ningfu W, Huaihong C, Jing W. Expression of miR-126 and miR-508-5p in endothelial progenitor cells is associated with the prognosis of chronic heart failure patients. *International Journal of Cardiology* 2013;168:2082-2088.
- [212] Loyer X, Potteaux S, Vion AC, Guérin CL, Boulkroun S, Rautou PE, Ramkhalawon B, Esposito B, Dalloz M, Paul JL, Julia P, Maccario J, Boulanger CM, Mallat Z, Tedgui A. Inhibition of microRNA-92a prevents endothelial dysfunction and atherosclerosis in mice. *Circ Res* 2014;114:434-443.

- [213] Seok HY, Chen J, Kataoka M, Huang ZP, Ding J, Yan J, Hu X, Wang DZ. Loss of MicroRNA-155 protects the heart from pathological cardiac hypertrophy. *Circ Res* 2014;114:1585-1595.
- [214] Welten SM, Bastiaansen AJ, de Jong RC, de Vries MR, Peters EA, Boonstra MC, Sheikh SP, La Monica N, Kandimalla ER, Quax PH, Nossent AY. Inhibition of 14q32 MicroRNAs miR-329, miR-487b, miR-494, and miR-495 increases neovascularization and blood flow recovery after ischemia. *Circ Res* 2014;115:696-708.
- [215] Takahashi Y, Satoh M, Minami Y, Tabuchi T, Itoh T, Nakamura M. Expression of miR-146a/b is associated with the Toll-like receptor 4 signal in coronary artery disease: effect of renin-angiotensin system blockade and statins on miRNA-146a/b and Toll-like receptor 4 levels. *Clin Sci (Lond)* 2010;119:395-405.
- [216] Montgomery RL, Hullinger TG, Semus HM, Dickinson BA, Seto AG, Lynch JM, Stack C, Latimer PA, Olson EN, van Rooij E. Therapeutic inhibition of miR-208a improves cardiac function and survival during heart failure. *Circulation* 2011;124:1537-1547.
- [217] Yang B, Lin H, Xiao J, Lu Y, Luo X, Li B, Zhang Y, Xu C, Bai Y, Wang H, Chen G, Wang Z. The muscle-specific microRNA miR-1 regulates cardiac arrhythmogenic potential by targeting GJA1 and KCNJ2. *Nat Med* 2007;13:486-491.
- [218] Fujita J. Report of the American Heart Association (AHA). Scientific sessions 2012, Los Angeles. *Circ J* 2013;77:35-40.
- [219] Sun X, Fu X, Han W, Zhao Y, Liu H. Can controlled cellular reprogramming be achieved using microRNAs? *Ageing Res Rev* 2010;9:475-483.
- [220] Jungebluth P, Alici E, Baiguera S, Le Blanc K, Blomberg P, Bozoky B, Crowley C, Einarsson O, Grinnemo KH, Gudbjartsson T, Le Guyader S, Henriksson G, Hermanson O, Juto JE, Leidner B, Lilja T, Liska J, Luedde T, Lundin V, Moll G, Nilsson B, Roderburg C, Stromblad S, Sutlu T, Teixeira AI, Watz E, Seifalian A, Macchiarini P. Tracheobronchial transplantation with a stem-cell-seeded bioartificial nanocomposite: a proof-of-concept study. *Lancet* 2011;378:1997-2004.

Chapter 64

**INSIGHTS FROM EPIGENOMICS ANALYSIS
OF SPERM: SPERM DEVELOPMENT
AND MALE INFERTILITY**

***Dheepa Balasubramanian¹, PhD,
Eisa Tahmasbpour², PhD
and Ashok Agarwal^{3,*}, PhD, HCLD, ABB, EMB, ACE***

¹Case Western Reserve University, Cleveland, OH, US

²Chemical Injury Research Center, Baqiatallah Medical Science University, Tehran, Iran

³Center for Reproductive Medicine, Glickman Urological and Kidney Institute,
Cleveland Clinic and Obstetrics and Gynecology and Women's Health Institute,
Cleveland, OH, US

ABSTRACT

Extensive characterization has been performed on the genes and genetic mutations that are involved in spermatogenesis and male infertility, but the vast role of the sperm's transcriptome and epigenome in male reproduction has yet to be completely explored. Recent research has established that epigenetic remodeling of the sperm is necessary for development and for its function following fertilization. The histone- retained regions of the sperm have been recently shown to carry the bivalent marks of activating histone H3 lysine K4 trimethylation and the repressive H3 lysine K27 trimethylation. These bivalent histone marks facilitate the dynamic changes in stage-specific gene expression during sperm development. The paternal epigenome bears unique and important epigenetic modifications determined to be potentially important to the developing embryo, but the complete scope of its contribution is still emerging. Additionally, advances in assisted reproductive techniques have also suggested that alterations in the epigenetic profile in infertile men are transmitted to the developing embryo. This review will highlight the latest advances in epigenome profiling of the chromatin modifications during the development of immature to a mature sperm as well as provide a glimpse into the future role of epigenetic mechanisms in the generation of new germ cells/gametes from induced pluripotent stem cells, to treat male infertility.

* Corresponding Author's Email: agarwaa@ccf.org.

Keywords: epigenetics, male infertility, epigenome, histone modification

ABBREVIATIONS

AR	Androgen receptor
ART	Assisted reproductive technology
AZF	Azoospermia a/b/c
BRDT	Bromodomain testis-specific protein
CFTR	Cystic fibrosis transmembrane conductance regulator
DAZ	Deleted in Azoospermia
DMR	differentially methylated regions
DNMT1/3a/3b	DNA methyltransferase
ESC	embryonic stem cells
HAT	histone acetyltransferases
HDAC	histone deacetylases
ICSI	Intra-cytoplasmic sperm injection
IGF2	Insulin-like growth factor
IVF	In vitro fertilization
KLF4	Krueppel-like factor 4
iPS/iPSCs	Induced pluripotent stem cells
MTHFR	methylene tetrahydrofolate reductase
MIWI2	PIWI protein
PRM1/2	Protamine 1/2
PGC	Primordial germ cells
PGCLC	primordial germ cell-like cells
SAM	S-adenosyl methionine
TP	transition proteins
OCT4	octamer-binding transcription factor 4
SCOS	sertoli cell-only syndrome
SSC	spermatogonial stem cells
SNRPN	small nuclear ribonucleoprotein polypeptide

INTRODUCTION

1. SPERM CELL DEVELOPMENT

In the past, it was believed that the sperm is but a vehicle to transfer the male genome into the oocyte, but increasing evidence now suggests that sperm cells play a substantial role in initiating embryo development, and environmental influences on the paternal genome can adversely affect male fertility as well as babies conceived via assisted reproductive technology (ART) [1].

Spermatogenesis involves the precise and well-controlled differentiation of germinal cells from spermatogonia to spermatozoa [2, 7]. This highly orchestrated process is a continuous process, characterized by a pre-meiotic, meiotic and a post-meiotic stage. Spermatogonia divide by mitosis, followed by meiosis through the formation of primary spermatocytes. As spermatocytes, they replicate DNA and undergo meiosis. During the meiosis stages, pairing of homologous chromosome, recombination, replacement of somatic histones with variants occur, and finally, at the post-meiotic stage, spermatids form into compact spermatozoa [3, 4].

Active transcription occurs during the stages of spermatocytes and round spermatids, but the later stages of spermatogenesis are relatively transcriptionally inert due to the compaction of the chromatin. Transcriptional, post-transcriptional, epigenetic regulatory networks direct the multiple stages of germ cell specification, differentiation and development [5]. Sperm chromatin is essential for sperm function as well as for embryonic development; defective sperm chromatin leads to abortion, decreased fertility, loss of genomic integrity or assisted reproductive failure [6].

2. EPIGENETIC REPROGRAMMING OF MALE GERM CELLS

Epigenetics is dominantly involved in the preservation of cellular identity and lineage fidelity, through the different mechanisms of alterations in chromatin structure, modifications of DNA and histones, remodeling of nucleosomes etc. Global reprogramming of epigenetic marks, in addition to meiotic chromatin organization and recombination as well as compaction of sperm genome, in the primordial germ cells (PGC) occurs during gametogenesis and in zygote following fertilization (Figure 1) [2]. The first reprogramming event occurs in the developing gonad. The reversible nature of epigenetic processes allows for the resetting of epigenetic information in the germ cells, in order to equip the sperm and oocyte to direct embryonic and post-natal development [7].

Dynamic epigenetic alteration is central to differentiation of mammalian sperm; however, the nature of these changes largely remains unknown. Spermatogenesis is a continuous and precisely controlled process that is accompanied by a dramatic reorganization of chromatin from a nucleosomal histone-based structure to a structure largely based on protamines [8]. During the spermatogenesis process, upon histone H4 hyperacetylation, nucleosomes are disassembled, DNA breaks occur and incorporation of non-canonical histone variants is made possible [9]. Histones are widely replaced by highly basic proteins, first by transition proteins and subsequently by the two protamines PRM1 and PRM2 [10].

Upon fertilization, the products of germ-cell development, the oocyte and sperm cell, fuse to form a zygote, a totipotent structure that can develop into a whole new organism [11]. Following fertilization, histones seem to remain associated with the paternal genome, in order, to contribute to zygotic chromatin despite the extensive reorganization of the sperm chromatin. Also, the sperm DNA is decondensed from its highly compacted and transcriptionally quiescent state, and expands to the inducible state found in the paternal pronucleus [12]. The paternal histones could therefore serve as a template for the incorporation of newly synthesized histones during replication in the zygote. Accordingly, the programmatic chromatin packaging in sperm could potentially deliver epigenetic information to the oocyte and the zygote post-fertilization [13].

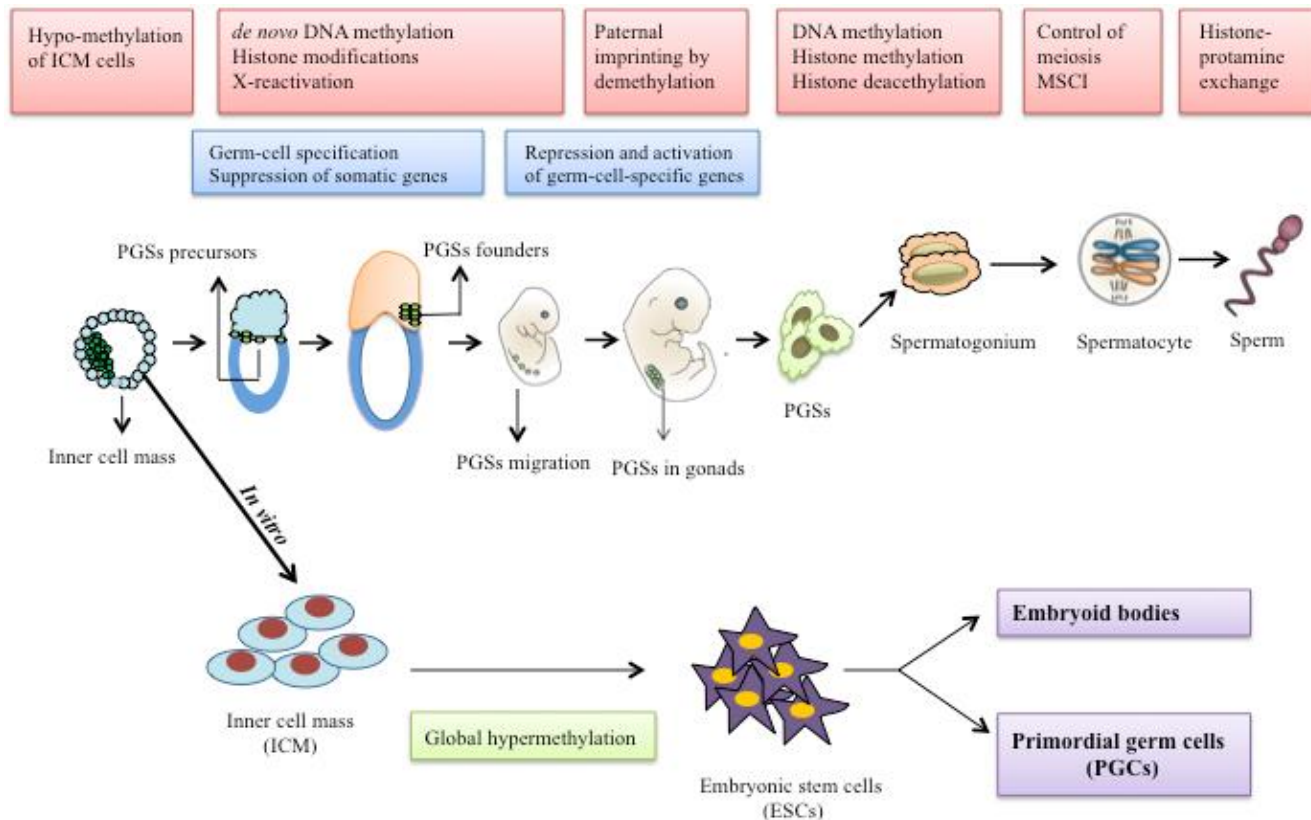


Figure 1. Epigenetics govern the regulation of processes from the initial stages of spermatogenesis to embryo development.

2.1. DNA Methylation

DNA methylation occurs largely on CpG islands located upstream of genes [14]. It is a process in which a methyl group from S-adenosyl methionine (SAM) is transferred to a cytosine residue to create 5-methyl cytosine (5mC). This process is catalyzed by a family of closely related DNA methyl transferases (DNMT1, DNMT3a and DNMT3b) [15]. The passive process takes place during replication of newly synthesized DNA strands by DNMT1.

The purpose of DNA demethylation in PGCs is to prevent transmission of inappropriate epigenetic information to the next generation. It is interesting to note that genome-wide methylation patterns in sperm differ markedly from that of somatic cells, highlighting the sperm-specific role of DNA methylation [8]. DNA methylation patterns are first acquired during gametogenesis. Evidence for the significance of DNA methylation for male germ cell development has been demonstrated in the mouse, where gene targeting of enzymes involved in catalyzing DNA methylation, the DNA (cytosine-5)-methyltransferases (DNMTs), results in loss of germ cell methylation, and failure of meiosis and infertility [16, 17, 18]. Recent studies have shown that mono methyl, dimethyl and trimethyl modifications of H3K4, H3K9 or H3K27 display tightly controlled temporal expression and ensure proper progression through spermatogenesis [19]. The level of H3K4 methylation peaks in the spermatogonial stem cell stage, and a targeted loss of H3K4 methylation, caused by reduction of Mll2 (an H3K4 methyl transferase) activity results in a dramatic reduction in the number of spermatocytes, suggesting that H3K4 methylation is essential for the exit from the stem cell stage and commitment to become a spermatocyte [20, 21, 22]. In contrast, H3K9 methylation and H3K27 methylation are low in the stem cell and increase during meiosis, persisting long after meiosis is complete, presumably to ensure gene-silencing [23]. Methylation on lysine 9 of histone H3 (H3K9me) is also associated with the sex chromosomes, euchromatin and heterochromatin in the late pachytene stage; however, the levels of H3K9 methylation drop upon completion of meiosis [24, 25]. This reduction in H3K9me is concurrently associated with an increase in H3K4me levels [7].

2.2. Histone Modifications

Histone modifications include the various post-translational alterations on the lysine-rich tail region of histones, especially of H3 and H4 histones [26]. However, histone acetylation, methylation and phosphorylation are more relevant during sperm development, spermatogenesis, fertilization and embryo development (Figure 2) [27].

The compactness of DNA is determined by histone proteins that are bound to DNA [28]. During the post-meiotic maturation of spermatids into spermatozoa or spermiogenesis, the genome of the male germ cell is entirely restructured. Indeed, while the core of the spermatid extends and condenses, most somatic histones are removed and replaced progressively by proteins called “transition” or TP, which are then replaced with specific nuclear proteins of sperm nucleus known as protamines [29, 30]. Protamines aid in the tight packaging of chromatin and are protective against the effects of harmful physical or chemical stressors [29]. In humans, histone to protamine transformation is considered to be incomplete, and the residual histones are highly acetylated and located in the annular region [29]. However, a small amount

of histone proteins is retained in spermatids, and chemical modification of these histones play an important role in sperm formation.

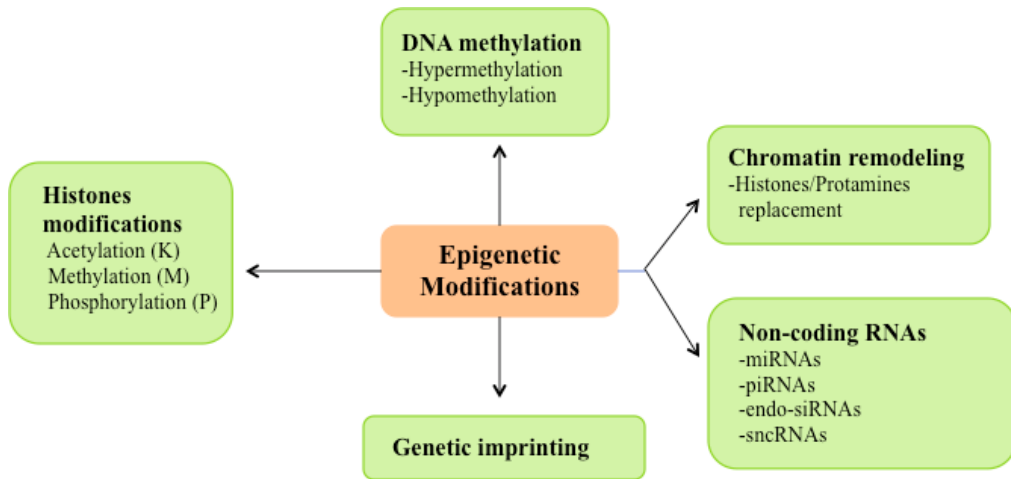


Figure 2. Epigenetic modifications involved in Sperm development.

During spermiogenesis, only 4% of the genome is retained in nucleosomes, while the rest of the histones are replaced with protamines. But the retained histones are overwhelmingly found at CpG islands near key developmental genes. The fetal germ cells contain several genes in a poised conformation, with both the H3K4me3 and H3K27me3 modifications [23, 25, 31]. Based on these findings, it can be presumed that the poised state of chromatin in germ cells are necessary for germ cell identity, to prepare for totipotency following fertilization and possibly to prevent DNA methylation at key developmental promoters [22].

The acetylated histone marks H3K4ac and H3K39ac are associated with transcriptional activation. There is growing interest in histone acetyl transferase (HAT) and histone deacetylase (HDAC) proteins involved in histone acetylation and spermatogenesis because they modulate gene expression and serve as a trigger for chromatin reorganization and remodeling [19, 32]. Waves of histone acetylation occur throughout human spermatogenesis. For example, hyperacetylation of histone-H4 was observed in spermatogonia, pre-leptotene spermatocyte and spermatids [19]. It facilitates the interaction of non-histone chromosomal proteins, such as transcription factors, with the chromatin. In spermatids, hyperacetylation of histone-H4 recruits bromodomain protein (BRDT) that triggers nuclear reorganization. Recent studies have demonstrated a statistically significant reduction in the percentage of spermatogonia with Hypac-H4 or Lys12ac-H4 [33]. Furthermore, hypac-H4 is believed to trigger nuclear reorganization and probably facilitate exchange of histones by protamines that allow condensation of the spermatid nucleus [33]. Another study demonstrated increased acetylation of histone H4 may be associated with a histone-to-protamine substitution, suggesting specific relationship between histone H4 modification and gene expression during spermatogenesis [19].

Every stage of spermatocyte development shows specific patterns of H4 modification. In pre-leptotene spermatocytes, the expression of H4me, H4K20me2, H4K5ac, H4K8ac and H4K12ac was high, and that of H4me3 was relatively low. Whether acetylation itself is only associated with gene activation is still not certain because it has been recently shown that

protamines also contain acetyl groups. Studies have also alluded to other post-translational modifications, such as S42 and K49 acetylation of protamine-1 and K64 acetylation on protamine-2 [34]. The fact that protamines carry marks associated with transcriptional activation is intriguing, as these proteins are thought to ensure tight packaging of the DNA in sperm cells and contribute to transcriptional silencing. Although histone acetylation is a characteristic feature of transcriptional active genes, it is known that spermatozoa are transcriptionally inactive. This obvious inconsistency resulted in the hypothesis that histone acetylation represents an epigenetic mark that is transmitted from sperm to oocyte and involved in the regulation of gene expression in the early embryo [34].

2.3. Chromatin Remodeling

Chromatin remodeling is a process in which ATP-dependent chromatin remodeling complexes (SWI/SNF, ISW1 and MI-2) use energy to alter the location and structure of nucleosomes [35]. The remodeling process activates or represses gene expression, as required for proper meiotic development and maturation of gametes [28]. It is a critical step in gamete development, and the compacted structure of the chromatin transmits vital information to the embryo to guide it through development. The lack of proper DNA packaging in sperm has been associated with infertility in mice [36]. During the chromatin repackaging process in spermatogenesis, about 85% of the histones are replaced by protamines [12]. In the initial stages of spermiogenesis, histones are hyperacetylated and undergo other modifications [36]. Subsequently, at the final stage of spermatogenesis, the nucleosomal structure is progressively disassembled, then replaced by transition proteins (TPs) and finally by protamines [37].

The incorporation of protamines into sperm chromatin induces DNA compaction that is important for the formation of spermatozoa [30]. However, it is known that protamines are phosphorylated before binding to DNA and that substantial dephosphorylation takes place concomitant with nucleo-protamine maturation. Indeed, mutation of the calmodulin-dependent protein kinase Camk4 that phosphorylates protamine 2, results in defective spermiogenesis and male infertility [38]. At fertilization, the mature sperm cells are packaged densely with protamines, whereas maternal genome is packaged with histones. Subsequently, upon fertilization, the highly compact nucleoprotamine structure must be unpacked and reorganized into a nucleosomal structure [39]. Recent studies have shown that epigenetic errors in these processes are a possible cause of male infertility. Both protamines 1 and 2 are essential for sperm function, and the haploinsufficiency of either P1 or P2 results in a reduced amount of the respective protein, abnormalities in the structure of the chromatin, DNA damage and infertility [28]. Furthermore, it is known that the optimal protamine-1 to protamine-2 ratio (P1/P2 ratio) is critical and well regulated [40]. Indeed, the P1/P2 ratio in fertile men ranges from 0.8–1.2 and deviation from this ratio has been shown to lead to infertility [41]. A change in either direction of this ratio adversely affects semen quality, DNA integrity and fertility in men. Patients with abnormally depressed or elevated P1/P2 ratios are characterized by poor sperm concentration, motility and morphology as well as decreased fertilization capabilities. Impaired spermatogenesis has been reported to be associated with aberrant H4 acetylation [42]. A study by Sonnack et al., showed that men exhibiting qualitative or quantitative infertility have significantly decreased levels of histone H4 acetylation associated with impaired spermatogenesis [43]. Hyperacetylation of histone H4 is required in the transition from histones to protamines. This step decreases the affinity of the interaction between the sperm histones and DNA to allow for the exchange of transition proteins to occur [28]. H4 hyperacetylation

was also observed in infertile men exhibiting sertoli-cell-only syndrome (SCOS) [44]. A significant difference in transcript expression of chromatin remodeling factors between normal spermatogenesis and the stage of round spermatid maturation arrest has been found in a previous study, suggesting impaired epigenetic information and aberrant transcription during sperm development. This could represent one possible reason for the developmental arrest of round spermatids [45].

2.4. Small RNAs

Spermatozoa contain numerous populations of RNAs. Small RNAs that have been enriched in sperm include microRNA (miRNA), endogenous small interfering RNAs (endo-siRNA), small non-coding RNA (sncRNA) and PIWI-interacting RNAs (piRNAs). Alterations in any one of the populations of RNAs in sperm indicate sperm abnormalities [3]. The importance of piRNAs in germline is very evident through their role of silencing of transposable elements during DNA demethylation in PGCs [46]. Mice knockout studies, where Dicer 1 was specifically deleted in spermatogonia, revealed prominent defects during haploid differentiation, specifically, in nuclei elongation and chromatin organization. Speculated roles for endo-siRNAs in germ cells include silencing and regulation of heterochromatin formation and maintenance. Among the piRNAs, MILI and MIWI2 together with pre-pachytene piRNAs, participate in silencing of transposable elements in fetal and neonatal germ cells. In the absence of these piRNAs in mice, it was shown that transposons were uncontrollably expressed, leading to meiotic arrest and sterility eventually [47].

3. EPIGENETIC DYSREGULATION: MALE INFERTILITY

3.1. Male Infertility

Human infertility, caused largely by a deficiency of sperm cells, is a major health problem that affects ~15% of couples globally [48]. It is a multifactorial syndrome encompassing a wide variety of disorders (Figure 3). Anatomic defects underlying male infertility include varicocele, vesicular damage due to torsion, obstruction of testicular sperm passage and ejaculatory failures, genital tract infections, gametogenesis dysfunction, molecular genetics disorders, endocrine disturbances and immunologic problems [48-50]. Additionally, factors such as life style, environment and smoking have also been reported to affect gamete and embryo development [51, 52]. In order to understand potential regulatory mechanisms involved in disease pathogenesis, it is essential to identify the molecular genetic factors involved in the etiology and physiology of male infertility.

Deletions in the Azoospermia Factor a, b and c (AZFa, AZFb and AZFc) regions of the long arm of the Y chromosome have been identified as the most common cause of Y-linked male factor infertility, particularly spermatogenic failure [53, 54]. In addition to Y-linked genes, many autosomal and X-linked genes such as CFTR, androgen receptor (AR) genes, estrogen receptors (ESR) genes, ubiquitin-specific protease-26 (USP26) gene, are also

necessary for normal sexual development, testis determination, testis descent, spermatogenesis and eventually fertility in men [48, 55].

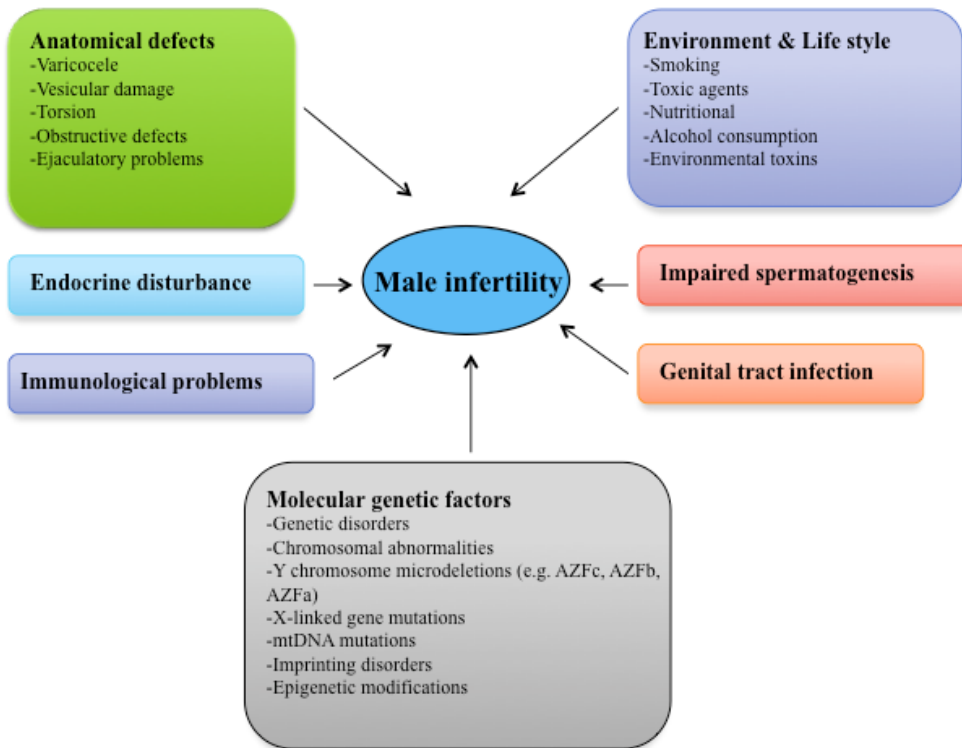


Figure 3. Most common causes of male infertility. Male Infertility is a multifactorial disease, with varied causative factors including genetic, molecular, epigenetic, immunological, environmental and life styles.

The advent of new technologies has spurred research on the role of epigenetics in spermatogenesis and male infertility. Epigenetics refer to process of gene regulation devoid of changes in DNA sequence [56]. Epigenetics is dominantly involved in the preservation of cellular identity and lineage fidelity, through the different mechanisms of alterations in chromatin structure, modifications of DNA and histones, remodeling of nucleosomes etc. Studies have demonstrated that numerous genes can be regulated through epigenetic mechanisms in the testes, indicating a direct influence of epigenetic mechanisms on the process of spermatogenesis, sperm development or maturation, fertilization process as well as male fertility [27]. It has become increasingly clear that epigenetic regulation of gene expression is critical during spermatogenesis and fertilization process [57].

Studies have made it abundantly clear that, in fathers exposed to environmental toxins or pesticides, sperm quality and chromatin integrity can be negatively affected through the primordial germ cells (PGCs), where there are some protected portions of the genome containing repetitive elements and single copy sequences that have the potential to transmit DNA methylation across generations or during differentiation, from PGCs to spermatogonia, de novo methylation occurs at imprinted loci or during development from spermatogonium to

spermatocytes and during reprogramming that happens at fertilization, through the 4-10% of retained histones [58-60].

3.2. Genomic Imprinting

Genomic imprinting, namely, the expression of alleles in a sex-specific manner, is characterized by epigenetic modifications including differential methylation in the CpG islands in either the maternal or paternal alleles [28]. It determines which genes from the parental and maternal genomes are expressed in the embryo and is critical for normal development [61]. Parental imprints are erased in the PGCs of the developing embryo, and re-established during gametogenesis in a sex-dependent manner, maintained through fertilization, pre- and post-implantation embryonic development [62]. Imprint re-establishment occurs at late fetal stages in male germ cells and after birth in growing oocytes [64].

In humans, fetal spermatogonia seem to be mostly unmethylated at *H19* differentially methylated regions, although spermatogonia in adult testis demonstrate significant methylation in this region [28]. Imprinted genes, including the paternally imprinted *GTL2* and *H19* loci, have been previously examined by Kobayashi et al., in 97 infertile men. The importance of genomic imprinting during spermatogenesis has been reinforced by the association of decreased methylation of the paternal *IGF2/H19* imprinting control region 1 (ICR1) and *GTL2* imprints in spermatozoa of men with disturbed spermatogenesis [60]. In another study by Poplinski et al., fertile men had a high degree of *IGF2/H19* ICR1 and a low degree of *MEST* methylation [64]. Low sperm counts were clearly associated with *IGF2/H19* ICR1 hypomethylation and, even stronger, with *MEST* hypermethylation. These results suggest that idiopathic male infertility is strongly associated with imprinting defects at *IGF2/H19* ICR1 and *MEST*, with aberrant *MEST* methylation being a strong indicator for sperm quality. Previous studies have also suggested that Prader-Willi Syndrome and Angelman syndrome are caused by the specific loss of paternally expressed genes [2].

ART techniques such as ICSI and round spermatid injection (ROSI) may increase the incidence of imprinting disorders and adversely affect embryonic development by using immature spermatozoa that may not have established proper imprints or global methylation yet. Andrologists have voiced concern about concealing reproductive defects through ART that might have negative consequences at the epigenetic level [2].

3.3. DNA Methylation

DNA methylation occurs largely on CpG islands found upstream of genes [14]. It is a process in which a methyl group from S-adenosyl methionine (SAM) is transferred to a cytosine residue to create 5-methyl cytosine (5mC). This process is catalyzed by a family of closely related DNA methyl transferases (DNMT1, DNMT3a and DNMT3b) [65].

The establishment and maintenance of correct DNA methylation patterns is essential for fertility, embryo development and viability of the offspring. Abnormal methylation patterns in human spermatozoa have been reported in men with infertility and low sperm counts [66]. Recent studies have demonstrated that CpG methylation occurs in selected promoters of the fertile group, while methylation levels in subfertile patients vary considerably [64]. Sperm

DNA methylation levels are related to conventional sperm parameters, as well as, sperm chromatin and DNA integrity [67]. Recently, hypermethylation of promoters of several genes including *MTHFR*, *PAX8*, *NTF3*, *SFN*, *HRAS*, *JHM2DA*, *IGF2*, *H19*, *RASGRF1*, *GTL2*, *PLAG1*, *D1RAS3*, *MEST*, *KCNQ1*, *LIT1* and *SNRPN* have been reported to be associated with poor semen parameters or male infertility [28]. Improper DNA methylation of various genes has been implicated in abnormal semen parameters, as well as several instances of male infertility. Houshdaran et al., demonstrated for the first time, that poor sperm concentration, motility and morphology were associated with broad DNA hypermethylation across a number of loci including *PLAG1*, *DIRAS3*, *MEST*, *PAX8*, *NTF3*, *SFN* and *HRAS*. They suggested that the underlying mechanism for these epigenetic changes may be improper erasure of DNA methylation during epigenetic reprogramming of the male germ line [68]. In another study by Khazamipour et al., 53% of men with nonobstructive azoospermia revealed hypermethylation of the *MTHFR* promoter, while none of the men with obstructive azoospermia exhibited hypermethylation of this gene promoter, suggesting that *MTHFR* hypermethylation is a specific epigenetic aberration that may specifically contribute to certain types of male infertility [69]. In addition, Wu et al., in a study reported that hypermethylation of *MTHFR* gene promoter in sperm was associated with idiopathic male infertility [70]. Abnormal DNA methylation of *H19* and *MEST* imprinted genes has also been shown to be associated with oligozoospermia, suggesting that spermatogenesis may be particularly vulnerable to changes in the methyl pool brought about by deficiency in *MTHFR* enzyme [71]. Hammoud et al., examined CpG methylation patterns in infertile (oligozoospermic and those with abnormal protamine ratios) and fertile donors at seven imprinted loci including *LIT1*, *MEST*, *SNRPN*, *PLAGL1*, *PEG3*, *H19*, and *IGF2* [72]. At six of the seven imprinted genes, the overall DNA methylation patterns were significantly altered in both infertile patient populations. They demonstrated a link between abnormal spermatogenesis and abnormal methylation of genes. Contradictory reports demonstrated high methylation at the *H19/IGF2* ICR1 locus and low methylation at the *MEST* locus in normozoospermia in one study, while another study revealed low methylation of the *H19/IGF2* ICR1 locus and high methylation of the *MEST* locus in association with low sperm concentrations [64]. Similarly, Boissonnas et al., found that many patients with teratozoospermia and oligoasthenoteratozoospermia exhibited hypomethylation at variable CpG islands at the *H19* locus [73]. However, recent studies have shown that hypermethylation at *MEST* was more strongly linked with poor sperm quality than hypomethylation at *H19/IGF2* ICR1 [28]. They suggested that sperm from infertile patients, especially those with oligospermia, may carry a higher risk of transmitting incorrect primary imprints to their offspring, highlighting the need for more research into epigenetic changes, when considering ART. The X-linked *RHOX* cluster encodes a set of homeobox genes that are selectively expressed in the male and female reproductive tract. Some members of the *Rhox* cluster are crucial for normal spermatogenesis, germ cell survival and male fertility. A recent study has demonstrated that the human *RHOX* gene cluster serves as an excellent marker for idiopathic male infertility. This study revealed a strong association between human *RHOX* cluster hypermethylation and the severity of poor sperm quality [74]. It appears that, DNA methylation directly represses the transcription of human *RHOX* gene family members. Incorrect DNA methylation of the *DAZL* promoter CpG island has been shown to be associated with defective human sperm [66]. *DAZ* gene family is crucial for normal spermatogenesis, as deletion of *DAZ* is estimated for 10% of cases of men with spermatogenic defect [75]. Targeted disruption of *DNMTs* genes in mice resulted in embryonic lethality and infertility [76]. Hammoud et al.,

showed a significant association between male factor infertility and alterations in sperm DNA methylation at imprinted loci [72]. They reported that patients with male factor infertility had significantly increased methylation alteration at LIT1, MEST, SNRPN, PLAGL1, PEG3, H19, and IGF2 genes [72]. In another study, DNA methylation at MEST gene was significantly associated with oligozoospermia, decreased bi-testicular volume and increased FSH levels [79]. Increased abnormal DNA methylation in the ART sample (41%) has been reported previously [80]. The reduced DNA methylation of the H19 ICR has been reported negatively correlated with sperm concentration [81]. A significant decrease in DNA methylation at the H19 DMR (differentially methylated regions) was found in testicular sperm of azoospermic men compared with proven fertile men [16]. In another study by Vieweg et al., sperm DNA methylation of normozoospermic men displayed low standard deviation, while a higher variability in the percentage of DNA methylation was detected in the promoters of subfertile men [68]. These data suggest that abnormalities in DNA methylation level in human sperm not only impact on sperm parameters quality and male infertility, but it could also represent a new approach to investigate the fertilizing ability of sperm in an assisted reproduction procedure, especially when sperm samples with normal characteristics are used [82].

4. EPIGENETICS IN ART (ASSISTED REPRODUCTIVE TECHNIQUES)

Assisted Reproductive technology (ART) includes all artificial methods used to achieve pregnancy [83]. Although a boon to infertile couples, ART-related manipulations to oocytes and embryos, have to be reviewed and designed with caution, as they coincide with the development of sex-specific genomic imprints [84]. The invasive as well as the non-invasive procedures performed during ART, including hormonal stimulation, egg retrieval, intracytoplasmic sperm injection (ICSI), micro-manipulation of gametes, *in vitro* fertilization (IVF), *in vitro* oocyte maturation, exposure to culture medium and centrifugation could potentially harm the early embryos and gametes. The process of ART coincides with crucial developmental phases of the pre-implantation embryo, and is therefore, implicated in inducing epigenetic changes that could affect fertility in subsequent generations and result in the birth of children with a higher risk of infertility, congenital abnormalities and morbidity [49]. It is thus imperative for the clinicians involved in the treatment of these couples to initiate genetic evaluation and counseling prior to any therapeutic procedures.

Until recently, the main evidence that ART like IVF/ICSI causes abnormalities in offspring, was corroborated by low birth weight babies, or premature births [83, 84]. Zhang et al., utilized microarray studies to show that poor perinatal outcomes in ART children arise from changed expression of genes involved in immune response, cell differentiation, compared to control patients [85]. But more recent investigations have shed light into differential expression of DNA methylation-associated genes that have been detected in offspring conceived by ART procedures. The commonly investigated list of imprinted genes for methylation-dependent disruptions include KCNQ1OT1, IGF2, SNRPN, MEST/PEG1, PEG10, PEG3, L3MBTL [87]. Children conceived from infertile couples, through ART have a higher propensity to develop imprinting disorders, since it has been shown that imprinted genes are more susceptible to environmental stresses [87]. A small but statistically significant subset of genes have altered

DNA methylation profiles in children derived through ART, and the authors [87] propose that these significant differences in gene expression could be accounted for, by either the abnormal epigenetic marks borne by infertile couples or from the ART process itself. But, overall, conflicting studies on the effects of ART on offspring have made it impossible to conclusively point to major differences between ART and normal children [88, 89].

In addition to short-term outcomes, long-term outcomes, with regard to ART-conceived children is also important to scrutinize. It is interesting to note that several studies have observed taller ART children, at the ages of 6-10 years, probably due to the pre- or early implantation factors. But, with respect to gonadal development or fertility status in ART-conceived children, there was basically no change between ART offspring and spontaneously conceived children [90, 91, 92].

5. EPIGENETIC MAKE-UP OF INDUCED PLURIPOTENT STEM CELLS/PGC FROM ADULT CELLS

In order to enhance our understanding of male reproduction and its associated diseased condition, male infertility, it is crucial to recapitulate gametogenesis and embryogenesis, *in vitro* [93, 94]. The last decade in reproductive research has seen important strides in this field using induced germ cells. The *in vitro* induction of germ cells from autologous iPS (induced pluripotent cells) and embryonic stem cells will also function as an additional route to conceive, for infertile couples, who cannot produce any gametes of their own [94]. Needless to say, the differentiation of human iPS cells will also enable in the molecular elucidation of diseases in gynecology, pediatrics and obstetrics.

Human IPS cells can be generated from somatic cells by the ectopic expression of OCT4, SOX2, KLF4 and c-MYC [95]. Recently, additional reports have demonstrated that human iPS cells can be induced to differentiate into the PGCs, gonocytes, spermatocytes, and into round spermatid –like cells [94]. It is however, important to note that PGCs derived from iPS cells have shown inefficient imprinting, as well as a propensity to acquire trisomies in chromosomes 12 and 17, possibly hinting to epigenetic instabilities. Additionally, human iPS cells seem to acquire higher number of CNVs, and revealed copy number variants at early passages compared to human ES cells. High-resolution methylation studies have also suggested that some iPS cell lines possess somatic memory, raising concerns about aberrant methylation and imprinting [10, 11, 18].

As a landmark discovery in this field, Saitou et al., accomplished the difficult task of generating PGC-like cells (PGCLC) from mouse ESC/iPSC cells. In order to prove that these PGCLCs could generate germ cells, they placed the PGCLCs into neonatal mice lacking their own germ cells. Very interestingly, mature sperms were produced from the testes, and could even be fertilized to produce their own offspring [95]. Though the PGCLCs could form functional sperm, there were minor differences in PGCLC-derived sperms compared to PGC-derived sperms [97]. But it is necessary to proceed with caution, since it is very likely that, they will be subject to genetic/epigenetic instabilities during iPS cell generation process, since the generation of such germ cells seem to occur against the Weismann barrier (hereditary information moves only from germ cells to somatic cells) [94].

An impairment of epigenetic reprogramming in germ cells can lead to loss of germ cells, faulty embryos or defectively developed offspring from these faulty embryos. In mice, during reprogramming of PGCs, they initiate a transcriptional network for pluripotency by predominantly acquiring epigenetic marks such as histone H3 lysine 9 dimethylation (H3K9me2), histone H3 lysine 27 trimethylation (H3K27me3) and demethylation of 5-methylcytosine, in addition to, silencing somatic programs [97, 98]. As a result, the PGCs acquire low levels of genome-wide 5-methylcytosine.

Recent progress in germ cell induction could lead to a new form of ART, where patients with no viable gametes will have the opportunity to produce fertile spermatozoa from autologous iPS cells, which can subsequently be used for IVF/ICSI. Utilizing spermatogonial stem cells (SSCs) from autologous iPS cells could also initiate in vivo spermatogenesis in otherwise infertile males [99].

CONCLUSION

The latest and most promising development is the creation of the first human artificial primordial germ cells (PGC) from reprogrammed skin cells by stem cell biologist Mitinori Saitou's group [100]. Initial studies performed on these artificial PGC's showed that their epigenetic and protein profiles were very similar to those of primordial germ cells obtained from aborted fetuses. In the next step in these studies, sperm or eggs were generated from these artificial PGC's in mice. It is interesting to envisage the future where these iPSc will be utilized to treat male infertility [101].

A conceivable future for the fertility treatment field will be to obtain sperm cells or eggs cells derived from skin cells of sterile men or women, which could then be a viable option to enable infertile couples to conceive, albeit, with the associated ethical concerns [102]. This technology will greatly benefit patients suffering from idiopathic male infertility, as well as the male cancer survivors. But, in order, to maximally benefit from these exciting techniques, it is imperative to gain a better understanding of the epigenetic modifications that occur during the generation of these iPSc cultures, as well as better investigate the imprints of the sperm chromatin, on a long-term basis in offspring [82].

REFERENCES

- [1] Schagdarsurengin U., Paradowska A. and Steger K. Analysing the sperm epigenome: roles in early embryogenesis and assisted reproduction. *Nature Reviews Urology*. 2012;9:609-19.
- [2] Kota SK and Feil R. Epigenetic transitions in germ cell development and meiosis. *Developmental Cell*. 2010;19:675-86.
- [3] Rousseauz S., Caron C., Govin J., Lestrat C., Faure AK. and Khocbin S. *Establishment of male-specific epigenetic information*. 2005; 345:39-153.
- [4] Hess RA and De Franca LR. Spermatogenesis and cycle of the seminiferous epithelium. *Adv Exp Med Biol*. 2008;636:1-15.

- [5] Lopez JG., Alonso L., Cardenas DB., Alvarez HA., Hourcade JD., Martinez S., Enriquez MAB and Del Mazo J. Diversity and functional convergence of small non coding RNAs in male germ cell differentiation and fertilization. *RNA*. 2015;21:946-62.
- [6] Kumar M., Kumar K., Jain S., Hassan T and Dada R. Novel insights into the genetic and epigenetic paternal contribution to the human embryo. *Clinics*. 2013;68:5-14.
- [7] Shaoqin G, Zhenghui Z, Xueqian Z and Yuan H. Epigenetic modifications in human spermatozoon and its potential role in embryonic development. *Hereditas / Zhongguo yi chuan xue hui bian ji*. 2014;36:439-46.
- [8] Meikar O., Ros MD and Kotaja N. Epigenetic regulation of male germ cell differentiation. *Sub-cellular biochemistry*. 2013; 61:119-138.
- [9] Yuen BT, Bush KM, Barrilleaux BL, Cotterman R and Knoepfler PS. Histone H3.3 regulates dynamic chromatin states during spermatogenesis. *Development*. 2014;141:3483-94.
- [10] Zamudio NM, Chong S and O'Bryan MK. Epigenetic regulation in male germ cells. *Reproduction*. 2008;136:131-46.
- [11] Nazor KL., Altun G., Lynch C., Tran H., Harness JV., Slavin I., Garitaonandia I., Muller FJ., Wang YC. and Boscolo FS. Recurrent variations in DNA methylation in human pluripotent stem cells and their differentiated derivatives. *Cell Stem Cell*. 2012;10:620-34.
- [12] Lister R., Pelizzola M., Kida YS., Hawkins RD., Nery JR., Hon G., Antosiewicz-Bourget J., O' Malley R., Castanon R. and Klugman S. Hotspots of aberrant epigenomic reprogramming in human induced pluripotent stem cells. *Nature*. 2011;471:68-73.
- [13] Oliva R. Protamines and male infertility. *Human reproduction update*. 2006;12:417-3.
- [14] Benchaib M., Ajina M., Braun V., Niveleau A and Guerin JF. DNA sperm methylation in assisted reproductive techniques. *Gynecologie, obstetrique & fertilite*. 2006;34:836-9.
- [15] Cisneros FJ. DNA methylation and male infertility. *Frontiers in bioscience: a journal and virtual library*. 2004;9:1189-200.
- [16] Minor A., Chow V. and Ma S. Aberrant DNA methylation at imprinted genes in testicular sperm retrieved from men with obstructive azoospermia and undergoing vasectomy reversal. *Reproduction*. 2011;141:749-57.
- [17] Miller D., Brinkworth M, and Iles D. Paternal DNA packaging in spermatozoa: more than the sum of its parts? DNA, histones, protamines and epigenetics. *Reproduction*. 2010;139:287-301.
- [18] Ohi Y., Qin H., Hong C., Blounin L., Polo JM., Guo T., Qi Z., Downey SL., Manos PD. and Rossi DJ. Incomplete DNA methylation underlies a transcriptional memory of somatic cells in human iPS cells. *Nat. Cell Biol*. 2011;13: 541-49.
- [19] Shirakata Y., Hiradate Y., Inoue H., Sato E. and Tanemura K. Histone H4 modification during mouse spermatogenesis. *The Journal of reproduction and development*. 2014;60:383-7.
- [20] Boissonnas CC., Jouannet P. and Jammes H. Epigenetic disorders and male subfertility. *Fertility and sterility*. 2013;99:624-31.
- [21] Castillo J., Estanyol JM., Ballesca JL. and Oliva R. Human sperm chromatin epigenetic potential: genomics, proteomics, and male infertility. *Asian journal of andrology*. 2015;17:601-9.

- [22] Lesch BJ and Page DC. Poised chromatin in the mammalian germ line. *Development*. 2014; 141:3619-26.
- [23] Vavouri T. and Lehner B. Chromatin organization in sperm may be the major functional consequence of base composition variation in the human genome. *PLoS Genet*. 2011;7:e1002036.
- [24] Drewell RA., Bush EC., Remnant EJ., Wong GT., Beeler SM., Stringham JL., Lim J. and Oldroyd BP. The dynamic DNA methylation cycle from egg to sperm in the honey bee *Apis mellifera*. *Development*. 2014;141:2702-11.
- [25] Erkek S., Hisano M, Liang CY., Gill M., Murr R., Dierker J., Schubeler D., van der Vlag J., Stadler MB, and Peters AHFM. Molecular determinants of nucleosome retention at CpG-rich sequences in mouse spermatozoa. *Nat. Struct. Mol. Biol*. 2013; 20: 868-75.
- [26] Yuen BT., Bush KM., Barrilleaux BL., Cotterman R. and Knoepfler PS. Histone H3.3 regulates dynamic chromatin states during spermatogenesis. *Development*. 2014;141:3483-94.
- [27] Rousseaux S., Reynoird N., Escoffier E., Thevenon J., Caron C and Khochbin S. Epigenetic reprogramming of the male genome during gametogenesis and in the zygote. *Reproductive biomedicine online*. 2008;16:492-503.
- [28] Rajender S., Avery K. and Agarwal A. Epigenetics, spermatogenesis and male infertility. *Mutation research*. 2011;727:62-71.
- [29] Oliva R and Ballesca JL. Altered histone retention and epigenetic modifications in the sperm of infertile men. *Asian journal of andrology*. 2012;14:239-40.
- [30] Gaucher J., Reynoird N., Montellier E., Boussouar, Rousseaux S and Khochbin S. From meiosis to postmeiotic events: The secrets of histone disappearance. *FEBS J*. 2010;277:599-604.
- [31] Vastenhouw NL and Scheir AF. Bivalent histone modifications in early embryogenesis. *Curr Opin Cell Biol*. 2012;24:374-386.
- [32] Pan XY., Li ZX., Wang ZC., Wang XN., Huang BY., Dou ZH and Sun YM. Histone modifications during spermatogenesis and male infertility. *Zhongguo yi xue ke xue yuan xue bao Acta Academiae Medicinae Sinicae*. 2014;36:108-13.
- [33] Kleiman SE., Bar-Shira Maymon B., Hauser R., Botchan A., Paz G., Yavetz H. and Yogev L. Histone H4 acetylation and AZFc involvement in germ cells of specimens of impaired spermatogenesis. *Fertility and sterility*. 2008;89:1728-36.
- [34] Vieweg M., Dvorakova-Hortova K., Dudkova B., Waliszewski P., Otte M., Oels B., Hajimohammad A., Turley H., Schorsch M., Schuppe HC., Weidner W., Steger K. and Paradowska-Dogan A. Methylation analysis of histone H4K12ac-associated promoters in sperm of healthy donors and subfertile patients. *Clinical epigenetics*. 2015;7:31.
- [35] Zamudio NM., Chong S and O'Bryan MK. Epigenetic regulation in male germ cells. *Reproduction*. 2008;136:131-46.
- [36] Oliva R., Bazett-Jones D., Mezquita C and Dixon GH. Factors affecting nucleosome disassembly by protamines in vitro. Histone hyperacetylation and chromatin structure, time dependence, and the size of the sperm nuclear proteins. *The Journal of biological chemistry*. 1987;262:17016-25.
- [37] Oliva R and Dixon GH. Vertebrate protamine genes and the histone-to-protamine replacement reaction. *Progress in nucleic acid research and molecular biology*. 1991;40:25-94.

- [38] Wu JY and Means AR. Ca(2+)/calmodulin-dependent protein kinase IV is expressed in spermatids and targeted to chromatin and the nuclear matrix. *The Journal of biological chemistry*. 2000;275:7994-9.
- [39] Oliva R and Mezquita C. Histone H4 hyperacetylation and rapid turnover of its acetyl groups in transcriptionally inactive rooster testis spermatids. *Nucleic acids research*. 1982;10:8049-59.
- [40] Torregrosa N., Dominguez-Fandos D., Camejo MI., Shirley CR., Meistrich ML., Ballesca JL. and Olivia R. Protamine 2 precursors, protamine 1/protamine 2 ratio, DNA integrity and other sperm parameters in infertile patients. *Human reproduction*. 2006;21:2084-9.
- [41] Aoki VW., Liu L and Carrell DT. Identification and evaluation of a novel sperm protamine abnormality in a population of infertile males. *Human reproduction*. 2005;20:1298-306.
- [42] Faure AK., Pivot-Pajot C., Kerjean A., Hazzouri M., Pelletier R., Peoc'h M., Sele B., Khochbin S. and Rousseaux S. Misregulation of histone acetylation in Sertoli cell-only syndrome and testicular cancer. *Molecular human reproduction*. 2003;9:757-63.
- [43] Sonnack V., Failing K., Bergmann M and Steger K. Expression of hyperacetylated histone H4 during normal and impaired human spermatogenesis. *Andrologia*. 2002;34:384-90.
- [44] Dada R., Kumar M., Jesudasan R., Fernandez JL., Gosalvez J and Agarwal A. Epigenetics and its role in male infertility. *Journal of assisted reproduction and genetics*. 2012;29:213-23.
- [45] Steilmann C., Cavalcanti MC., Bartkuhn M., Pons-Kuhnemann J., Schuppe HC., Weidner W., Steger K and Paradowska A. The interaction of modified histones with the bromodomain testis-specific (BRDT) gene and its mRNA level in sperm of fertile donors and subfertile men. *Reproduction*. 2010;140:435-43.
- [46] Siomi MC., Sato K., Pezic D and Aravin AA. PIWI-interacting small RNAs: the vanguard of genome defence. *Nature Reviews Molecular Cell Biology*. 2011;12: 246-58.
- [47] Carmell MA., Girard A., van de Kant HJ., Bourc'his D., Bestor TH., de Rooji DG. and Hannon GJ. MIWI2 is essential for spermatogenesis and repression of transposons in the mouse male germ line. *Dev. Cell*. 2007;4: 503-14.
- [48] Tahmasbpour E., Balasubramanian D and Agarwal A. A multi-faceted approach to understanding male infertility: gene mutations, molecular defects and assisted reproductive techniques (ART). *Journal of assisted reproduction and genetics*. 2014;31:1115-37.
- [49] Agarwal A., Mulgund A., Hamada A and Chyatte MR. A unique view on male infertility around the globe. *Reproductive biology and endocrinology. RB&E*. 2015;13:37.
- [50] Ko EY., Sabanegh ES, Jr. and Agarwal A. Male infertility testing: reactive oxygen species and antioxidant capacity. *Fertility and sterility*. 2014;102:1518-27.
- [51] Benoff S., Jacob A and Hurley IR. Male infertility and environmental exposure to lead and cadmium. *Human reproduction update*. 2000;6:107-21.
- [52] Colagar AH., Jorsaraee GA and Marzony ET. Cigarette smoking and the risk of male infertility. *Pakistan journal of biological sciences. PJBs*. 2007;10:3870-4.
- [53] Zhang YS., Dai RL., Wang RX., Zhang ZH., Fadlalla E and Liu RZ. Azoospermia factor microdeletions: occurrence in infertile men with azoospermia and severe oligozoospermia from China. *Andrologia*. 2014;46(5):535-40.

- [54] Ambulkar PS., Sigh R., Reddy M., Varma PS., Gupta DO., Shende MR. And Asoke KP. Genetic Risk of Azoospermia Factor (AZF) Microdeletions in Idiopathic Cases of Azoospermia and Oligozoospermia in Central Indian Population. *Journal of clinical and diagnostic research: JCDR*. 2014;8:88-91.
- [55] Lo Giacco D., Chianese C., Ars E., Ruiz-Castane E., Forti G and Krausz C. Recurrent X chromosome-linked deletions: discovery of new genetic factors in male infertility. *Journal of medical genetics*. 2014;51:340-4.
- [56] Shaoqin G., Zhenghui Z., Xueqian Z and Yuan H. Epigenetic modifications in human spermatozoon and its potential role in embryonic development. *Hereditas / Zhongguo yi chuan xue hui bian ji*. 2014;36:439-46.
- [57] Xu YN., Cui XS., Tae JC., Jin YX. and Kim NH. DNA synthesis and epigenetic modification during mouse oocyte fertilization by human or hamster sperm injection. *Journal of assisted reproduction and genetics*. 2011;28:325-33.
- [58] Miranda CL., Gomez PR., Rojas G., Cruz I., Berrueta L., Salmen S., Colmenares M., Barreto S., Baiza A., Zavala L., Morales Y., Molina Y., Valeri L., Contreras CA and Osuna JA. Occupational exposure to organophosphate and carbamate pesticides affects sperm chromatin integrity and reproductive hormone levels among Venezuelan farm workers. *J Occup. Health*. 2013; 55: 195-203.
- [59] Martenies SE and Perry MJ. Environmental and occupational pesticide exposure and human sperm parameters: a systematic review. *Toxicology*. 2013; 307:66-73.
- [60] Soubry A., Hoyo C., Jirtle RL. and Murphy SK. A paternal environmental legacy: Evidence for epigenetic inheritance through the male germ line. *Bioessays*. 2014;36:359-371.
- [61] Kobayashi H., Sato A., Otsu E., Hiura H., Tomatsu C., Utsunomiya T., Sasaki H., Yaegashi N. and Arima T. Aberrant DNA methylation of imprinted loci in sperm from oligospermic patients. *Human molecular genetics*. 2007;16:2542-51.
- [62] Reik W and Walter J. Genomic imprinting :parental influence on the genome. *Nat Rev Gnet*. 2001;2:21-32.
- [63] Delaval K and Feil R. Epigenetic regulation of mammalian genomic imprinting. *Curr Opin Genet Dev*. 2004;14:188-95.
- [64] Poplinski A., Tuttelmann F., Kanber D., Horsthemke B. and Gromoll J. Idiopathic male infertility is strongly associated with aberrant methylation of MEST and IGF2/H19 ICR1. *International journal of andrology*. 2010;33:642-9.
- [65] Cisneros FJ. DNA methylation and male infertility. *Frontiers in bioscience: a journal and virtual library*. 2004;9:1189-200.
- [66] Navarro-Costa P., Nogueira P., Carvalho M., Leal F., Cordeiro I., Calhaz-Jorge C. goncalves J. and Plancha CE. Incorrect DNA methylation of the DAZL promoter CpG island associates with defective human sperm. *Human reproduction*. 2010;25:2647-54.
- [67] Montjean D., Zini A., Ravel C., Belloc S., Dalleac A., Copin H., Boyer P., McElreavey K and Benkhalifa M. Sperm global DNA methylation level: association with semen parameters and genome integrity. *Andrology*. 2015;3:235-40.
- [68] Houshdaran S., Cortessis VK., Siegmund K., Yang A., Laird PW. and Sokol RZ. Widespread epigenetic abnormalities suggest a broad DNA methylation erasure defect in abnormal human sperm. *PloS One*. 2007;2:e1289.
- [69] Khazamipour N., Noruzinia M., Fatehmanesh P., Keyhanee M. and Pujol P. MTHFR promoter hypermethylation in testicular biopsies of patients with non-obstructive

- azoospermia: the role of epigenetics in male infertility. *Human reproduction*. 2009;24:2361-4.
- [70] Wu W., Shen O., Qin Y., Niu X., Lu C., Xia Y., Song L., Wang S and Wang X. Idiopathic male infertility is strongly associated with aberrant promoter methylation of methylenetetrahydrofolate reductase (MTHFR). *PloS one*. 2010;5:e13884.
- [71] Kumar R., Venkatesh S., Kumar M., Tanwar M., Shasmsi MB., Kumar R., Gupta NP. Sharma RK., Talwar P and Dada R. Oxidative stress and sperm mitochondrial DNA mutation in idiopathic oligoasthenozoospermic men. *Indian journal of biochemistry & biophysics*. 2009;46:172-7.
- [72] Hammoud SS., Purwar J., Pflueger C., Cairns BR and Carrell DT. Alterations in sperm DNA methylation patterns at imprinted loci in two classes of infertility. *Fertility and sterility*. 2010;94:1728-33.
- [73] Boissonnas CC., Abdalaoui HE., Haelewyn V., Fauque P., Dupont JM., Gut I., Vaiman D., Jouannet P., Tost J and Jammes H. Specific epigenetic alterations of IGF2-H19 locus in spermatozoa from infertile men. *European journal of human genetics: EJHG*. 2010;18:73-80.
- [74] Richardson ME., Bleiziffer A., Tuttelmann F., Gromoll J. and Wilkinson MF. Epigenetic regulation of the RHOX homeobox gene cluster and its association with human male infertility. *Human molecular genetics*. 2014;23:12-23.
- [75] Kuo PL., Wang ST., Lin YM., Lin YH., Teng YN and Hsu CC. Expression profiles of the DAZ gene family in human testis with and without spermatogenic failure. *Fertility and sterility*. 2004;81:1034-40.
- [76] Webster KE., O'Bryan MK., Fletcher S., Crewther PE., Aapola U., Craig J., Harrison DK., Aung H., Phutikanit N., Lyle R., Meachem SJ., Antonarakis SE., de Krestser DM., Hedger MP., Peterson P., Carroll BJ and Scott HS. Meiotic and epigenetic defects in Dnmt3L-knockout mouse spermatogenesis. *Proceedings of the National Academy of Sciences of the United States of America*. 2005;102:4068-73.
- [77] Klaver R., Tuttelmann F., Bleiziffer A., Haaf T., Kliesch S. and Gromoll J. DNA methylation in spermatozoa as a prospective marker in andrology. *Andrology*. 2013;1:731-40.
- [78] Kobayashi H., Hiura H., John RM., Sato A., Otsu E., Kobayashi N. Suzuki R., Suzuki F., Hayashi C., Utsunomiya T., Yaegashi N., Arima T. DNA methylation errors at imprinted loci after assisted conception originate in the parental sperm. *European journal of human genetics : EJHG*. 2009;17:1582-91.
- [79] Li JB., Li B., Liang XX., Wang J., Ma YF., Zhang YQ., Liu Z., Min Bh., Ma XH and Wang XH. Correlation of DNA methylation status of imprinted gene H19 ICR with oligozoospermia and asthenozoospermia *National journal of andrology*. 2013;19:511-7.
- [80] Benchaib M., Braun V., Ressnikof D., Lornage J., Durand P., Niveleau A. and Guerin JF. Influence of global sperm DNA methylation on IVF results. *Human reproduction*. 2005;20:768-73.
- [81] Nagano M., Patrizio P. and Brinster RL. Long-term survival of human spermatogonial stem cells in mouse testes. *Fertil. Steril*. 2002; 78:1225-33.
- [82] Easley CA., Phillips BT., McGuire MM., Barringer JM., Valli H., Hermann BP., Simerly CR., Rajkovic A., Miki T., Orwig KE. and Schatten GP. Direct differentiation of human pluripotent stem cells into haploid spermatogenic cells. *Cell*. 2012;2: 440-6.

- [83] Henningsen AK., Pinborg A., Lidegaard O., Vestergaard C., Forman JL. and Nyboe AA. Perinatal outcome of singleton siblings born after assisted reproductive technology and spontaneous conception: Danish national sibling-cohort study. *Fertil Steril.* 2011; 95: 959-63.
- [84] Lu YH., Wang N and Jin F. Long-term follow-up of children conceived through assisted reproductive technology. *J. Zhejiang Univ Sci. B.* 2013;14:359-71.
- [85] Zhang Y., Cui Y., Zhou Z., Sha J., Li Y. and Liu J. Altered global expressions of human placenta subjected to assisted reproductive technology treatments. *Placenta.* 2010; 31:251-258.
- [86] Uyar A and Seli E. The impact of assisted reproductive technologies on genomic imprinting and imprinting disorders. *Curr. Opin. Obstet. Gynecol.* 2014; 26:210-21.
- [87] Katari S., Turan N., Bibikova M., Erinie O., Chalian R., Foster M., Gaughan JP, Coutifaris C. and Sapienza C. DNA methylation and gene expression differences in children conceived in vitro or in vivo. *Hum Mol Genet.* 2009;18: 3769-78.
- [88] Tierling S., Souren NY., Gries J., Lo Porto C., Groth M., Lutsik P., Neitzel H., Utz-Billing I., Gillessen-Kaesbach G., Kentenich H., Griesinger G., Sperling K., Schwinger E. and Walter J. Assisted reproductive technologies do not enhance the variability of DNA methylation imprints in human. *J Med Genet* 2010;47:371-76.
- [89] Rancourt RC., Harris HR. and Michels KB. Methylation levels at imprinting control regions are not altered with ovulation inductions or in vitro fertilization in a birth cohort. *Hum Reprod.* 2012; 27:2208-16.
- [90] Belva F., Henriët S., Liebaers I., van Steirteghem A., Celestin-Weistreich S. and Bonduelle M. Medical outcome of 8-year old singleton ICSI children (born ≥ 32 weeks gestation) and a spontaneously conceived comparison group. *Hum Reprod.* 2007;22:506-15.
- [91] Belva F., Roelants M., Painter R., Bonduelle M., Devroey P and de Schepper J. Pubertal development in ICSI children. *Human Reprod.* 2012;27:1156-61.
- [92] de Schepper J., Belva F., Schiettecatte J., Anckaert E., Tournaye H. and Bonduelle M. Testicular growth and tubular function in prepubertal boys conceived by intracytoplasmic sperm injection. *Horm Res.* 2009;71:359-63.
- [93] Ishii T. Human iPS cell-derived germ cells: Current status and clinical potential. *J. CLin. Med.* 2014; 3: 1064-83.
- [94] Hayashi K., Ohta H., Kurimoto K., Aramaki S. and Saitou M. Reconstitution of the mouse germ cell specification pathway in culture by pluripotent stem cells. *Cell.* 2011; 146:519-32.
- [95] Takahashi K., Tanabe K., Ohnuki M., Narita M., Ichisaka T., Tomoda K. and Yamanaka S. Induction of pluripotent stem cells from adult human fibroblasts by defined factors. *Cell.* 2007; 131:861-72.
- [96] Reik W and Walter J. Genomic imprinting: Parental influence on the genome. *Nat Rev Genet.* 2001;2:21-32.
- [97] Seisenberger S., Andrews S., Krueger F., Arand J., Walter J., Santos F., Popp C., Thienpont B., Dean W. and Reik W. The dynamics of genome-wide DNA methylation reprogramming in mouse primordial germ cells. *Mol. Cell.* 2012. 48:849-62.
- [98] Seki Y., Hayashi K., Itoh K., Mizugaki M., Saitou M and Matsui Y. Extensive and orderly reprogramming of genome-wide chromatin modifications associated with

specification and early development of germ cells in mice. *Develop. Biol.* 2005. 278:440-58.

- [99] Nagano M., Patrizio P. and Brinster RL. Long-term survival of human spermatogonial stem cells in mouse testes. *Fertil. Steril.* 2002; 78:1225-33.
- [100] Saitou M., Kagiwada S. and Kurimoto K. Epigenetic reprogramming in mouse pre-implantation development and primordial germ cells. *Development.* 2012. 139:15-31.
- [101] Gafni O., Weinberger L., Mansour AA., Manor YS., Chomsky E., Ben-Yosef D., Kalma Y., Viukov S., Maza I., Zviran A., Rais Y., Shipony Z., Mukamel Z., Krupalnik V., Zerbib M., Geula S., Caspi I., Schneir D., Schwartz T., Gilad S., Amann-Zalcenstein D., Benjamin S., Amit I., Tanay A., Massarwa R., Novershtem N. and Hanna JH. Derivation of novel human ground state naïve pluripotent stem cells. *Nature.* 2013; 504:282-6.

Chapter 65

EPIGENETICS, INFLAMMATION AND INFLAMMATORY RELATED-DISEASES: A GENERAL LOOK

Müge Sayitoğlu* and Yücel Erbilgin

Istanbul University, Institute of Experimental Medicine,
Department of Genetics, Istanbul, Turkey

ABSTRACT

Epigenetic is defined as the study of mitotically and/or meiotically heritable changes in the gene function without changing DNA sequence and playing crucial roles both in normal development and human diseases. The molecular basis of epigenetic process consists of histone modifications, DNA methylation, positioning of histone variants, and non-coding RNAs. Genome-wide patterns of DNA and chromatin modifications together called as 'epigenomes'. Epigenomes undergo precise, coordinated, reversible changes through the developmental stages and so contributes to the lineage and tissue specific expression of genes. In addition to the tissue specific impact, environmental factors such as nutrients, toxins, infections and hypoxia can also influence epigenomes. Distinct or global changes in the epigenetic landscape are hallmarks of chronic inflammation associated diseases. Alteration in methylation status of CpG sites, monoallelic silencing, and other epigenetic regulatory mechanisms have been observed in key inflammatory response genes. Epigenetic changes including DNA methylation, histone modification and noncoding RNA expression, were found associated with acute and chronic inflammatory disorders. Recently, therapies targeting epigenetic mechanisms are trending options for the treatment of chronic and degenerative disorders. Epigenetic drugs that are applied on animal models and some clinical trials are displaying positive therapeutic effects. Not only mono therapies but also combined usage of HDAC or DNMT inhibitors could be the next step for the epigenetic therapeutic modulations. Scope of this chapter is to provide an overview of the epigenetic modifications in inflammation and inflammation driven diseases.

* Corresponding Author's Email: mugeay@istanbul.edu.tr.

ABBREVIATIONS

DNA	deoxyribonucleic acid
RNA	ribonucleic acid
CpG	cytosine-guanine nucleotides
DNMT	DNA methyltransferases
SLE	systemic lupus erythematosus
RA	rheumatoid arthritis
MS	multiple sclerosis
HMTs	histone methyltransferases
HDMs	demethylases
HATs	histone acetyltransferases
HDACs	histone deacetylases
LncRNAs	Long noncoding RNAs
miRNA	microRNA
T1D	type 1 diabetes
CVD	cardiovascular diseases
IBD	inflammatory bowel disease
CD	Crohn's disease
UC	ulcerative colitis
ACPA	against citrullinated peptide antigens
RASFs	RA synovial fibroblasts
PBMCs	peripheral blood mononuclear cells
5-azaC	5-azacytidine cytosine
IFN	type 1 interferon
PADI2	peptidyl arginine deiminase type 2
MBP	myelin basic protein
ROS	reactive oxygen species
HIV	human immunodeficiency virus
EBV	Epstein Barr virus
SIRS	systemic inflammatory response syndrome
SSI	severe systemic inflammation
RNS	reactive nitrogen species
UVB	ultraviolet B
JAK	Janus-activated kinase
MAP	mitogen-activated protein
DMV	DNA methylation valleys

1. INTRODUCTION

Inflammation (Latin, *inflammatio*) is a part of the complex biological protective response of the body that involves immune cells, blood vessels, and molecular mediators. Inflammation develops acute or chronic response of immune system and it is modulated by genetic, epigenetic, macrobiotic and environmental events. Although many variations exist among

inflammatory diseases, a consistency occurs in local and systemic acute inflammation (like abscess and sepsis) and chronic inflammation like metabolic syndrome of obesity, cancer, atherosclerosis, autoimmunity, and aging.

One of the earliest reports which is suggested by a Roman physician named Cornelius Celsus, therein described patient symptoms against to tissue injury were *rubor* (redness), *tumor* (swelling, caused by increased permeability of the microvasculature), *calor* (heat), and *dolor* (pain). After centuries, scientists have begun to understand the physiological and cellular basis of inflammation. The inflammatory cells are not only important for host defense against pathogens; but also they play beneficial roles in tissue maintenance and homeostasis through the process of phagocytosis. Acute inflammation occurs over seconds, minutes, hours, and days, and chronic inflammation occurs over longer times. Inflammation is mainly described by redness, heat, pain and swelling, against pathogens, damaged cells, irritant and/or tissue repair/healing. Inflammatory response can be vascular or cellular. During vascular response histamine mediated hyperemia occurs, blood vessels dilate, fluid moves into tissue platelets forms a clot to trap the injurious agent. In the cellular response, white blood cells move to the site of the injury (called chemotaxis), neutrophils, monocytes/ macrophages, eosinophils, basophils and lymphocytes involve in the inflammation process [1, 2]. Inflammation has a slow onset, persists for a long period of time, and becomes chronic. The symptoms in chronic inflammation are not as severe as in acute inflammation, but the condition is persistent. Chronic inflammation is associated with a group of multifactorial complex disorders such as cancer, diabetes, neurological, cardiovascular and pulmonary disorders that will be discussed further.

Stages of inflammation can change in hours so the process needs a strict gene expression regulation. During the basal state, genes are repressed until a diverse receptor system, like toll-like receptors (TLRs), senses a threat. Whenever a threat is noticeable, cells rapidly transcribe pro-inflammatory genes to induce acute inflammation. If the threat is transient, genes return to the basal state within hours. If it is severe, the incitement phase is replaced within 4-6 hours by a gene-specific epigenetic reprogram that may endure days to weeks. Duration may extent up to 21 days in human. This sustained epigenetic regulation period produces variable inflammatory clinical phenotypes and predicts outcome [2].

Nuclear factor kappa B (NF- κ B) pathway is the first outstanding molecular mechanism, when we focus on to the molecular bases of inflammation. Canonical NF- κ B pathway has complex roles in innate immunity, inflammation and oncogenesis, which is conserved in all multicellular animals [3]. As a part of the immune defense, NF- κ B activation leads to target and eliminate transformed cells [4]. Microbial infections, TLR activation and proinflammatory cytokines such as interleukine (IL)-1 and Tumor Necrosis Factor-alpha (TNF- α) activate NF- κ B pathway. This activation induces the expression of inflammatory cytokines, adhesion molecules, key enzymes of prostaglandin synthase pathway (COX-2), nitric oxide (NO) synthase, angiogenic factors and antiapoptotic genes [5]. Termination of the transcriptional activity of NF- κ B is mainly achieved by NF- κ B, up regulating its own inhibitors such as I κ B α , A20 and CYLD [6, 7]. I κ B α enters the nucleus and removes NF- κ B from the DNA and sends it to the cytosol. In acute inflammation, this self-regulation usually ends with complete deactivation of NF- κ B. However, in chronic inflammatory states, the persistent presence of NF- κ B activating stimuli seems to outperform the inhibitory feedback circuits, rooting to an increased constitutive activity of NF- κ B. Depending on the accessibility of the genome regulated by epigenetic and genomic alterations, different transcription factors can be expressed

after eNF- κ B activation. The crosstalk between different pathways leads development of cancers and other diseases [1, 3].

The importance of the interconnection between germ line DNA and epigenetic modifications in inflammation is still emerging. The inflammatory regulatory response requires complex regulatory networks at genetic and epigenetic levels. Thousands of inflammation and innate immunity genes contribute directly or indirectly in acute or chronic inflammatory processes. Transcription factors of the NF- κ B, FOXP3, IRF, STAT families including epigenetic phenomena, play critical roles in the regulation of inflammatory genes. The proposed epigenetic mechanisms regulate acute and chronic inflammation in different ways. The epigenetic silencing state does not exist during chronic inflammation and the reason for this paradox is unknown, but it has important therapeutic implications. Therefore, we investigate the effect of epigenetic modifications in inflammation and related diseases as below.

1.1. DNA Methylation and Inflammation

DNA methylation, a dynamic process involving methylation and demethylation events, occurs in different regions of the genome and is extremely important for embryogenesis, cellular proliferation and differentiation [8]. In eukaryotes, DNA methylation only occurs at cytosine residues and catalyzed by the DNA methyltransferase (DNMT1, DNMT3a, and DNMT3b) enzymes [9]. In DNA methylation, cytosine is converted to methyl-cytosine at CpG site and about 60% of human genes have CpG islands at their promoters. The changes in DNA methylation, such as hypomethylation mostly associated with chromosome instability and activation of transposable elements, beside that hypermethylation of promotor associated CpGs often linked with suppression of gene expression and cancers [10]. DNA demethylation is necessary for the epigenetic reprogramming of somatic nuclei and active demethylation is associated with the activation of immune cells [11, 12].

DNA methylation is the most widely studied mechanism in autoimmune diseases. A range of disease associated with chronic inflammation, including systemic lupus erythematosus (SLE) and rheumatoid arthritis (RA), display global hypomethylation associated with a concurrent decreased expression of methylation-related genes. Moreover hypomethylation patterns were also observed at interferon-regulated genes, such as *IRF5*, *IFIT2*, *STAT1* and *USP18*, in CD4⁺ T cells in the context of SLE. However, not all epigenetic alterations in lymphocytes are subject to hypomethylation. It has been reported that regulatory regions of *FOXP3* in peripheral blood CD4⁺ T cells from patients with SSc, RA and T1DM is hypermethylated, affecting the expression of key transcription factor which is required for the generation of regulatory T cells [13, 14].

Alterations in DNA methylation are known to cooperate with genetic events and to be involved in various diseases [15]. Therefore, understanding the roles of DNA methylation is important for understanding disease progression. Applications of next generation techniques have led to accumulate more data on DNA methylation in both normal and disease cells. Methylation is a reversible epigenetic mark and highly informative to described gene regulation in normal and diseased cells, and it can potentially function as a biomarker [16].

1.2. Histone Modifications and Inflammation

Histones are highly alkaline proteins that package and order the DNA into structural units named nucleosomes. Nucleosomes consist of about 146 bp of DNA wrapped 1.65 turns around an octameric histone core, comprised of 2 copies each of histones H2A, H2B, H3 and H4. Nucleosomes are stabilized by linker histones H1 and H5 [17]. The core histones have a globular domain and protruding N-terminal tails, both of which are subjected to post-translational modifications [18, 19].

There are three well studied mechanisms that modulate nucleosome structure and chromatin dependent cellular processes; core histone post-translational modifications, chromatin remodeling and composition of nucleosomes. Core histone post-translational modifications, including acetylation, ubiquitination, phosphorylation, methylation, sumoylation and ribosylation affect histone interactions, change dynamics of nucleosome and subsequently regulate gene expression. In general, acetylation of lysine in the promoter regions of histone triggers the gene expression on the other hand; histone deacetylation, biotinylation, and sumoylation inhibit the gene expression. Methylation and ubiquitination mechanisms can be activator or repressor of gene expression depending on target histone residues [20, 21]. DNA methylation and histone methylation are tightly controlled entwined events. For instance, DNA methylation requires the histone H3 N-terminal tail with an unmethylated lysine 4 (H3K4). Tri-methylation of histones H3 on lysine 9 (H3K9me3) participates in DNA methylation mediated by DNMT1 [22].

Epigenetic modifications in chromatin's structure affect gene expression such as loosely packed form of DNA which is called euchromatin, shows higher transcriptional activity in contrast, tightly packed form of DNA, heterochromatin presents lower transcriptional activity [23]. Various chromatin remodeling complexes can disrupt and remodel the nucleosome structure by increasing DNA accessibility [24]. Two groups of chromatin-modifying complexes have been described; ATP-dependent complex, covalent histone modifications. The balance of histone modifications is established and maintained by a range of enzymes including, but not limited to histone methyltransferases (HMTs) and demethylases (HDMs), histone acetyltransferases (HATs) and deacetylases (HDACs). Histone modifications are important regulators of gene transcription and modulate dynamic processes that affect nucleosomes. Long noncoding RNAs have also been shown to be necessary for targeting histone-modifying activities [25].

Aberrant histone modifications contribute to human diseases such as cancer, neurodegenerative diseases and infections. Including inflammation (IL-1 β , TNF- α , LPS ect.) many stimulants can promote histone acetylation. Histone acetyl transferases (HATs) activate inflammatory genes, whereas histone deacetylase (HDACs) activity represses the inflammatory gene expression. Promoters of pro-inflammatory cytokines (such as IL-1, IL-2, IL-8, and IL-12) are rapidly acetylated and become transcriptionally active. HDACs regulate transcription of both pro-inflammatory and anti-inflammatory cytokines via their co-repressor complexes and transcription factors such as *FOXP3*, *STATs*, *GATAs*, *ZEB1*, and *NF- κ B* [22].

In addition to nuclear function, histones also act like an endogenous signal when they locate at extra nuclear space. As a response to stress, immune cells, cerebellar neurons, Schwann cells and microglia present H2A, H2B, H3, H4, and H1 on their cell surface or cytoplasm. Levels of circulating histones as well as nucleosomes are increased in cancer, inflammation, and

infection, which suggest that histones are therapeutic targets for infectious and inflammatory disorders [26, 27].

1.3. Non-Coding RNAs and Inflammation

1.3.1. Long Non-Coding RNAs

Long noncoding RNAs (lncRNAs) vary widely in size, genomic localization, and biogenesis and can be classified according to their molecular mechanism of action. lncRNAs function through a variety of different molecular mechanisms, and most commonly play a scaffolding role to promote proper recruitment and positioning of protein regulators both in the nucleus and in the cytoplasm. Loss-of-function approaches have found that lncRNAs regulate the biology of both innate and adaptive immune cells during inflammatory responses. Some nuclear lncRNAs that contribute to the regulation of DNA methylation in the context of inflammation have been identified. Some lncRNAs implicated in histone modification, particularly methylation, that play role in processes such as cell cycle, inflammation, and senescence [28].

Some inflammation related lncRNAs and their functions are summarized below;

NeST is involved in inflammation during microbial infection, binds to the H3K4 methyltransferase component WDR5 (WD repeat domain 5) and recruits it to the *IFN γ* locus. *IFN γ* methylation increases with advancing age, *NeST* could contribute to the inflammatory response and infection in elderly.

17A is up regulated in cerebral tissues derived from Alzheimer disease patients as well as in response to inflammatory stimuli such as *IL-1 α* . *17A* is encoded within the *G-protein-coupled receptor 51 (GPR51)* gene. *17A* provides an interesting link between GPRs and age-associated neurodegeneration.

Lethe, in mouse embryonic fibroblasts (MEFs), lncRNAs are found to be regulated following treatment with TNF- α , a pro-inflammatory cytokine, which is associated with inflammation, aging, age-related diseases, and cellular senescence. Among these, *Lethe* was found to be particularly important for the pro-inflammatory state of aging tissues by providing a key negative feedback loop: binding of *Lethe* to the NF- κ B subunit RelA reduce the production of inflammatory proteins.

THRIL, TNF- α was also found induced by lncRNA *THRIL* (TNF- α - and hnRNPL-related immunoregulatory lincRNA). This lncRNA interacts with the TNF- α promoter and regulates its expression in *THP1* macrophages. Together, lncRNAs *lethe* and *THRIL* are involved in inflammation in a TNF- α -dependent manner.

Lnc-IL7R, was identified as a regulator of the lipopolysaccharide (LPS)-induced inflammatory response. Depletion of *Lnc-IL7R* reduces trimethylation of histone H3 at lysine 27 (H3K27me3) causing a reduction in the levels of inflammatory mediators including E-selectin, VCAM-1, IL-6 and IL-8.

The TGF- β /Smad3 pathway is involved in the inflammatory response. It has been identified that Smad3 associated lncRNAs are related to renal inflammation. The analysis identified several lncRNAs altered in Smad3 knockout mice. As this pathway is impaired with age and Smad3 regulates senescence phenotype, this network of proteins and lncRNAs differentially expressed may play a role in aging related inflammation ('inflammaging').

1.3.2. Micro RNAs

MicroRNAs are small, single-stranded ncRNAs and function to repress gene expression and influence virtually all organ systems in vertebrates. As an epigenetic regulatory mechanism, miRNAs, enable increased human complexity despite a genome size that is similar to less complex organisms. This appears to include critical roles in establishing proper inflammatory set points and facilitating optimal responses and resolution by our immune system.

Recent studies demonstrated that aberrantly expressed miRNAs exert a significant role in inflammatory genes. RNA interference mechanism culminates with the up or down regulation of hundreds of immune response genes, an essential first step in coordinating inflammatory responses. miRNAs act as a modulator of inflammation signals or as mediators between inflammation and cancer. The experimental data indicated that miRNAs have been shown to be expressed in immune cells, to target proteins involved in inflammation regulation, and consequently to affect the severity of the response. There are two major ways for the regulations: To impacting the development of inflammatory cell subsets (e.g., Th2 versus Th17) or to establish the level of immune cell function (e.g., controlling how much cytokine is made by DCs).

Certain fundamental attributes of miRNAs make them ideally suited to regulate chronic inflammatory conditions. MicroRNAs modulate immune cell differentiation and responses. Inflammation related miRNAs play important roles in various inflammatory processes. The opposing effects of miRNAs on inflammatory responses indicate that, the immune system utilizes multiple miRNAs to properly balance its functional capacity, creating a tension between activation and repression that can be precisely tuned [29]. It is still unknown how multiple miRNAs are in cooperation to balance inflammatory responses, and how miRNA networks can work together to optimize responses.

miRNAs have been implicated in regulating macrophage lineage skewing during inflammatory responses. Macrophages can be skewed toward either pro-inflammatory subtypes (M1), or less inflammatory subtypes (M2). *Mir-125* has been shown to repress M1 skewing while promoting the M2 fate. *Mir-223* has also been implicated in control of granulocyte activation, and *mir-223*^{-/-} mice display overactive immune responses and develop inflammatory lung pathology. Age dependent inflammatory phenotypes are being explored among miRNAs. It has been discovered that mice lacking *mir-146a* develop an age-dependent, chronic inflammatory disease that is spontaneous, life shortening, associated with inflammatory cytokines and autoantibodies, and that involves a variety of hematopoietic abnormalities and malignancies associated with the aging [30].

Details of epigenetic mechanisms, which play driver roles in inflammation and related diseases, will be discussed as below.

2. EPIGENETICS AND INFLAMMATORY DISORDERS

Epigenetic modifications in inflammatory autoimmune diseases have been recognized and attracted recently. Although there are many components (such as genomic variations and environmental factors), which lead predisposition of autoimmunity and related phenotypes, there are several reported epigenetic alterations as well.

The inflammatory response is a double-edged sword; immune responses are necessary for immune defense, wound repair and tissue homeostasis, if these responses become deregulated, they initiate a chronic reaction leading to chronic status. This condition, referred as *chronic low-grade inflammation*, and adversely contributes to many diseases such as aging, obesity, type 1 diabetes (T1D), rheumatoid arthritis (RA), systemic lupus erythematosus (SLE), neurodegeneration, and cardiovascular diseases (CVD) [13].

Epigenetic alterations contribute immune response and those alterations related autoimmune diseases are summarized in Table 1. In the last decade, epigenetic studies provide new insights into autoimmune diseases. The identification of specific epigenetic markers may lead to lighten uncharacterized mechanisms. Identification of clinical significance of the epigenetic markers may be utilized for diagnostic biomarkers and therapeutic benefits for the autoimmune disorders. In this section we will give some examples to the epigenetic regulations, malfunctions and markers of known inflammatory disorders;

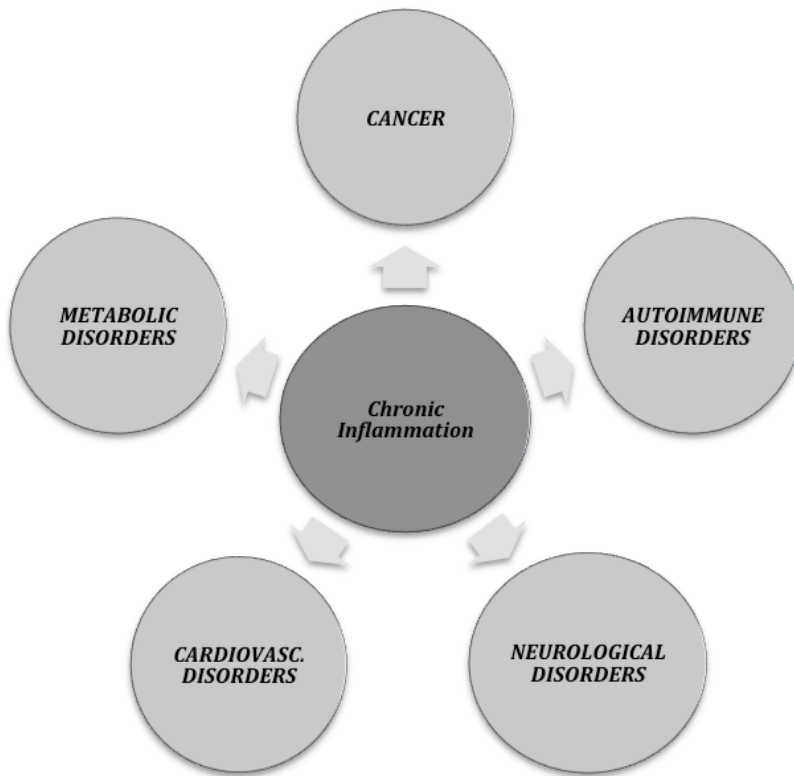


Figure 1. Inflammation mediated diseases.

2.1. Inflammatory Bowel Disease (IBD)

Inflammatory bowel disease (IBD) is a group of inflammatory conditions of the colon and small intestine. IBD is classified as an autoimmune disease, and Crohn's disease (CD) and ulcerative colitis (UC) is the main types of the IBD. Deregulated immunologic response causes IBD and subsequently long lasting active IBD evokes to development of colorectal carcinoma in many cases. So far the data supports the notion that epigenetic factors can act as fine-tuners

of gene expression in diseases such as in IBD. Particular combinations of epigenetic changes could result in a pathogenic phenotype through activating genes that promote chronic inflammation or inhibiting anti-inflammatory gene expression in IBD. Regarding the interaction between environmental, genetic and epigenetic mechanisms seem to have great importance in the IBD. Intestinal micro biota are likely the most important environmental factor in IBD as targets of the inflammatory response [31]. It is a challenge to measure the impact of the environmental factors that potentially can contribute to IBD; thus far epigenetic factors can mediate gene-environment interactions involved in pathogenesis.

Intestinal bacteria can regulate the epithelial gene expression and the immune response by epigenetic mechanisms. Bacterial and host epigenetics can also affect host genetics to trigger inflammatory processes. A number of studies have similarly reported the bacteria induce histone modifications (see in part 2.6), thereby inflammatory response modulation and other cellular processes of the epithelium in the gastrointestinal tract [32].

The first evidence of DNA methylation, as an epigenetic mechanism in UC was reported in 1996 and found that incorporation of the 3^H -methyl groups into DNA was 10-fold higher in UC patients than controls and significantly higher in histologically active than inactive disease. A progressive increase in methylation status was found associated with neoplastic progression in UC [33]. Another reported promoter methylation on the *E-CADHERIN* gene, which is coding a specific calcium ion-dependent cell adhesion protein, was detected in 93% of dysplastic samples from UC. The result was confirmed by *E-cadherin* synthesis levels, found reduced in samples with dysplasia and normal in samples without dysplasia, suggesting that long standing inflammation is related to hypermethylation of the gene promoter and that DNA methylation may be used as a biomarker for detecting high risk patients for developing colorectal cancer [34].

Both targeted gene methylation and methylome array based studies have detected numerous colonic mucosal associates of inflammation in IBD [35, 36]. Several studies confirmed that *ER*, *MYOD*, *GSPG2* and *p16* genes were found hyper methylated in dysplastic than normal epithelium. Severe disease phenotypes of UC was found related to *MDR1* hypermethylation [37]. A tumor suppressor gene death-associated protein kinase (*DAPK*) hypermethylation was also found to be associated to the inflammation of mucosa in UC patients and a correlation is observed between higher methylation and severe inflammation [38]. DNA methylation has been found to be related to many different clinical aspects, such as disease severity, disease duration, disease phenotype, disease extent, steroid use, steroid dependence or refractoriness, age of onset, number of hospitalizations and finally active inflammation and dysplasia [39, 40].

There is limited data on concerning the contribution of DNA methylation status in CD pathogenesis. Lin et al. compared normal to inflamed tissue from CD and UC patients and found significant differences in DNA methylation of several CpG loci such as *SERPINA5*, *TNFRSF1A*, *AATK*, *GABRA5*, *MAPK10* and *STAT5A* [35]. One of the first genome wide methylation studies analyzed the methylation profiles of peripheral blood samples from women and children with CD. Several genes identified differentially methylated in patients with IBD and controls. Ontology analysis highlighted several pathways associated with IBD, including immune system processes, immune response, and host response to bacteria, whereas canonical pathway analysis indicated the involvement of Th17 cell pathways [41].

Histone modifications are complex processes and less extensively studied in IBD than DNA methylation. The patterns of histone acetylation in the colon of rats with colitis and humans with CD have been described and increased acetylation of H4 (at lysine residues 8 and 12) was determined in the inflamed mucosa and in Peyer's patches [42]. A gene-specific pattern of histone modification is observed for collagen type I, the most abundant component of the fibrotic extracellular matrix, can be induced by fibrosis relevant cytokines (IL1b, transforming growth factor beta, and tumor necrosis factor alpha) suggesting that epigenetic factors regulate fibrotic gene transcription relevant to IBD with a fibrostenosing phenotype [32]. Several anti-inflammatory mechanisms of HDACi have been proposed, one of these is T-regulatory cells, which mediate immune tolerance and repress the excessive inflammation, is linked to FOXP3 gene expression. Administration of HDACi to increased T-regulatory cell differentiation and suppression of bowel inflammation, as a result of acetylation of lysines in the Foxp3 [43]. Another mechanism could involve their effects on acetylation of proteins in the NF- κ B pathway [44].

Studies in animal models have shown that intestinal miRNAs regulate gut homeostasis [45]. Studies reported miRNA expression profiles in the peripheral blood and gut biopsy specimens of healthy controls versus adult/pediatric patients with active and inactive IBD. The miRNA expression profiles found deregulated at both the tissue level and in peripheral blood of UC and CD patients. Moreover IBD-related glucocorticoid therapy can modify the expression profile of different miRNAs [46, 47]. Specific miRNA profiles and identification of their targets might provide additional insight into IBD pathogenesis and help to predict IBD susceptibility, progression, relapse, and response to therapy. The miRNAs associated with IBD could also serve also as fine tuners of gene expression. The levels of specific mRNAs shown to be tightly controlled by miRNAs, alterations of the threshold mRNA levels results in marked changes in protein synthesis. Deregulation of the intestinal inflammatory response that results from the disruption in the balance between miRNA activity and its specific target mRNAs, e.g., genes with important functions in intestinal homeostasis [48].

Changes in miRNAs in human IBD were first described in in sigmoid biopsy specimens from patients with active UC. *MiR-192*, which is expressed in colonic epithelial cells, was found significantly reduced in tissues of patients with active UC. *MiR-192* levels were found increased in colon tissues of patients with UC leading to the expression of *macrophage inhibitory peptide 2a*, a CXC chemokine expressed in the epithelial cells [49]. *MiR-150* is up regulated in mice with dextran sulfate sodium induced colitis in colon tissues from patients with UC; its levels correlate inversely with those of its target c-Myb, which has a role in apoptosis [50]. Another miRNA, *miR-21*, which promotes inflammation, has been reported in several studies of patients with active UC and CD colitis, along with *miR-155* [51]. *MiR-196* is overexpressed in the inflamed epithelium of patients with CD [52]. Several micro RNAs have been found to be significantly up-regulated or down-regulated in 2 or more studies, including *miRs-16*, *miR-28*, *miR-149*, *miR-151*, *miR-199*, and *miR-534* in UC and CD [53]. There are several other miRNAs found related to intestinal fibrosis; *miR-29*, *miR-200b*, *miR-21*, *miR-192* or associated with epithelial barrier and immune function in IBD pathogenesis; *miR-192*, *miR-21*, *miR-126*, *miR-155*, *miR-106a* [48].

The importance of epigenetic factors in IBD derives from the mechanisms of action of some therapeutics used for the treatment of IBD. There is increasing evidence that glucocorticoids change the chromatin structure of genes that code mediators of inflammation.

Glucocorticoids can increase transcription of anti-inflammatory genes which results from the uncoiling of DNA wound around histone by acetylation of the histone residues. Glucocorticoids may also lead to deacetylation of histones on the genes of pro-inflammatory cytokines, resulting in tighter coiling of DNA and reduced access of transcription factors to their binding sites, and suppressing the target gene expression [54].

2.2. Rheumatoid Arthritis (RA)

Another chronic auto inflammatory disease, rheumatoid arthritis (RA), is characterized by synovial hyperplasia, chronic inflammation and progressive destruction of the synovial joints. RA has an unknown etiology and affecting about 1% of the population [55]. There are two subsets of RA, according to the existence of the antibody against citrullinated peptide antigens (ACPA). The presence or absence of ACPA is one of the best clinical predictors for disease outcome. The most important genetic risk factors, accounting for up to 50% of the overall risk for RA, are mainly confined to the human leukocyte antigen locus [56]. All kinds of epigenetic alterations occur in RA including DNA methylation, histone acetylation, and miRNAs has contribution [57] and continuation of the disease [58, 59].

The RA synovial fibroblasts (RASFs) play a major role in the initiation and continuation of the disease [58, 59]. These cells display an activated, aggressive and invasive phenotype [60, 61]. A global hypomethylation of the genes on these cells could be responsible for the overexpression of inflammatory cytokines in synovial fluid leading to disease phenotype [62] and these epigenetic biomarkers might provide the missing link between RA, risk factors and a lack of therapy response. Global DNA hypomethylation has also been observed in T cells and peripheral blood mononuclear cells (PBMCs) from RA patients [63].

L1 is one of the major classes of repetitive elements in the genome. Since they are methylated in the normal synovial tissue they used as biomarkers. In synovial tissue from patients with RA, L1 is hypomethylated as a consequence of reduced expression of *DNMTs*. This reduction of methylation in inflammatory response promoter genes causes an overexpression of growth factors and receptors, adhesion molecules, and cytokines, and cause irreversible phenotypic changes in synovial fibroblasts. The other hypomethylated gene, *IL-6*, is a proinflammatory cytokine that participates in B cell response. When it is hypomethylated, an overexpression of *IL-6* will occur and trigger the overexpression of pro-inflammatory cytokines at the same time and this is associated with a local hyper activation of the inflammation circuit [64, 65]. Another hypomethylated gene, *IL-1* is associated with elevated *IL-10* expression in RA and by regulating *IL-10* transcription it can be responsible for the pathogenesis of RA. Not only global but also single gene hypomethylations, such as *chemokine ligand 12 (CXCL-12)* in RASFs and (*Interleukin 1 receptor, type II*) *IL1R2*, has also been reported in RA [66, 67]. Genome-wide evaluation revealed many hypomethylated loci in RASFs including *CHI3L1*, *CASP1*, *STAT3*, *MAP3K5*, *MEFV*, and *WISP3* [12].

Besides hypomethylation, CpG hypermethylations in the regulatory regions of some genes lead to aberrant transcription in RA. Hypermethylation of *EBF3* and *IRX1* genes interacts with transforming growth factor beta (TGF- β) pathway components and reduce mRNA expression in RASFs [68]. The promoter region of the *death receptor 3 gene (DR3)*, a member of the apoptosis inducing FAS family, was shown to have significant hypermethylation of CpG

islands in RA patients, which may enhance resistance to apoptosis in RA synovial fibroblasts [57, 69].

The gender bias (3:1/F:M) of RA suggests an X-chromosome role in the development of RA. An increased presence of a skewed X-chromosome inactivation pattern has been shown in the peripheral blood of a RA patient. Instead of a random X-chromosome inactivation, 80% or more of the cells exhibited inactivation of the same X-chromosome. CD40 ligand (CD40L) is primarily expressed at the surface of activated CD4⁺ T helper cells and plays a crucial role in the immune response. A promoter hypomethylation and increased expression of CD40L has been detected in CD4⁺ T helper cells of female RA patients. Hypomethylation of genes on the inactive X chromosome may contribute to female-biased RA patients [70, 71]. The exposure to *IL-1* decreases the mRNA levels of *DNMT1* and *DNMT3a* in RASFs and also inhibits global methylation of RA fibroblast-like synoviocytes (FLS), suggesting the contribution to differential DNA methylation in RASFs through altered expression of *DNMTs*.

RA synovial tissue is characterized by an imbalance between HAT and HDAC activity. Cartilage destruction that is mediated by matrix metalloproteinases (MMPs) and enzymes from the ADAMTS family, are regulated by chromatin modifications and histone acetylations. HDAC inhibitors inhibit cartilage degradation by blocking the induction of certain MMPs by proinflammatory cytokines. Hyper acetylation of synovial cell histones, induce cyclin dependent kinase inhibitors (p16 and p21) expression via decreased *tumor necrosis factor-alpha* (*TNF-α*) expression (72, 73). Additionally the hyper acetylation of histones down regulate HIF-1α (hypoxia inducible factor) and vascular endothelial growth factor (VEGF) to block angiogenesis in synovial cells [74].

DNA methylation has been shown to regulate the microRNAs (miRNAs) expressions in RA. Demethylation with 5-azacytidine (5-azaC) increases the expression of *miR-203* in RA [75]. The increased levels of *miR-203* result in elevated secretion of MMP-1 and IL-6 via the NF-κB pathway, and thereby, contribute to the activated phenotype of RASFs. MeCP2 is a chromatin-binding protein that preferentially binds to methylated CpGs and is highly abundant in the synovium of RA rats. It has also been shown that increasing global methylation levels are regulated by overexpression of *MeCP2* in a RA rat model [76]. The promoter of *miR-124a* hypermethylation implicated in the proliferation of FLS [77]. The recent genome-wide methylation studies have reported hypomethylation and hypermethylation of multiple genes in tissue-derived FLS from patients with RA [78-82].

Micro RNAs have also been implicated in RA [63]. Primary human samples have increased expression of *miR-155*, *miR-146*, and *miR-223* and mouse arthritis models have demonstrated the functional relevance of *miR-155* and *miR-182* for regulating B and T cell function during disease [83-86]. *MiR-146* and *miR-155*, are shown to be induced by proinflammatory stimuli such as *IL-1*, *TNFα*, and *Toll-like receptors (TLRs)*. Another miRNA implicated in RA is *miR-124*, which targets *cyclin-dependent kinase 2 (CDK-2)*. In the basal state, *CDK2* represses cell proliferation and arrests the cell cycle at the G1 phase, but in pathologic conditions such as RA, its level decreases. *MiR-124* also targets *monocyte chemo attractant protein 1 (MCP-1)*, which is responsible for mononuclear phagocytes into the joint. Thus this miRNA increases cell proliferation and MCP-1 production in RA [87]. *MiR-203* is overexpressed in RASF, enhancing the secretion of MMP-1 and interleukin-6 [75]. Tumor necrosis factor-α induces *microRNA-18a* and activates RASF through a feedback loop in NF-κB signaling [88].

2.3. Systemic Lupus Erythematosus (SLE)

Systemic Lupus Erythematosus (SLE) is a systemic autoimmune disease with multiple organ involvement, which is characterized by autoantibody response to nuclear or cytoplasmic antigens. It is defined with the production of increased amount of autoantibodies, which potentially drive immune complex related inflammation in various tissues and organs [89]. Several genetic and epigenetic factors contribute to the development of SLE. Epigenetic disruption results in overexpression of certain genes of the key immune cells, such as T cells, has been associated with the pathophysiology of SLE. Self-antigens resulting in inflammation mediated multiple organ damage, displays global H3 and H4 hypoacetylation in the CD4⁺ T cells of patients [90, 91].

Like RA, SLE is more common in females with peak disease prevalence between menarche and menopause. Estrogen increases the risk of SLE in women via increasing the type 1 interferon (IFN) production and survival of auto reactive B-lymphocytes [92, 93]. Epigenetic modifications might explain the female prevalence for SLE. Impaired DNA methylation presented on the inactive X-chromosomes of SLE patients was suggested to contribute to the female predominance of SLE [94].

The roles of DNA methylation in SLE first described in the 1960s and since then it become a very hot research topic [95]. Several studies have shown that there is a global hypomethylation of promoter regions, which contain the genes that are overexpressed in the SLE like *ITGAL*, *CD40LG*, *PRF1*, *CD70*, *IFGMR2*, *MMP14*, *LCN2*, and in the ribosomal RNA gene promoter (18S and 28S) [94, 96, 97]. The DNA hypomethylation may also affect the chromatin structure of T-cells and leads the overexpression of these genes. The genes overexpression will cause cell hyperactivity and perpetuation of inflammatory response [98]. Genome-wide methylation studies have provided new insights into the role of DNA methylation in the pathogenesis of SLE. The regulatory regions for interferon-responsive genes in naive CD4⁺ T cells from patients with SLE are hypomethylated and ‘poised’ for expression; transcription does not occur until T-cell activation [99].

Histone modifications in SLE have been studied in animal models and in human diseases. These studies showed that, during apoptosis, histones could be modified to make them immunogenic. The antibodies are directed against components of the cell nucleus that are exposed at the cell surface during apoptosis [100, 101]. The nucleosomes, the primary pioneer antigen in SLE, are released in patients with SLE as a result of a corrupted apoptosis or an insufficient clearance of apoptotic residue. During apoptosis, the nucleosome is modified and creates more immunogenic epitopes leading to the formation of autoantibodies against unmodified chromatin components [102]. Histone modifications such as histone 3 lysine 4 trimethylation (H3K4me3), histone 3 lysine 8 (H4K8ac) triacetylation, histone 3 lysine 27 trimethylation (H3K27me3), and histone 2B lysine 12 acetylation (H2BK12ac) cause an increase in apoptotic nucleosomes. These apoptotic nucleosomes generate auto-immunogenicity that can cause activation of antigen presenting cells and autoantibody production with a subsequent inflammatory response [103]. There are other studies that have shown acetylation patterns of histone H3 and H4 in active SLE CD4⁺ T cells [104]. Monocytes, which are important in SLE renal disease, have been shown to have an altered acetylation pattern of histone H4 thus increasing the expression of interferon (IFN) genes that play a key role in SLE pathogenesis [89, 105].

Epigenetic reprogramming has potential preventive and therapeutic purposes for inflammation and autoimmunity. Two methylation inhibitors, procainamide and hydralazine, were found to induce a lupus-like disease after long-term administration in wild-type mice, and the disease disappeared after the withdrawal of these drugs. DNA demethylation has been found in SLE CD4⁺ T cells, but not in CD8⁺ T cells or peripheral blood mononuclear cells (PBMCs) [106]. The recent researches have focused on specific cell types as well as certain SLE relevant genes. Among the other inhibitors procainamide, is a competitive inhibitor of DNMT1 and hydralazine inhibits T and B cell signal regulated kinase pathways that play an important role in the regulation of methylation [107, 108]. These two therapeutic mechanisms produce a reduction in DNMTs that will enhance the genetic expression of adhesion molecules on lupus-drug-induced lymphocytes [109].

Several studies suggest a role for miRNAs in the pathogenesis of SLE. MicroRNAs control the differentiation and function of B cells, which are considered key elements in the pathogenesis of systemic lupus erythematosus (SLE) [110]. For example the under expression of miR-1246 through the AKT-P53 signaling pathway, leads to the expression of *EBF1*, and promote the activation of B cells [111].

Peripheral blood mononuclear cells from patients with SLE showed differential expressions in some miRNAs (increased *miR-189*, *miR-61*, *miR-78*, *miR-21*, *miR-142-3p*, *miR-342*, *miR-299-3p*, *miR-198*, and *miR-298* and decreased *miR-196a*, *miR-17-5p*, *miR-409-3p*, *miR-141*, *miR-383*, *miR-112*, and *miR-184*) [112]. The expression of *miR-21* which targets an important autoimmune gene *RASGRP1*, is increased in CD4⁺ T cell of SLE patients and mediates the RAS–MAPK pathway, down regulating *DNMT1* expression [113]. Similarly, increased levels of *miR-21* found contributed to the B cell hyper responsiveness and aberrant T cell response in SLE by targeting the *PTEN* and *PDCD4* respectively [114-116]. Besides *miR-21*, *miR-126* was also reported to modulate DNA methylation in SLE CD4⁺ T-cells by directly targeting DNMT1 [106].

MiR-146a is a well-recognized negative regulator in autoimmunity. *MiR-146a* was found under expressed in SLE [117] and this was possibly due to a germline genetic variant in the *miR-146a* promoter [118]. This genetic variant (rs57095329) in the promoter region of *miR-146a* confers association with its expression level. The SLE risk-associated G allele of the variant is linked to a reduced *miR-146a* expression in PBMCs, possibly by affecting the protein binding affinity and activity of the promoter. A reverse correlation of *miR-146a* levels with the expression of interferon inducible genes and SLE disease activity was reported, indicating a critical role of *miR-146a* in excessive type I interferon production and signaling activity in SLE [117]. *MiR-146a* has been shown to negatively modulate type-I IFN production in macrophages through targeting TLR signaling genes *TRAF6*, *IRAK1* and *IRAK2* [119]. *MiR-146a* is often bundled with TLR7/9 stimulation, which plays a central role in SLE pathogenesis. Apart from PBMCs, TLR7/9 stimulation in plasmacytoid dendritic cells (pDCs) also induces *miR-146a* expression [120].

SLE patients have multiple clinical features caused by inflammation. *MiR-23b* was found under expressed in the affected tissues of patients with SLE. It has been reported that *miR-23b* targets *TAB2*, *TAB3* and *IKK- α* to inhibit *IL17*, *TNF* or *IL1 β* signaling [121]. Many miRNAs contribute to the reduced IL-2 production in lupus patients. A significant reduction in *miR-31* expression in SLE T cells is positively correlated with the lowered IL-2 production [122].

2.4. Multiple Sclerosis (MS)

Multiple Sclerosis (MS) is another disease characterized by myelin destruction followed by a progressive degree of neurodegeneration. Inflammation is an important mechanism of CNS damage in MS, and proper transcriptional control of the immune responses is mediated in part by epigenetic regulation.

Peptidyl arginine deiminase type II (PADI2) has been found hypomethylated in MS patients [123]. *PADI2* plays a key role in the citrullination process of myelin basic protein (MBP) that promotes protein auto cleavage and increases the probability of creating new epitopes and also modulates the immune response. In MS, an increase has been found in demethylase enzyme activity, which will cause hypomethylation of the *PADI2* promoter region [123, 124]. Hypomethylation increases *PADI2* expression and MBP citrullination process via the production of immunodominant peptides, which will increase the autocleavage of MBP, causing irreversible changes in its biological properties, and will produce proteolytic digestion, myelin instability, and a chronic inflammation response [125]. *SHP-1* gene is the negative regulator of pro-inflammatory signals. Increased promoter methylation in leukocytes in more than one-third of the MS subjects leading to decreased *SHP-1* expression and increased leukocyte-mediated inflammation [126]. CD4⁺ T-cells of relapsing-remitting multiple sclerosis patients showed *FOXP3* and *IL-17A* demethylation. *FOXP3* demethylation inhibits Th1 and Th2 cell differentiation and promotes Th17 cell lineage commitment, whereas the *IL-17A* leads to increased differentiation towards the Th17 cell lineage. Epigenetic control of the Th1/Th2/Th17 balance through to determine disease status [127, 128]. Genome-wide methylome analysis confirmed a global DNA hypermethylation of CD8⁺ T cells for MS patients [129].

The oligodendrocyte identity is modulated by posttranslational modifications of histones. In patients with MS, there is a shift toward histone acetylation in the white matter. Hyperacetylation of H3 in the promoter region of inhibitory genes produce high levels of transcriptional inhibitors of oligodendrocyte differentiation such as *TCF7L2*, *ID2*, and *SOX2* [130]. H3 hyperacetylation is accompanied by an increase in number and activation state of microglia and astrocytes. While the expression of histone deacetylases showed *HDAC8* and *HDAC11* up-regulation, only a subgroup of patients showed overexpression of a number of genes including *TCF7L2*, *SOX2* and *IDS* [131].

Many studies have focused on miRNA involvement in MS pathogenesis. Among those, the *miR-326* plays a critical role in the pathogenesis of MS via upregulation of the Th-17 cell differentiation by targeting Ets-1, which acts as a negative regulator of Th-17 differentiation. *MiR-326* was found significantly upregulated in patients with relapsing-remitting MS, which produced more severe symptoms [132]. Other miRNAs involved in MS are *miR-34a* and *miR-155*, are also upregulated in active MS lesions and contribute to MS pathogenesis by targeting CD47, which is a “don’t eat me” signal Macrophages with low levels of CD47 are released via inhibitory control signals, leads to increased phagocytosis of myelin. *miR-155* also promotes development of inflammatory Th1 and Th17 cells [133].

Differentially expressed *miR-17-5p*, *miR-497*, *miR-193*, and *miR-126* have been identified in different lymphocyte subsets including CD4⁺ T cells, CD8⁺ T cells, B cells, and CD4⁺ CD25⁺ Treg cells from patients with MS. Direct involvement and contribution of deregulated miRNAs in MS still needs additional investigation [133, 134]. It is noteworthy that all miRNAs are involved in the pathogenesis of the disease. There are miRNAs that can serve as prognostic

markers such as the expression of *miR-18b* and *miR-599* is related to relapse and *miR-96* is involved in remission [135]. *MiR-124*, which is expressed in microglia, is brain specific and it might reduce the activation of myelin-specific T cells with a marked suppression of the disease, which would make it a key regulator of microglia quiescence and a good prognostic factor for MS [136].

2.5. Diabetes

The environmental factors and nutrition play an important role in the pathogenesis of diabetes. However, the knowledge on molecular mechanisms by which these factors trigger β -cell dysfunction and diabetes are still limited. Recent discoveries give an impression that epigenetic changes in response to environmental stimuli may play an important role in the development of diabetes. Chronic exposure to high concentrations of glucose, leading to oxidative stress and inflammation, may induce changes in the regulation of gene expression on insulin secretion and increased apoptosis. The combination of inflammation, mitochondrial dysfunction and ROS production provides an attractive model for the pathogenesis of diabetes and may explain the widespread alterations of epigenetic regulation of gene expression during the glycemic control [137, 138].

Type 1 Diabetes (T1D) is a T-cell-mediated autoimmune disease that develops in genetically susceptible individuals and affects their endocrine pancreas. There are some mechanisms by which epigenetics may play an important role in T1D by modulating lymphocyte maturation, cytokine gene expression and by differentiation of subtype Th cells ruled by epigenetic controls. In this autoimmune disease, in contrast to SLE and RA, there is a global hypermethylation activity caused by altered metabolism of homocysteine [139]. Glucose and insulin levels are determinants of methylation, they alter homocysteine metabolism by increasing cell homocysteine production [140, 141]. When there is an increase in the levels of homocysteine, methionine in cells will be catalyzed by DNMTs in S-adenosylmethionin, which will enhance DNMT activity, and lead to increased global DNA methylation. Elevation of maternal homocysteine during pregnancy as a result of a low protein diet can produce an altered methionine metabolism that will cause a decrease in islet mass and vascularity in the fetus with a subsequent glucose intolerance in adult life [142]. The reduced *Foxp3* expression was found to be associated with *Foxp3* promoter hypermethylation in adult autoimmune diabetes patients [143].

There are limited studies associated with histone modifications and T1D. Patients with T1D show a subset of genes with an increase in histone 3 lysine 9 dimethylation (H3K9me2) in lymphocytes. This subset of genes includes the *CLTA4*, which is a T1D susceptibility gene and has increased methylation of H3K9 in its promoter region. Other genes that have altered H3K9me2 are *transforming growth factor-beta (TGF- β)*, *NF- κ B*, *p38 (mitogen-activated protein kinase)*, *toll-like receptors (TLRs)*, and *IL-6*. The transcription factor *NF- κ B* is also upregulated by H3K4 methyltransferase thus causing an increase in inflammatory gene expression in diabetic mice. All these genes are associated with autoimmune and inflammation related pathways [144, 145]. Histone modifications are also among the mechanisms that cause cardiovascular complications in T1D patients. Chemical modification of H3K4 and H3K9 has been found to be related to the gene expression conferred by hyperglycemia. Transient hyperglycemia promotes gene activating epigenetic changes and signaling events critical in the

development and progression of vascular complications. These epigenetic changes are H3K4 and H3K9 methylation in genes associated with vascular inflammation [146, 147].

There are some hypotheses that the function of regulatory T cells (Tregs) is influenced by changes in the expression of specific miRNAs in T1D. *MiR-510* expression was found to be increased and *miR-342* and *miR-191* expressions were found to be decreased in Tregs of diabetic patients. There are other studies, which demonstrate that miRNAs may be the cause of cytokine mediated beta-cell cytotoxicity. This cytotoxicity is achieved when IL-1B and TNF- α induce the expression of *miR-21*, *miR-34a*, and *miR-146a* in pancreatic islets thus producing beta-cell failure by increasing proinflammatory cytokines [148]. *MiR-101a* and *miR-30b* are key players and contribute to cytokine-mediated β -cell dysfunction in the development and progression of type 1 diabetes. IL-1 β induces *miR-101a* and *miR-30b* expression and participates in β -cell dysfunction, including decreased insulin content, gene expression, and increased β -cell death. *MiR-101a* and *miR-30b* reduce pro-insulin expression and insulin content by directly targeting the transcriptional factor Neurod1. In addition, these miRNAs associated with diminished expression level of the antiapoptotic protein Bcl2 [149]. A significant increase of *miR-21* has been observed in the plasma and urine of pediatric T1D. In a type 1 diabetes mouse model, high glucose induced *miR-21* de-repress TORC1-pathway and leads to renal pathologies [150]. In a type 2 diabetes model *miR-21* down regulates phosphorylated SMAD-7, leading to up-regulation of TGF- β and nuclear factor- β signaling pathways, increasing renal fibrosis and inflammation [151].

2.6. Bacterial Infections

Recent studies revealed that bacteria could affect the chromatin structure and transcriptional program of host cells by influencing diverse epigenetic events (i.e., histone modifications, DNA methylation, chromatin-associated complexes, noncoding RNAs, and RNA splicing factors). Bacterial-induced epigenetic deregulations may affect host cell function either to promote host defense or to allow pathogen persistence. Their effects might generate specific, long-lasting imprints on host cells, leading to a memory of infection that influences immunity. The selective activation or silencing of specific genes not only depends on transcription factors, but also on their cross talk with epigenetic modulators, which regulate DNA accessibility by controlling the chromatin structure.

The importance of DNA methylation events associated with bacterial infections is attracting attention. The best documented example is *H. pylori* infection leading aberrant DNA methylation in the human gastric mucosa, strikingly at promoters of genes found to be methylated in gastric cancer cells [152]. *H. pylori* associated hypermethylation is seen at the *CDH1*, *USF1*, *USF2*, *WWOX*, *MLH1* genes [153]. *H. pylori* mediated inflammation triggered lymphocyte and macrophage infiltration, which appears to have a key role in induction of methylation [154]. Among signals resulting from chronic inflammation, elevated levels of interleukine 1b (IL1b) and nitric oxide (NO) are proposed to contribute to influence the recruitment of DNMTs.

In the intestine, following chronic inflammation, some *PRC2* target genes are subject to aberrant DNA methylation [155]. Bacteria-induced DNA methylation can also affect genes involved in cell proliferation. In human uroepithelial cells, infection with uropathogenic *Escherichia coli* (UPEC) results in the up regulation of DNA methyltransferase (DNMT)

activity and DNMT1 expression, and induces CpG methylation and down-regulation of *CDKN2A*, a G1-cell-cycle inhibitor regulator [156]. This may increase uroepithelial cell proliferation and pathogen persistence, by infection stimulated host cell apoptosis.

Other organs can also be targeted by bacteria-mediated epigenetic changes, including placenta. Maternal infection with *Campylobacter rectus* induces hypermethylation of the imprinted *IGF2* gene promoter in murine placental tissue [157] suggesting that bacterial infections during pregnancy might epigenetically affect genes that play an important role in fetal development.

Mechanisms controlling gene expression at the chromatin level show high levels of complexity. Various bacterial products can affect them in many ways, through activation of signaling cascades or directly in the nucleus. The most of the reported chromatin modifications induced by bacteria are histone modification (acetylation/[deacetylation and phosphorylation/dephosphorylation) events generated through activation of host cell signaling cascades. Among the host signaling pathways that certain bacteria activate, MAPKs, NF- κ B, and PI3K pathways are known to activate the kinases. *B. vulgatus* induces TGF- β anti inflammatory pathway, and H3 de-acetylation via HDAC recruitment at proinflammatory gene promoters [158]. Bacterial toxins also dampen the host innate immune responses by inhibiting H3 phosphorylation/acetylation. The lethal toxin (LT) from *Bacillus anthracis*, the agent of anthrax, cleaves and inactivates MAPKKs leading to disruption in MAPK signaling [159].

H. pylori is also capable of modifying histones via the modulation of the MAPK pathway. The peptidyl prolyl cis, trans isomerase HP0175 secreted by *H. pylori* binds the innate immunity receptor TLR4 and activates MAPKs in monocytes [160]. Another study showed that exposure of *H. pylori* to gastric epithelial cells promotes release of HDAC1 from the promoter of the cell-cycle regulator gene *p21^{WAF}*, hyperacetylation of H4, and increased expression of *p21^{WAF}*. Chromatin alterations might contribute to the effects of *H. pylori* on cell cycle progression, cellular proliferation, and cell death [161].

Pore-forming toxins use another signaling pathway to change histone marks. He-La cells, *L. monocytogenes* listeriolysin O (LLO) promotes histone deacetylation and dephosphorylation in a subset of immune genes such as *CXCL2*, *MKP2* and *IFIT3* which cause repression [162]. Other pore-forming toxins, such as PFO of *Clostridium perfringens*, PLY of *Streptococcus pneumoniae*, and Aerolysin from *Aeromonas hydrophila* have similar effects, leading to histone modification [163].

Bacteria can also produce metabolites acting as inhibitors of chromatin-modifying enzymes. One such product is butyric acid, a short-chain fatty acid acting as a potent inhibitor of HDACs [164]. The adverse effect of *Porphyromonas gingivalis* in reactivating latent viruses, such as human immunodeficiency virus (HIV) and Epstein – Barr syndrome (EBV), result from production of butyrate by this bacterium [165]. It is proposed that viral genes maintained silent by HDAC-containing complexes are reactivated following inhibition of HDACs by butyric acid. Butyric acid also exerts anti-inflammatory effects on the host, via epigenetic up-regulation of anti-inflammatory genes. These observations give the idea of the usage of butyrate-producing probiotic bacteria as immunosuppressors [166].

miRNAs are important regulators of immune responses that are induced in response to pathogenic bacteria, such as *H. pylori*, *Salmonella typhimurium*, *L. monocytogenes*, and *Mycobacterium bovis* BCG [167, 168]. Whether a bacterial stimulus induces expression of miRNA or endogenous siRNA acts at the chromatin level in the nucleus is unknown. *MiR-155*, *miR-146*, *miR-125*, *let-7* and *miR-21* which are involved in immune responses and protection

against harmful effects of inflammation, play important roles in the response of host cells to bacteria [169]. These miRNAs are involved in immune responses as well as in the protection against harmful effects of inflammation. miRNAs are involved in the regulation of proliferation, differentiation and apoptosis pathways induced by pathogenic bacteria in their host cells. A role for miRNAs that control bacteria populations and their production of metabolites leading to inflammatory conditions is beginning to emerge. It has got important implications both within the gut and in peripheral tissues.

Table 1. Selected epigenetic deregulations and inflammatory conditions

DISEASE	DNA METHYLATION	HISTONE MODIFICATION	Micro RNA
IBD	<i>ER, MYOD, GSPG2, p16, CDH1, GDNF, E-CAHERIN, DAPK, MDR1, SERPINA5, TNFRSF1A, AATK, GABRA5, MAPK10, STAT5A</i> hypermethylation [34-38]	<i>COL1, FOXP3</i> [32,43]	<i>miR-125, miR-155, miR-192, miR-150, miR-21, miR-196</i> [49-52]
RA	<i>FOXP3, SFRP1, SFRP4, EBF3, IRX1, DR3</i> hypermethylation [57, 68, 69] <i>L1 repeats, IL-6, IL-1, CXCL-12 CD40LG, IL1R2, CHI3L1, CASP1, STAT3, MAP3K5, MEFV, and WISP3</i> hypomethylation [12, 64-67, 70, 71]	<i>EXH2, SITR1, ADAMTS, p16, p21, VEGF, HIF-1a</i> [72-74]	<i>miR-21, miR-30a-3p, miR-146a, miR-155, miR-223, miR-182, miR-203, miR-124a, miR-18a</i> [4, 14, 83-87]
SLE	<i>IL-10, IL-13, IFIT2, IRF5, STAT1, CD5, HRES1, IFNGR2, 18S and 28S</i> hypomethylation [93-96]	<i>IFN</i> genes [88, 102-104]	<i>Let-7a, miR-21, miR-29, miR-146, miR-182, miR-17-92, miR-155, miR-1246, miR126, miR-23b, miR-31</i> [110, 115-119, 120,121]
MS	<i>PADI2</i> -hypomethylation [122, 123] <i>FOXP3</i> and <i>IL-17A</i> hypomethylation [126] <i>SHP1</i> -hypermethylation [125] Global hypermethylation in CD4+ T cell [128]	<i>TCF7L2, ID2, IDS</i> and <i>SOX2</i> [125, 126]	<i>miR-21, miR-29b, miR-115-miR-326-, miR-124, miR34a, miR-17-5p, miR-497, miR-193, and miR-126, miR-18b, miR-599, miR-96</i> [127-131]
Diabetes	<i>FOXP3</i> - hypermethylation [142] Global hypermethylation [138]	<i>CLTA4 TGF-β, NF-κB, p38, TLRs, and IL-6. NF-κB</i> [138, 139]	<i>miR-29, miR-510 miR-146a, miR-342, miR-191, miR-21, miR-34a, miR-101a and miR-30b</i> [142-144]
Bacterial infection	<i>H. pylori</i> induced hypermethylation seen at the <i>CDH1, USF1, USF2, WWOX, MLH1</i> [147] <i>E. coli</i> (UPEC) induced hypermethylation of <i>CDKN2A</i> [150]	<i>B. vulgatus</i> induces TGF-1β pathway [152] <i>B. anthracis</i> inactivates MAPKKs [153] <i>H. pylori</i> induces cell-cycle regulator gene <i>p21</i> ^{WAF} [155]	<i>MiR-155, miR-146, miR-125, let-7 and miR-21</i> (168)

IBD; inflammatory bowel disease, RA; rheumatoid arthritis, SLE; systemic lupus erythematosus, MS; multiple sclerosis

3. EPIGENETICS, INFLAMMATION AND CANCER

Cancer has traditionally been described as a disease that is driven by the accumulation of genetic variations [170]. In addition to genetic alterations, it is well known now that epigenetic changes also play an important role in the development and progression of cancers [15, 171]. Self-sufficient proliferation, insensitivity to anti-proliferative signals, evasion of apoptosis, limitless replicative capacity, the maintenance of vascularization, invasion and metastasis are the hallmarks of cancer [170, 172]. In fact, the genetic background of cancer is relatively straightforward: mutation of tumor suppressors and/or oncogenes causes either loss or gain of function and abnormal expression. However, the epigenetic pathway to cancer is more complex and comprised by chromatin structure including DNA methylation, histone variants and modifications, nucleosome remodeling as well as small non-coding regulatory elements. During the carcinogenesis, the epigenome goes through multiple alterations, including a genome-wide loss of DNA methylation, frequent increases in promoter methylation of CpG islands, changes in nucleosome occupancy and modification profiles [173, 174]. Outcome of next generation sequencing studies has been the discovery of many inactivating mutations in genes that have potential to disrupt DNA methylation, histone modification and gene expression. The current studies have indicated that genetic and epigenetic mechanisms are not two separate states in cancer; they intertwine and take advantage of each other during tumorigenesis. Alterations in epigenetic mechanisms can lead to genetic mutations and genetic mutations in epigenetic regulators lead to an altered epigenome [174].

Inflammation is part of our natural healing process and there are two types of inflammation; chronic and acute. Acute inflammation plays a beneficial role against infection and injury. However, inadequate resolution of inflammation and uncontrolled inflammatory reactions can arouse a state of chronic inflammation and may play role in cancer development and progression [5, 175]. The link between cancer and inflammation has long been recognized. Accumulating evidence from preclinical and clinical studies suggested that persistent inflammation functions as a driving force in the journey to cancer [176]. It was estimated earlier that out of 2.2 million cancer cases diagnosed in the world on average more than 15% cases could be rooted to infection [177]. For instance, cancer of stomach, liver, gallbladder, prostate, and pancreas are usually related to gastric inflammation [176]. There is about 1.2 million new patients diagnosed with colorectal cancer and patients suffering from inflammatory bowel disorders, such as ulcerative colitis and Chron's disease, have an increased risk of developing colorectal cancer [176, 178]. Beside that management of colitis with anti-inflammatory therapy reduces this risk [179, 180].

The association between inflammation and cancer has two paradigm; inflammation cause cancer (extrinsic mechanism), cancer cells produce toxins and alert the immune system, when cancer has started, inflammation makes the tumor grow, creates a blood supply then spreads the cancer (intrinsic mechanism) [181]. Chronic hepatitis based hepatocellular carcinoma and colon cancer caused by inflammatory bowel diseases are two examples for extrinsic mechanisms [178, 182]. The intratumoral inflammatory microenvironment acts as a favorable ground for premalignant cells to attain malignant properties by allowing tumor cells to escape from host immune attack, inducing epigenetic changes, enforcing epithelial to mesenchymal transition, and providing unlimited growth potential, thereby creating a vicious cancer-

inflammation-cancer axis [176]. For instance, the mutation of RAS family proteins trigger RAS-RAF signaling and results in aggressive tumor formation [183, 184].

Several biochemical processes that altered during chronic inflammation have been implicated in carcinogenesis. These include shifting cellular redox balance toward oxidative stress, induction of genomic instability, increase DNA breakage, stimulation of cell proliferation, metastasis and angiogenesis, deregulation of cellular epigenetic control of gene expression. One of the mechanisms by which consistence inflammation can initiate tumorigenesis is the generation of oxidants and subsequent DNA damage leading to activation of oncogenes and/or silencing of tumor suppressors [185]. A wide variety of inflammatory and innate immune cells are often gathered at the site of infection. As a response to proinflammatory stimuli, activated inflammatory/immune cells generate reactive oxygen species (ROS) and reactive nitrogen species (RNS), which cause DNA damage. DNA damage can lead to activation of oncogenes [186]. Chronic exposure to ultraviolet (UV) B radiation is known to trigger inflammatory tissue damage and skin cancer through the activation of *ras* oncogene and loss of function of *p53* tumor suppressor gene [187]. NO, another reactive species, plays a role in inflammation related carcinogenesis by modification of DNA and inactivation of DNA repair enzymes [186]. 8-Oxo-7,8-dihydro-2'-deoxyguanosine (8-oxo-dG), a major biochemical hallmark of oxidative and mutagenic DNA damage, has been found to be produced in association with *H.pylori* induced gastric and TNF- α -induced pulmonary carcinogenesis. In general, inflammation driven activation of various protein kinases that include Janus-activated kinase (JAK), Akt, and mitogen-activated protein (MAP) kinases unbounded transmits growth signals, allowing cells to acquire a malignant phenotype [181, 188, 189].

Beside the mutations in the genomic DNA, epigenetic alterations can change gene expression via DNA methylation, histone modifications and non-coding regulatory elements. Traditionally, DNA methylation studies mostly have focused on CpG islands. DNA methylation, the addition of a methyl group to the 5-position of cytosine base in the DNA, represents a critical epigenetic control of gene expression [174, 190, 191]. The CpG hypermethylation of E-cadherin gene in intestinal metaplasia in *H. pylori* infected patients suggests DNA hypermethylation as a preneoplastic event in gastric cancer [192, 193]. Moreover, *H. pylori* infection has been shown to cause DNA involvement of epigenetic alterations in inflammation-associated cancers. Gene silencing via promoter hypermethylation in tumor suppressor genes *p16*, *RUNX3*, *MLH1* and *HPP1* are seen in ulcerative colitis and Barretts esophagus, which are closely associated with gastric carcinogenesis [193, 194]. Tissue specific expressed and imprinted genes can show hypomethylation, which may contribute to cancer cell phenotype. Recent genome wide methylation analyses have allowed to systematically discovering the key features of tumor methylomes. These include not just hypermethylated CpG islands, but also large (several 100 kb to several Mb) partially methylated domains that are localized at intergenic regions [195]. Another novel feature of methylomes is DNA methylation valleys (DMV) that are strongly hypomethylated in most tissues, and become hypermethylated in several cancers [196]. Abu-Remaileh et al. has shown the dynamic hypermethylation of DMVs and identify it as an early event in inflammation induced intestinal tumorigenesis. Aberrant methylation of DMVs may be novel epigenetic programs that links intestinal inflammation to colon cancer [197].

Epigenetic mechanisms may modulate the expression of pro-inflammatory cytokine TNF- α , interleukins, tumor suppressor genes, oncogenes and autocrine and paracrine activation of the transcription factor NF- κ B [3, 198]. In areas of tissue inflammation, activated neutrophils

and eosinophils release HOCl and HOBr, which react with DNA to produce 5-chlorocytosine and 5-bromocytosine, respectively [176, 199]. Neither methyl-binding proteins nor DNA methyltransferase-1 (DNMT-1) can distinguish between these inflammation-damaged 5-halocytosines and 5-methylcytosine. Thus, the formation and persistence of 5-halocytosine residues in the DNA of cells at the site of inflammation may lead to inappropriate de novo DNA methylation and represent another important link between inflammation and cancer development [176, 199]. Treatment of human multiple myeloma KAS 6/1 cells with a proinflammatory cytokine IL-6 was resulted an increased expression of *DNMT-1* and hypermethylation of the *p53* promoter. Zebularine is the DNMT inhibitor and restores the normal *p53* function by demethylating *p53* promoter [200].

On the other hand, hypomethylation of epidermal growth factor (*EGFR*) gene in IL-6-transfected malignant cholangiocytes is involved in increasing *EGFR* expression, thereby promoting growth of cholangiocarcinoma cells [201]. Moreover, silencing of cytokine signaling by epigenetic modification revealed resistance to apoptosis in cholangiocarcinoma cells via sustained inflammatory signaling mediated by IL-6/signal transducer and activators of transcription (STAT-3) and subsequent expression of myeloid cell lymphoma-1 (Mcl-1) [176, 202].

Gene expression changes can be modulated by the modification of histones and their variants. Indeed, aberrant histone modifications are already known hallmarks of different cancer types [27]. Histone related enzymes HDACs and HATs have been linked to chronic inflammatory responses and cancer [203, 204]. The acetylation of lysine residues on the N-terminus of histones by HATs activates gene transcription; on the other hand removal of an acetyl group from lysine residues in histone tails by HDACs leads to transcriptional suppression of genes. Hypermethylation itself may also lead to histone deacetylation and chromatin condensation, thereby instigating a transcriptionally silenced state [205, 206]. Transcriptional regulation of inflammatory cytokines and related pathways such as NF- κ B by proinflammatory stimuli depends on histone regulations. Therefore, the inflammation induced alterations in histones and linking upregulation of COX-2 and NF- κ B suggest that inflammation can change the cellular epigenetic machinery, thereby indirectly induce the genetic instability of cancer cells [176]. Beside intranuclear functions, histones can also be released into the extracellular space by damaged and activated cells, presenting toxic or pro-inflammatory activity *in vivo* and *in vitro* [26, 27]. When extracellular histones released, they selectively bind to Toll-like receptors to produce pro-inflammatory cytokines (TNF- α and IL-6), which in turn accelerates inflammatory response and tissue damage [207].

Several miRNAs can behave as oncogenes or tumor suppressors in cancer. Cytokine regulated miRNAs can function as a critical link between inflammation and cancer. UC patients revealed the upregulation of several miRNAs, including *miR-21* and *miR-155* according to healthy controls [208]. Upregulation of *miR-let-7a* may play a role in inflammation-associated cholangiosarcoma. Stable overexpression of IL-6 induces *miR-let-7a*, which contributes phosphorylation of STAT3 in malignant cholangiocytes [181, 209]. As an alternative way, IL-6 can enhance the growth of human cholangiocarcinoma cells by downregulating *miR-370* [210]. Therefore, the role of miRNA in inflammation and cancer mostly depends on mechanisms by which tumor cells and tumor microenvironment regulate miRNA biogenesis and miRNA-mediated epigenetic control of proinflammatory gene expression [181].

CONCLUSION

Interactions of cells in the innate and adaptive immune systems, and inflammatory mediators orchestrate aspects of the acute and chronic inflammation that underlie the diseases of many organs. Understanding of molecular mechanisms in those processes will not only provide important insights into the aetiology of chronic inflammatory diseases, but also provide new targets for the development of novel anti-inflammatory drugs based on triggering pathways.

Epigenetic regulation of gene expression is a novel approach to treat diseases. Epigenetic reprogramming has potential preventive and therapeutic purposes in inflammation and autoimmunity. Inhibition of histone deacetylase enzymes (HDAC) initially used in the treatment of cancer, but recent reports suggesting inhibitors of HDAC (HDACi) could be utilized in treating inflammatory diseases as well. Chromatin regulatory protein families have also provided alternative potential therapeutic targets for new anti-cancer and anti-inflammatory drugs. Many of the inhibitors that have been described to date have targeted chromatin modulators. However, their non-selectivity and consequent side effects are still limiting factors in their broader clinical use. A number of effective small molecular inhibitors of individual human methyltransferases and lysine demethylases have been developed. Inhibition of chromatin reader domains is showing great promise as a selective pharmacological strategy.

As miRNAs have been functionally connected to the development of chronic inflammation, alterations to miRNA levels could be a diagnostic way to detect the presence of chronic inflammatory states in patients. miRNAs are currently being used as diagnostics for at least some types of diseases, including some forms of chronic inflammation such as colitis and IBD. The identification of miRNAs in blood serum as well as other biological fluids, samples obtained through non-invasive methods, has opened the door for diagnosis of various diseases. Specific secreted miRNAs in the serum can be used to diagnose chronic inflammatory diseases such as IBD.

The widely use of new generation whole genome technologies and meta-analysis, lead to understand epigenetic mechanism and their roles in disease development. Enlightenment of genetic and epigenetic modifications in inflammation will guide to create new therapeutic approach in practice. However, functional studies are still emerging to validate molecular targets for smart drugs development.

REFERENCES

- [1] Libby P. Inflammatory mechanisms: the molecular basis of inflammation and disease. *Nutr Rev*, 2007, 65(12 Pt 2), S140-6.
- [2] Eming, SA; Krieg, T; Davidson, JM. Inflammation in wound repair: molecular and cellular mechanisms. *J Invest Dermatol*, 2007, 127(3), 514-25.
- [3] Hoesel, B; Schmid, JA. The complexity of NF- κ B signaling in inflammation and cancer. *Mol Cancer*, 2013, 12(86).
- [4] Disis, ML. Immune regulation of cancer. *J Clin Oncol*, 2010, 28(29), 4531-8.

- [5] Balkwill, F; Mantovani, A. Inflammation and cancer: back to Virchow? *Lancet*, 2001, 357(9255), 539-45.
- [6] Wertz, IE; O'Rourke, KM; Zhou, H; Eby, M; Aravind, L; Seshagiri, S; et al. Deubiquitination and ubiquitin ligase domains of A20 downregulate NF-kappaB signalling. *Nature*, 2004, 430(7000), 694-9.
- [7] Trompouki, E; Hatzivassiliou, E; Tsichritzis, T; Farmer, H; Ashworth, A; Mosialos, G. CYLD is a deubiquitinating enzyme that negatively regulates NF-kappaB activation by TNFR family members. *Nature*, 2003, 424(6950), 793-6.
- [8] Holliday, R; Pugh, JE. DNA modification mechanisms and gene activity during development. *Science*, 1975, 187(4173), 226-32.
- [9] Jones, PA; Liang, G. Rethinking how DNA methylation patterns are maintained. *Nat Rev Genet*, 2009, 10(11), 805-11.
- [10] Jones, PA. Functions of DNA methylation: islands, start sites, gene bodies and beyond. *Nat Rev Genet*, 2012, 13(7), 484-92.
- [11] Simonsson, S; Gurdon, J. DNA demethylation is necessary for the epigenetic reprogramming of somatic cell nuclei. *Nat Cell Biol*, 2004, 6(10), 984-90.
- [12] Lai, AY; Mav, D; Shah, R; Grimm, SA; Phadke, D; Hatzi, K; et al. DNA methylation profiling in human B cells reveals immune regulatory elements and epigenetic plasticity at Alu elements during B-cell activation. *Genome Res*, 2013, 23(12), 2030-41.
- [13] Zhang, Z; Zhang, R. Epigenetics in autoimmune diseases: Pathogenesis and prospects for therapy. *Autoimmun Rev*, 2015, 14(10), 854-63.
- [14] Raisch, J; Darfeuille-Michaud, A; Nguyen, HT. Role of microRNAs in the immune system, inflammation and cancer. *World J Gastroenterol*, 2013, 19(20), 2985-96.
- [15] Baylin, SB; Jones, PA. A decade of exploring the cancer epigenome - biological and translational implications. *Nat Rev Cancer*, 2011, 11(10), 726-34.
- [16] Schübeler, D. Function and information content of DNA methylation. *Nature*, 2015, 517(7534), 321-6.
- [17] Campos, EI; Reinberg, D. Histones: annotating chromatin. *Annu Rev Genet*, 2009, 43(559-99).
- [18] Luger, K; Richmond, TJ. The histone tails of the nucleosome. *Curr Opin Genet Dev*, 1998, 8(2), 140-6.
- [19] Singh, R; Mortazavi, A; Telu, KH; Nagarajan, P; Lucas, DM; Thomas-Ahner, JM; et al. Increasing the complexity of chromatin: functionally distinct roles for replication-dependent histone H2A isoforms in cell proliferation and carcinogenesis. *Nucleic Acids Res*, 2013, 41(20), 9284-95.
- [20] Morillon, A; Karabetsov, N; Nair, A; Mellor, J. Dynamic lysine methylation on histone H3 defines the regulatory phase of gene transcription. *Mol Cell*, 2005, 18(6), 723-34.
- [21] Eberharter, A; Becker, PB. Histone acetylation: a switch between repressive and permissive chromatin. Second in review series on chromatin dynamics. *EMBO Rep*, 2002, 3(3), 224-9.
- [22] Cantley, MD; Haynes, DR. Epigenetic regulation of inflammation: progressing from broad acting histone deacetylase (HDAC) inhibitors to targeting specific HDACs. *Inflammopharmacology*, 2013, 21(4), 301-7.
- [23] Jiang, C; Pugh, BF. Nucleosome positioning and gene regulation: advances through genomics. *Nat Rev Genet*, 2009, 10(3), 161-72.

- [24] Martens, JA; Winston, F. Recent advances in understanding chromatin remodeling by Swi/Snf complexes. *Curr Opin Genet Dev*, 2003, 13(2), 136-42.
- [25] Wang, KC; Yang, YW; Liu, B; Sanyal, A; Corces-Zimmerman, R; Chen, Y; et al. A long noncoding RNA maintains active chromatin to coordinate homeotic gene expression. *Nature*, 2011, 472(7341), 120-4.
- [26] Allam, R; Kumar, SV; Darisipudi, MN; Anders, HJ. Extracellular histones in tissue injury and inflammation. *J Mol Med (Berl)*, 2014, 92(5), 465-72.
- [27] Chen, R; Kang, R; Fan, XG; Tang, D. Release and activity of histone in diseases. *Cell Death Dis*, 2014, 5(e1370).
- [28] Grammatikakis, I; Panda, AC; Abdelmohsen, K; Gorospe, M. Long noncoding RNAs(lncRNAs) and the molecular hallmarks of aging. *Aging (Albany NY)*, 2014, 6(12), 992-1009.
- [29] Marques-Rocha, JL; Samblas, M; Milagro, FI; Bressan, J; Martínez, JA; Martí, A. Noncoding RNAs, cytokines, and inflammation-related diseases. *FASEB J*, 2015, 29(9), 3595-611.
- [30] Alexander, M; O'Connell, RM. Noncoding RNAs and chronic inflammation: Micro-managing the fire within. *Bioessays*, 2015, 37(9), 1005-15.
- [31] Manichanh, C; Borruel, N; Casellas, F; Guarner, F. The gut microbiota in IBD. *Nat Rev Gastroenterol Hepatol*, 2012, 9(10), 599-608.
- [32] Däbritz, J; Menheniott, TR. Linking immunity, epigenetics, and cancer in inflammatory bowel disease. *Inflamm Bowel Dis*, 2014, 20(9), 1638-54.
- [33] Glória, L; Cravo, M; Pinto, A; de Sousa, LS; Chaves, P; Leitão, CN; et al. DNA hypomethylation and proliferative activity are increased in the rectal mucosa of patients with long-standing ulcerative colitis. *Cancer*, 1996, 78(11), 2300-6.
- [34] Schuebel, KE; Chen, W; Cope, L; Glöckner, SC; Suzuki, H; Yi, JM; et al. Comparing the DNA hypermethylome with gene mutations in human colorectal cancer. *PLoS Genet*, 2007, 3(9), 1709-23.
- [35] Lin, Z; Hegarty, JP; Cappel, JA; Yu, W; Chen, X; Faber, P; et al. Identification of disease-associated DNA methylation in intestinal tissues from patients with inflammatory bowel disease. *Clin Genet*, 2011, 80(1), 59-67.
- [36] Kellermayer, M. [To the Editors: The Hungarian Healthcare Society's assessment and recommendations for medical specialty education in Hungary]. *Orv Hetil*, 2012, 153(4), 158.
- [37] Tahara, T; Shibata, T; Nakamura, M; Yamashita, H; Yoshioka, D; Okubo, M; et al. Effect of MDR1 gene promoter methylation in patients with ulcerative colitis. *Int J Mol Med*, 2009, 23(4), 521-7.
- [38] Häsler, R; Feng, Z; Bäckdahl, L; Spehlmann, ME; Franke, A; Teschendorff, A; et al. A functional methylome map of ulcerative colitis. *Genome Res*, 2012, 22(11), 2130-7.
- [39] Karatzas, PS; Mantzaris, GJ; Safioleas, M; Gazouli, M. DNA methylation profile of genes involved in inflammation and autoimmunity in inflammatory bowel disease. *Medicine (Baltimore)*, 2014, 93(28), e309.
- [40] Stylianou, E. Epigenetics: the fine-tuner in inflammatory bowel disease? *Curr Opin Gastroenterol*, 2013, 29(4), 370-7.
- [41] Nimmo, ER; Prendergast, JG; Aldhous, MC; Kennedy, NA; Henderson, P; Drummond, HE; et al. Genome-wide methylation profiling in Crohn's disease identifies altered

- epigenetic regulation of key host defense mechanisms including the Th17 pathway. *Inflamm Bowel Dis*, 2012, 18(5), 889-99.
- [42] Tsaprouni, LG; Ito, K; Powell, JJ; Adcock, IM; Pouchard, N. Differential patterns of histone acetylation in inflammatory bowel diseases. *J Inflamm (Lond)*, 2011, 8(1), 1.
- [43] Tao, R; de Zoeten, EF; Ozkaynak, E; Chen, C; Wang, L; Porrett, PM; et al. Deacetylase inhibition promotes the generation and function of regulatory T cells. *Nat Med*, 2007, 13(11), 1299-307.
- [44] Glauben, R; Siegmund, B. Inhibition of histone deacetylases in inflammatory bowel diseases. *Mol Med*, 2011, 17(5-6), 426-33.
- [45] McKenna, LB; Schug, J; Vourekas, A; McKenna, JB; Bramswig, NC; Friedman, JR; et al. MicroRNAs control intestinal epithelial differentiation, architecture, and barrier function. *Gastroenterology*, 2010, 139(5), 1654-64, 1664.e1.
- [46] Ventham, NT; Kennedy, NA; Nimmo, ER; Satsangi, J. Beyond gene discovery in inflammatory bowel disease: the emerging role of epigenetics. *Gastroenterology*, 2013, 145(2), 293-308.
- [47] Pekow, JR; Kwon, JH. MicroRNAs in inflammatory bowel disease. *Inflamm Bowel Dis*, 2012, 18(1), 187-93.
- [48] Chapman, CG; Pekow, J. The emerging role of miRNAs in inflammatory bowel disease: a review. *Therap Adv Gastroenterol*, 2015, 8(1), 4-22.
- [49] Wu, F; Zikusoka, M; Trindade, A; Dassopoulos, T; Harris, ML; Bayless, TM; et al. MicroRNAs are differentially expressed in ulcerative colitis and alter expression of macrophage inflammatory peptide-2 alpha. *Gastroenterology*, 2008, 135(5), 1624-1635 e24.
- [50] Bian, Z; Li, L; Cui, J; Zhang, H; Liu, Y; Zhang, CY; et al. Role of miR-150-targeting c-Myb in colonic epithelial disruption during dextran sulphate sodium-induced murine experimental colitis and human ulcerative colitis. *J Pathol*, 2011, 225(4), 544-53.
- [51] Fasseu, M; Treton, X; Guichard, C; Pedruzzi, E; Cazals-Hatem, D; Richard, C; et al. Identification of restricted subsets of mature microRNA abnormally expressed in inactive colonic mucosa of patients with inflammatory bowel disease. *PLoS One*, 2010, 5(10).
- [52] Brest, P; Lapaquette, P; Souidi, M; Lebrigand, K; Cesaro, A; Vouret-Craviari, V; et al. A synonymous variant in IRGM alters a binding site for miR-196 and causes deregulation of IRGM-dependent xenophagy in Crohn's disease. *Nat Genet*, 2011, 43(3), 242-5.
- [53] Wu, F; Guo, NJ; Tian, H; Marohn, M; Gearhart, S; Bayless, TM; et al. Peripheral blood microRNAs distinguish active ulcerative colitis and Crohn's disease. *Inflamm Bowel Dis*, 2011, 17(1), 241-50.
- [54] Petronis, A; Petroniene, R. Epigenetics of inflammatory bowel disease. *Gut*, 2000, 47(2), 302-6.
- [55] Jonsson, B; Kobelt, G; Smolen, J. The burden of rheumatoid arthritis and access to treatment: uptake of new therapies. *Eur J Health Econ*, 2008, 8 Suppl 2(S61-86).
- [56] Padyukov, L; Seielstad, M; Ong, RT; Ding, B; Ronnelid, J; Seddighzadeh, M; et al. A genome-wide association study suggests contrasting associations in ACPA-positive versus ACPA-negative rheumatoid arthritis. *Ann Rheum Dis*, 2011, 70(2), 259-65.
- [57] Klein, K; Ospelt, C; Gay, S. Epigenetic contributions in the development of rheumatoid arthritis. *Arthritis Res Ther*, 2012, 14(6), 227.

- [58] Karouzakis, E; Gay, RE; Michel, BA; Gay, S; Neidhart, M. DNA hypomethylation in rheumatoid arthritis synovial fibroblasts. *Arthritis Rheum*, 2009, 60(12), 3613-22.
- [59] Karouzakis, E; Gay, RE; Gay, S; Neidhart, M. Increased recycling of polyamines is associated with global DNA hypomethylation in rheumatoid arthritis synovial fibroblasts. *Arthritis Rheum*, 2012, 64(6), 1809-17.
- [60] Lafyatis, R; Remmers, EF; Roberts, AB; Yocum, DE; Sporn, MB; Wilder, RL. Anchorage-independent growth of synoviocytes from arthritic and normal joints. Stimulation by exogenous platelet-derived growth factor and inhibition by transforming growth factor-beta and retinoids. *J Clin Invest*, 1989, 83(4), 1267-76.
- [61] Kinne, RW; Boehm, S; Iftner, T; Aigner, T; Vornehm, S; Weseloh, G; et al. Synovial fibroblast-like cells strongly express jun-B and C-fos proto-oncogenes in rheumatoid- and osteoarthritis. *Scand J Rheumatol Suppl*, 1995, 101(121-5).
- [62] Fu, LH; Ma, CL; Cong, B; Li, SJ; Chen, HY; Zhang, JG. Hypomethylation of proximal CpG motif of interleukin-10 promoter regulates its expression in human rheumatoid arthritis. *Acta Pharmacol Sin*, 2011, 32(11), 1373-80.
- [63] Song, YJ; Li, G; He, JH; Guo, Y; Yang, L. Bioinformatics-Based Identification of MicroRNA-Regulated and Rheumatoid Arthritis-Associated Genes. *PLoS One*, 2015, 10(9), e0137551.
- [64] Kuchen, S; Seemayer, CA; Rethage, J; von Knoch, R; Kuenzler, P; Beat, AM; et al. The L1 retroelement-related p40 protein induces p38delta MAP kinase. *Autoimmunity*, 2004, 37(1), 57-65.
- [65] Garaud, S; Le Dantec, C; Jousse-Joulin, S; Hanrotel-Saliou, C; Saraux, A; Mageed, RA; et al. IL-6 modulates CD5 expression in B cells from patients with lupus by regulating DNA methylation. *J Immunol*, 2009, 182(9), 5623-32.
- [66] Karouzakis, E; Rengel, Y; Jungel, A; Kolling, C; Gay, RE; Michel, BA; et al. DNA methylation regulates the expression of CXCL12 in rheumatoid arthritis synovial fibroblasts. *Genes Immun*, 2011, 12(8), 643-52.
- [67] Bertoncelj, V; Pelipenko, J; Kristl, J; Jeras, M; Cukjati, M; Kocbek, P. Development and bioevaluation of nanofibers with blood-derived growth factors for dermal wound healing. *Eur J Pharm Biopharm*, 2014, 88(1), 64-74.
- [68] Park, SH; Kim, SK; Choe, JY; Moon, Y; An, S; Park, MJ; et al. Hypermethylation of EBF3 and IRX1 genes in synovial fibroblasts of patients with rheumatoid arthritis. *Mol Cells*, 2013, 35(4), 298-304.
- [69] Takami, N; Osawa, K; Miura, Y; Komai, K; Taniguchi, M; Shiraishi, M; et al. Hypermethylated promoter region of DR3, the death receptor 3 gene, in rheumatoid arthritis synovial cells. *Arthritis Rheum*, 2006, 54(3), 779-87.
- [70] Chabchoub, G; Uz, E; Maalej, A; Mustafa, CA; Rebai, A; Mnif, M; et al. Analysis of skewed X-chromosome inactivation in females with rheumatoid arthritis and autoimmune thyroid diseases. *Arthritis Res Ther*, 2009, 11(4), R106.
- [71] Liao, J; Liang, G; Xie, S; Zhao, H; Zuo, X; Li, F; et al. CD40L demethylation in CD4(+) T cells from women with rheumatoid arthritis. *Clin Immunol*, 2012, 145(1), 13-8.
- [72] Huber, LC; Stanczyk, J; Jüngel, A; Gay, S. Epigenetics in inflammatory rheumatic diseases. *Arthritis Rheum*, 2007, 56(11), 3523-31.
- [73] Nishida, K; Komiyama, T; Miyazawa, S; Shen, ZN; Furumatsu, T; Doi, H; et al. Histone deacetylase inhibitor suppression of autoantibody-mediated arthritis in mice via

- regulation of p16INK4a and p21(WAF1/Cip1) expression. *Arthritis Rheum*, 2004, 50(10), 3365-76.
- [74] Manabe, H; Nasu, Y; Komiyama, T; Furumatsu, T; Kitamura, A; Miyazawa, S; et al. Inhibition of histone deacetylase down-regulates the expression of hypoxia-induced vascular endothelial growth factor by rheumatoid synovial fibroblasts. *Inflamm Res*, 2008, 57(1), 4-10.
- [75] Stanczyk, J; Ospelt, C; Karouzakis, E; Filer, A; Raza, K; Kolling, C; et al. Altered expression of microRNA-203 in rheumatoid arthritis synovial fibroblasts and its role in fibroblast activation. *Arthritis Rheum*, 2011, 63(2), 373-81.
- [76] Miao, CG; Yang, YY; He, X; Li, J. New advances of DNA methylation and histone modifications in rheumatoid arthritis, with special emphasis on MeCP2. *Cell Signal*, 2013, 25(4), 875-82.
- [77] Zhou, Q; Long, L; Shi, G; Zhang, J; Wu, T; Zhou, B. Research of the methylation status of miR-124a gene promoter among rheumatoid arthritis patients. *Clin Dev Immunol*, 2013, 2013(524204).
- [78] Nakano, K; Whitaker, JW; Boyle, DL; Wang, W; Firestein, GS. DNA methylome signature in rheumatoid arthritis. *Ann Rheum Dis*, 2013, 72(1), 110-7.
- [79] Wilkinson, J. Study reveals a DNA methylome signature in rheumatoid arthritis. *Epigenomics*, 2012, 4(5), 481.
- [80] Ai, R; Whitaker, JW; Boyle, DL; Tak, PP; Gerlag, DM; Wang, W; et al. DNA Methylome Signature in Synoviocytes From Patients With Early Rheumatoid Arthritis Compared to Synoviocytes From Patients With Longstanding Rheumatoid Arthritis. *Arthritis Rheumatol*, 2015, 67(7), 1978-80.
- [81] de la Rica, L; Urquiza, JM; Gomez-Cabrero, D; Islam, AB; Lopez-Bigas, N; Tegner, J; et al. Identification of novel markers in rheumatoid arthritis through integrated analysis of DNA methylation and microRNA expression. *J Autoimmun*, 2013, 41(6-16).
- [82] Ballestar, E. Genetics: Insights into RA pathogenesis from DNA methylome analysis. *Nat Rev Rheumatol*, 2015, 11(7), 386-8.
- [83] Kurowska-Stolarska, M; Alivernini, S; Ballantine, LE; Asquith, DL; Millar, NL; Gilchrist, DS; et al. MicroRNA-155 as a proinflammatory regulator in clinical and experimental arthritis. *Proc Natl Acad Sci U S A*, 2011, 108(27), 11193-8.
- [84] Fulci, V; Scappucci, G; Sebastiani, GD; Giannitti, C; Franceschini, D; Meloni, F; et al. miR-223 is overexpressed in T-lymphocytes of patients affected by rheumatoid arthritis. *Hum Immunol*, 2010, 71(2), 206-11.
- [85] Nakasa, T; Miyaki, S; Okubo, A; Hashimoto, M; Nishida, K; Ochi, M; et al. Expression of microRNA-146 in rheumatoid arthritis synovial tissue. *Arthritis Rheum*, 2008, 58(5), 1284-92.
- [86] Stittrich, AB; Haftmann, C; Sgouroudis, E; Köhl, AA; Hegazy, AN; Panse, I; et al. The microRNA miR-182 is induced by IL-2 and promotes clonal expansion of activated helper T lymphocytes. *Nat Immunol*, 2010, 11(11), 1057-62.
- [87] Nakamachi, Y; Kawano, S; Takenokuchi, M; Nishimura, K; Sakai, Y; Chin, T; et al. MicroRNA-124a is a key regulator of proliferation and monocyte chemoattractant protein 1 secretion in fibroblast-like synoviocytes from patients with rheumatoid arthritis. *Arthritis Rheum*, 2009, 60(5), 1294-304.

- [88] Trenkmann, M; Brock, M; Gay, RE; Michel, BA; Gay, S; Huber, LC. Tumor necrosis factor alpha-induced microRNA-18a activates rheumatoid arthritis synovial fibroblasts through a feedback loop in NF-kappaB signaling. *Arthritis Rheum*, 2013, 65(4), 916-27.
- [89] Miceli-Richard, C. Epigenetics and lupus. *Joint Bone Spine*, 2015, 82(2), 90-3.
- [90] Mak, A; Kow, NY. The pathology of T cells in systemic lupus erythematosus. *J Immunol Res*, 2014, 2014(419029).
- [91] Richardson, BC; Patel, DR. Epigenetics in 2013. DNA methylation and miRNA: key roles in systemic autoimmunity. *Nat Rev Rheumatol*, 2014, 10(2), 72-4.
- [92] Hughes, GC; Choubey, D. Modulation of autoimmune rheumatic diseases by oestrogen and progesterone. *Nat Rev Rheumatol*, 2014, 10(12), 740-51.
- [93] Schwartzman-Morris, J; Putterman, C. Gender differences in the pathogenesis and outcome of lupus and of lupus nephritis. *Clin Dev Immunol*, 2012, 2012(604892).
- [94] Lu, Q; Wu, A; Tesmer, L; Ray, D; Yousif, N; Richardson, B. Demethylation of CD40LG on the inactive X in T cells from women with lupus. *J Immunol*, 2007, 179(9), 6352-8.
- [95] Cannat, A; Seligmann, M. Induction by isoniazid and hydralazine of antinuclear factors in mice. *Clin Exp Immunol*, 1968, 3(1), 99-105.
- [96] Oelke, K; Lu, Q; Richardson, D; Wu, A; Deng, C; Hanash, S; et al. Overexpression of CD70 and overstimulation of IgG synthesis by lupus T cells and T cells treated with DNA methylation inhibitors. *Arthritis Rheum*, 2004, 50(6), 1850-60.
- [97] Javierre, BM; Fernandez, AF; Richter, J; Al-Shahrour, F; Martin-Subero, JI; Rodriguez-Ubreva, J; et al. Changes in the pattern of DNA methylation associate with twin discordance in systemic lupus erythematosus. *Genome Res*, 2010, 20(2), 170-9.
- [98] Wu, H; Zhao, M; Chang, C; Lu, Q. The real culprit in systemic lupus erythematosus: abnormal epigenetic regulation. *Int J Mol Sci*, 2015, 16(5), 11013-33.
- [99] Coit, P; Jeffries, M; Altorok, N; Dozmorov, MG; Koelsch, KA; Wren, JD; et al. Genome-wide DNA methylation study suggests epigenetic accessibility and transcriptional poisoning of interferon-regulated genes in naïve CD4⁺ T cells from lupus patients. *J Autoimmun*, 2013, 43(78-84).
- [100] Schett, G; Smole, J; Zimmermann, C; Hiesberger, H; Hoefler, E; Fournel, S; et al. The autoimmune response to chromatin antigens in systemic lupus erythematosus: autoantibodies against histone H1 are a highly specific marker for SLE associated with increased disease activity. *Lupus*, 2002, 11(11), 704-15.
- [101] Bruns, A; Bläss, S; Hausdorf, G; Burmester, GR; Hiepe, F. Nucleosomes are major T and B cell autoantigens in systemic lupus erythematosus. *Arthritis Rheum*, 2000, 43(10), 2307-15.
- [102] van Bavel, CC; Dieker, JW; Tamboer, WP; van der Vlag, J; Berden, JH. Lupus-derived monoclonal autoantibodies against apoptotic chromatin recognize acetylated conformational epitopes. *Mol Immunol*, 2010, 48(1-3), 248-56.
- [103] Hu, N; Qiu, X; Luo, Y; Yuan, J; Li, Y; Lei, W; et al. Abnormal histone modification patterns in lupus CD4⁺ T cells. *J Rheumatol*, 2008, 35(5), 804-10.
- [104] van Bavel, CC; Dieker, JW; Kroeze, Y; Tamboer, WP; Voll, R; Muller, S; et al. Apoptosis-induced histone H3 methylation is targeted by autoantibodies in systemic lupus erythematosus. *Ann Rheum Dis*, 2011, 70(1), 201-7.
- [105] Zhang, Z; Song, L; Maurer, K; Petri, MA; Sullivan, KE. Global H4 acetylation analysis by ChIP-chip in systemic lupus erythematosus monocytes. *Genes Immun*, 2010, 11(2), 124-33.

- [106] Zhao, M; Liu, S; Luo, S; Wu, H; Tang, M; Cheng, W; et al. DNA methylation and mRNA and microRNA expression of SLE CD4⁺ T cells correlate with disease phenotype. *J Autoimmun*, 2014, 54(127-36).
- [107] Lee, BH; Yegnasubramanian, S; Lin, X; Nelson, WG. Procainamide is a specific inhibitor of DNA methyltransferase 1. *J Biol Chem*, 2005, 280(49), 40749-56.
- [108] Deng, C; Lu, Q; Zhang, Z; Rao, T; Attwood, J; Yung, R; et al. Hydralazine may induce autoimmunity by inhibiting extracellular signal-regulated kinase pathway signaling. *Arthritis Rheum*, 2003, 48(3), 746-56.
- [109] Cornacchia, E; Golbus, J; Maybaum, J; Strahler, J; Hanash, S; Richardson, B. Hydralazine and procainamide inhibit T cell DNA methylation and induce autoreactivity. *J Immunol*, 1988, 140(7), 2197-200.
- [110] Duroux-Richard, I; Cuenca, J; Ponsolles, C; Pineiro, AB; Gonzalez, F; Roubert, C; et al. MicroRNA Profiling of B Cell Subsets from Systemic Lupus Erythematosus Patients Reveals Promising Novel Biomarkers. *Int J Mol Sci*, 2015, 16(8), 16953-65.
- [111] Luo, S; Liu, Y; Liang, G; Zhao, M; Wu, H; Liang, Y; et al. The role of microRNA-1246 in the regulation of B cell activation and the pathogenesis of systemic lupus erythematosus. *Clin Epigenetics*, 2015, 7(1), 24.
- [112] Dai, Y; Huang, YS; Tang, M; Lv, TY; Hu, CX; Tan, YH; et al. Microarray analysis of microRNA expression in peripheral blood cells of systemic lupus erythematosus patients. *Lupus*, 2007, 16(12), 939-46.
- [113] Aletaha, D; Landewe, R; Karonitsch, T; Bathon, J; Boers, M; Bombardier, C; et al. Reporting disease activity in clinical trials of patients with rheumatoid arthritis: EULAR/ACR collaborative recommendations. *Arthritis Rheum*, 2008, 59(10), 1371-7.
- [114] Pan, W; Zhu, S; Yuan, M; Cui, H; Wang, L; Luo, X; et al. MicroRNA-21 and microRNA-148a contribute to DNA hypomethylation in lupus CD4⁺ T cells by directly and indirectly targeting DNA methyltransferase 1. *J Immunol*, 2010, 184(12), 6773-81.
- [115] Wu, XN; Ye, YX; Niu, JW; Li, Y; Li, X; You, X; et al. Defective PTEN regulation contributes to B cell hyperresponsiveness in systemic lupus erythematosus. *Sci Transl Med*, 2014, 6(246), 246ra99.
- [116] Stagakis, E; Bertsias, G; Verginis, P; Nakou, M; Hatzia Apostolou, M; Kritikos, H; et al. Identification of novel microRNA signatures linked to human lupus disease activity and pathogenesis: miR-21 regulates aberrant T cell responses through regulation of PDCD4 expression. *Ann Rheum Dis*, 2011, 70(8), 1496-506.
- [117] Tang, Y; Luo, X; Cui, H; Ni, X; Yuan, M; Guo, Y; et al. MicroRNA-146A contributes to abnormal activation of the type I interferon pathway in human lupus by targeting the key signaling proteins. *Arthritis Rheum*, 2009, 60(4), 1065-75.
- [118] Luo, X; Yang, W; Ye, DQ; Cui, H; Zhang, Y; Hirankarn, N; et al. A functional variant in microRNA-146a promoter modulates its expression and confers disease risk for systemic lupus erythematosus. *PLoS Genet*, 2011, 7(6), e1002128.
- [119] A B, R K, N R, Meza, V; Morelato, R; Piccinni, D. [Malignant melanoma of the oral cavity]. *Rev Fac Cien Med Univ Nac Cordoba*, 2008, 65(2), 70-3.
- [120] Karrich, JJ; Jachimowski, LC; Libouban, M; Iyer, A; Brandwijk, K; Taanman-Kueter, EW; et al. MicroRNA-146a regulates survival and maturation of human plasmacytoid dendritic cells. *Blood*, 2013, 122(17), 3001-9.

- [121] Zhu, S; Pan, W; Song, X; Liu, Y; Shao, X; Tang, Y; et al. The microRNA miR-23b suppresses IL-17-associated autoimmune inflammation by targeting TAB2, TAB3 and IKK-alpha. *Nat Med*, 2012, 18(7), 1077-86.
- [122] Fan, W; Liang, D; Tang, Y; Qu, B; Cui, H; Luo, X; et al. Identification of microRNA-31 as a novel regulator contributing to impaired interleukin-2 production in T cells from patients with systemic lupus erythematosus. *Arthritis Rheum*, 2012, 64(11), 3715-25.
- [123] Mastronardi, FG; Noor, A; Wood, DD; Paton, T; Moscarello, MA. Peptidyl argininedeiminase 2 CpG island in multiple sclerosis white matter is hypomethylated. *J Neurosci Res*, 2007, 85(9), 2006-16.
- [124] Mastronardi, FG; Wood, DD; Mei, J; Raijmakers, R; Tseveleki, V; Dosch, HM; et al. Increased citrullination of histone H3 in multiple sclerosis brain and animal models of demyelination: a role for tumor necrosis factor-induced peptidylarginine deiminase 4 translocation. *J Neurosci*, 2006, 26(44), 11387-96.
- [125] Calabrese, R; Zampieri, M; Mechelli, R; Annibali, V; Guastafierro, T; Ciccarone, F; et al. Methylation-dependent PAD2 upregulation in multiple sclerosis peripheral blood. *Mult Scler*, 2012, 18(3), 299-304.
- [126] Kumagai, C; Kalman, B; Middleton, FA; Vyshkina, T; Massa, PT. Increased promoter methylation of the immune regulatory gene SHP-1 in leukocytes of multiple sclerosis subjects. *J Neuroimmunol*, 2012, 246(1-2), 51-7.
- [127] Jadidi-Niaragh, F; Mirshafiey, A. Th17 cell, the new player of neuroinflammatory process in multiple sclerosis. *Scand J Immunol*, 2011, 74(1), 1-13.
- [128] Janson, PC; Linton, LB; Bergman, EA; Marits, P; Eberhardson, M; Piehl, F; et al. Profiling of CD4+ T cells with epigenetic immune lineage analysis. *J Immunol*, 2011, 186(1), 92-102.
- [129] Bos, SD; Page, CM; Andreassen, BK; Elboudwarej, E; Gustavsen, MW; Briggs, F; et al. Genome-wide DNA methylation profiles indicate CD8+ T cell hypermethylation in multiple sclerosis. *PLoS One*, 2015, 10(3), e0117403.
- [130] Pedre, X; Mastronardi, F; Bruck, W; Lopez-Rodas, G; Kuhlmann, T; Casaccia, P. Changed histone acetylation patterns in normal-appearing white matter and early multiple sclerosis lesions. *J Neurosci*, 2011, 31(9), 3435-45.
- [131] Pedre, X; Mastronardi, F; Bruck, W; López-Rodas, G; Kuhlmann, T; Casaccia, P. Changed histone acetylation patterns in normal-appearing white matter and early multiple sclerosis lesions. *J Neurosci*, 2011, 31(9), 3435-45.
- [132] Du, C; Liu, C; Kang, J; Zhao, G; Ye, Z; Huang, S; et al. MicroRNA miR-326 regulates TH-17 differentiation and is associated with the pathogenesis of multiple sclerosis. *Nat Immunol*, 2009, 10(12), 1252-9.
- [133] Junker, A; Krumbholz, M; Eisele, S; Mohan, H; Augstein, F; Bittner, R; et al. MicroRNA profiling of multiple sclerosis lesions identifies modulators of the regulatory protein CD47. *Brain*, 2009, 132(Pt 12), 3342-52.
- [134] De Santis, G; Ferracin, M; Biondani, A; Caniatti, L; Rosaria Tola, M; Castellazzi, M; et al. Altered miRNA expression in T regulatory cells in course of multiple sclerosis. *J Neuroimmunol*, 2010, 226(1-2), 165-71.
- [135] Otaegui, D; Baranzini, SE; Armañanzas, R; Calvo, B; Muñoz-Culla, M; Khankhanian, P; et al. Differential micro RNA expression in PBMC from multiple sclerosis patients. *PLoS One*, 2009, 4(7), e6309.

- [136] Ponomarev, ED; Veremeyko, T; Barteneva, N; Krichevsky, AM; Weiner, HL. MicroRNA-124 promotes microglia quiescence and suppresses EAE by deactivating macrophages via the C/EBP-a-PU.1 pathway. *Nat Med*, 2011, 17(1), 64-70.
- [137] Ling, C; Groop, L. Epigenetics: a molecular link between environmental factors and type 2 diabetes. *Diabetes*, 2009, 58(12), 2718-25.
- [138] Gilbert, ER; Liu, D. Epigenetics: the missing link to understanding β -cell dysfunction in the pathogenesis of type 2 diabetes. *Epigenetics*, 2012, 7(8), 841-52.
- [139] Meda, F; Folci, M; Baccarelli, A; Selmi, C. The epigenetics of autoimmunity. *Cell Mol Immunol*, 2011, 8(3), 226-36.
- [140] Schalinske, KL. Interrelationship between diabetes and homocysteine metabolism: hormonal regulation of cystathionine beta-synthase. *Nutr Rev*, 2003, 61(4), 136-8.
- [141] McCarty, MF. Insulin secretion as a potential determinant of homocysteine levels. *Med Hypotheses*, 2000, 55(5), 454-5.
- [142] Arany, E; Strutt, B; Romanus, P; Remacle, C; Reusens, B; Hill, DJ. Taurine supplement in early life altered islet morphology, decreased insulinitis and delayed the onset of diabetes in non-obese diabetic mice. *Diabetologia*, 2004, 47(10), 1831-7.
- [143] Wang, Z; Zheng, Y; Hou, C; Yang, L; Li, X; Lin, J; et al. DNA methylation impairs TLR9 induced Foxp3 expression by attenuating IRF-7 binding activity in fulminant type 1 diabetes. *J Autoimmun*, 2013, 41(50-9).
- [144] Miao, F; Smith, DD; Zhang, L; Min, A; Feng, W; Natarajan, R. Lymphocytes from patients with type 1 diabetes display a distinct profile of chromatin histone H3 lysine 9 dimethylation: an epigenetic study in diabetes. *Diabetes*, 2008, 57(12), 3189-98.
- [145] Li, Y; Reddy, MA; Miao, F; Shanmugam, N; Yee, JK; Hawkins, D; et al. Role of the histone H3 lysine 4 methyltransferase, SET7/9, in the regulation of NF-kappaB-dependent inflammatory genes. Relevance to diabetes and inflammation. *J Biol Chem*, 2008, 283(39), 26771-81.
- [146] Reddy, MA; Natarajan, R. Epigenetic mechanisms in diabetic vascular complications. *Cardiovasc Res*, 2011, 90(3), 421-9.
- [147] Reddy, MA; Natarajan, R. Role of epigenetic mechanisms in the vascular complications of diabetes. *Subcell Biochem*, 2013, 61(435-54).
- [148] Roggli, E; Britan, A; Gattesco, S; Lin-Marq, N; Abderrahmani, A; Meda, P; et al. Involvement of microRNAs in the cytotoxic effects exerted by proinflammatory cytokines on pancreatic beta-cells. *Diabetes*, 2010, 59(4), 978-86.
- [149] Zheng, Y; Wang, Z; Tu, Y; Shen, H; Dai, Z; Lin, J; et al. miR-101a and miR-30b contribute to inflammatory cytokine-mediated beta-cell dysfunction. *Lab Invest*, 2015.
- [150] Dey, N; Das, F; Mariappan, MM; Mandal, CC; Ghosh-Choudhury, N; Kasinath, BS; et al. MicroRNA-21 orchestrates high glucose-induced signals to TOR complex 1, resulting in renal cell pathology in diabetes. *J Biol Chem*, 2011, 286(29), 25586-603.
- [151] Zhong, X; Chung, AC; Chen, HY; Dong, Y; Meng, XM; Li, R; et al. miR-21 is a key therapeutic target for renal injury in a mouse model of type 2 diabetes. *Diabetologia*, 2013, 56(3), 663-74.
- [152] Maekita, T; Nakazawa, K; Mihara, M; Nakajima, T; Yanaoka, K; Iguchi, M; et al. High levels of aberrant DNA methylation in *Helicobacter pylori*-infected gastric mucosae and its possible association with gastric cancer risk. *Clin Cancer Res*, 2006, 12(3 Pt 1), 989-95.

- [153] Bierne, H; Hamon, M; Cossart, P. Epigenetics and bacterial infections. *Cold Spring Harb Perspect Med*, 2012, 2(12), a010272.
- [154] Hur, K; Niwa, T; Toyoda, T; Tsukamoto, T; Tatematsu, M; Yang, HK; et al. Insufficient role of cell proliferation in aberrant DNA methylation induction and involvement of specific types of inflammation. *Carcinogenesis*, 2011, 32(1), 35-41.
- [155] Hahn, MA; Hahn, T; Lee, DH; Esworthy, RS; Kim, BW; Riggs, AD; et al. Methylation of polycomb target genes in intestinal cancer is mediated by inflammation. *Cancer Res*, 2008, 68(24), 10280-9.
- [156] Tolg, C; Sabha, N; Cortese, R; Panchal, T; Ahsan, A; Soliman, A; et al. Uropathogenic *E. coli* infection provokes epigenetic downregulation of CDKN2A (p16INK4A) in uroepithelial cells. *Lab Invest*, 2011, 91(6), 825-36.
- [157] Bobetsis, YA; Barros, SP; Lin, DM; Weidman, JR; Dolinoy, DC; Jirtle, RL; et al. Bacterial infection promotes DNA hypermethylation. *J Dent Res*, 2007, 86(2), 169-74.
- [158] Haller, D; Holt, L; Kim, SC; Schwabe, RF; Sartor, RB; Jobin, C. Transforming growth factor-beta 1 inhibits non-pathogenic Gram negative bacteria-induced NF-kappa B recruitment to the interleukin-6 gene promoter in intestinal epithelial cells through modulation of histone acetylation. *J Biol Chem*, 2003, 278(26), 23851-60.
- [159] Bardwell, AJ; Abdollahi, M; Bardwell, L. Anthrax lethal factor-cleavage products of MAPK (mitogen-activated protein kinase) kinases exhibit reduced binding to their cognate MAPKs. *Biochem J*, 2004, 378(Pt 2), 569-77.
- [160] Pathak, SK; Basu, S; Bhattacharyya, A; Pathak, S; Banerjee, A; Basu, J; et al. TLR4-dependent NF-kappaB activation and mitogen- and stress-activated protein kinase 1-triggered phosphorylation events are central to *Helicobacter pylori* peptidyl prolyl cis-, trans-isomerase (HP0175)-mediated induction of IL-6 release from macrophages. *J Immunol*, 2006, 177(11), 7950-8.
- [161] Xia, G; Schneider-Stock, R; Diestel, A; Habold, C; Krueger, S; Roessner, A; et al. *Helicobacter pylori* regulates p21(WAF1) by histone H4 acetylation. *Biochem Biophys Res Commun*, 2008, 369(2), 526-31.
- [162] Hamon, MA; Batsché, E; Régnault, B; Tham, TN; Seveau, S; Muchardt, C; et al. Histone modifications induced by a family of bacterial toxins. *Proc Natl Acad Sci USA*, 2007, 104(33), 13467-72.
- [163] Hamon, MA; Cossart, P. K⁺ efflux is required for histone H3 dephosphorylation by *Listeria monocytogenes* listeriolysin O and other pore-forming toxins. *Infect Immun*, 2011, 79(7), 2839-46.
- [164] Riggs, MG; Whittaker, RG; Neumann, JR; Ingram, VM. n-Butyrate causes histone modification in HeLa and Friend erythroleukaemia cells. *Nature*, 1977, 268(5619), 462-4.
- [165] Imai, K; Ochiai, K. Role of histone modification on transcriptional regulation and HIV-1 gene expression: possible mechanisms of periodontal diseases in AIDS progression. *J Oral Sci*, 2011, 53(1), 1-13.
- [166] Licciardi, PV; Wong, SS; Tang, ML; Karagiannis, TC. Epigenome targeting by probiotic metabolites. *Gut Pathog*, 2010, 2(1), 24.
- [167] O'Connell, RM; Rao, DS; Baltimore, D. microRNA regulation of inflammatory responses. *Annu Rev Immunol*, 2012, 30(295-312).

- [168] Fassi Fehri, L; Koch, M; Belogolova, E; Khalil, H; Bolz, C; Kalali, B; et al. Helicobacter pylori induces miR-155 in T cells in a cAMP-Foxp3-dependent manner. *PLoS One*, 2010, 5(3), e9500.
- [169] Zeng, FR; Tang, LJ; He, Y; Garcia, RC. An update on the role of miRNA-155 in pathogenic microbial infections. *Microbes Infect*, 2015, 17(9), 613-21.
- [170] Hanahan, D; Weinberg, RA. Hallmarks of cancer: the next generation. *Cell*, 2011, 144(5), 646-74.
- [171] Beck, S; Bernstein, BE; Campbell, RM; Costello, JF; Dhanak, D; Ecker, JR; et al. A blueprint for an international cancer epigenome consortium. A report from the AACR Cancer Epigenome Task Force. *Cancer Res*, 2012, 72(24), 6319-24.
- [172] Colotta, F; Allavena, P; Sica, A; Garlanda, C; Mantovani, A. Cancer-related inflammation, the seventh hallmark of cancer: links to genetic instability. *Carcinogenesis*, 2009, 30(7), 1073-81.
- [173] Sharma, SV; Lee, DY; Li, B; Quinlan, MP; Takahashi, F; Maheswaran, S; et al. A chromatin-mediated reversible drug-tolerant state in cancer cell subpopulations. *Cell*, 2010, 141(1), 69-80.
- [174] You, JS; Jones, PA. Cancer genetics and epigenetics: two sides of the same coin? *Cancer Cell*, 2012, 22(1), 9-20.
- [175] Philip, M; Rowley, DA; Schreiber, H. Inflammation as a tumor promoter in cancer induction. *Semin Cancer Biol*, 2004, 14(6), 433-9.
- [176] Kundu, JK; Surh, YJ. Inflammation: gearing the journey to cancer. *Mutat Res*, 2008, 659(1-2), 15-30.
- [177] Pal, S; Bhattacharjee, A; Ali, A; Mandal, NC; Mandal, SC; Pal, M. Chronic inflammation and cancer: potential chemoprevention through nuclear factor kappa B and p53 mutual antagonism. *J Inflamm (Lond)*, 2014, 11(23).
- [178] Söderlund, S; Brandt, L; Lapidus, A; Karlén, P; Broström, O; Löfberg, R; et al. Decreasing time-trends of colorectal cancer in a large cohort of patients with inflammatory bowel disease. *Gastroenterology*, 2009, 136(5), 1561-7, quiz 1818-9.
- [179] Eaden, J; Abrams, K; Ekbom, A; Jackson, E; Mayberry, J. Colorectal cancer prevention in ulcerative colitis: a case-control study. *Aliment Pharmacol Ther*, 2000, 14(2), 145-53.
- [180] Eaden, JA; Mayberry, JF. Colorectal cancer complicating ulcerative colitis: a review. *Am J Gastroenterol*, 2000, 95(10), 2710-9.
- [181] Kundu, JK; Surh, YJ. Emerging avenues linking inflammation and cancer. *Free Radic Biol Med*, 2012, 52(9), 2013-37.
- [182] Werling, K; Szepesi, A; Szentirmay, Z; Schaff, Z; Tulassay, Z; Szalay, F. Effect of hepatitis C virus on hepatocyte proliferation and DNA ploidy in patients with chronic hepatitis C. *Z Gastroenterol*, 2000, 38(7), 553-4, 556-8.
- [183] Rajalingam, K; Schreck, R; Rapp, UR; Albert, S. Ras oncogenes and their downstream targets. *Biochim Biophys Acta*, 2007, 1773(8), 1177-95.
- [184] Strano, S; Dell'Orso, S; Di Agostino, S; Fontemaggi, G; Sacchi, A; Blandino, G. Mutant p53: an oncogenic transcription factor. *Oncogene*, 2007, 26(15), 2212-9.
- [185] Federico, A; Morgillo, F; Tuccillo, C; Ciardiello, F; Loguercio, C. Chronic inflammation and oxidative stress in human carcinogenesis. *Int J Cancer*, 2007, 121(11), 2381-6.
- [186] Marnett, LJ. Oxyradicals and DNA damage. *Carcinogenesis*, 2000, 21(3), 361-70.

- [187] Halliday, GM. Inflammation, gene mutation and photoimmunosuppression in response to UVR-induced oxidative damage contributes to photocarcinogenesis. *Mutat Res*, 2005, 571(1-2), 107-20.
- [188] Xu, H; Chaturvedi, R; Cheng, Y; Bussiere, FI; Asim, M; Yao, MD; et al. Spermine oxidation induced by *Helicobacter pylori* results in apoptosis and DNA damage: implications for gastric carcinogenesis. *Cancer Res*, 2004, 64(23), 8521-5.
- [189] Babbar, N; Casero, RA. Tumor necrosis factor- α increases reactive oxygen species by inducing spermine oxidase in human lung epithelial cells: a potential mechanism for inflammation-induced carcinogenesis. *Cancer Res*, 2006, 66(23), 11125-30.
- [190] Bergman, Y; Cedar, H. DNA methylation dynamics in health and disease. *Nat Struct Mol Biol*, 2013, 20(3), 274-81.
- [191] Esteller, M. Epigenetics in cancer. *N Engl J Med*, 2008, 358(11), 1148-59.
- [192] Chan, AO; Lam, SK; Wong, BC; Kwong, YL; Rashid, A. Gene methylation in non-neoplastic mucosa of gastric cancer: age or *Helicobacter pylori* related? *Am J Pathol*, 2003, 163(1), 370-1, author reply 371-3.
- [193] Chan, AO; Lam, SK; Wong, BC; Wong, WM; Yuen, MF; Yeung, YH; et al. Promoter methylation of E-cadherin gene in gastric mucosa associated with *Helicobacter pylori* infection and in gastric cancer. *Gut*, 2003, 52(4), 502-6.
- [194] Kawasaki, H; Abe, H. in cancer and inflammation. *Personalized Medicine Universe*, 2012, 1(1), 7-12.
- [195] Hansen, KD; Timp, W; Bravo, HC; Sabuncuyan, S; Langmead, B; McDonald, OG; et al. Increased methylation variation in epigenetic domains across cancer types. *Nat Genet*, 2011, 43(8), 768-75.
- [196] Xie, W; Schultz, MD; Lister, R; Hou, Z; Rajagopal, N; Ray, P; et al. Epigenomic analysis of multilineage differentiation of human embryonic stem cells. *Cell*, 2013, 153(5), 1134-48.
- [197] Abu-Remaileh, M; Bender, S; Raddatz, G; Ansari, I; Cohen, D; Gutekunst, J; et al. Chronic inflammation induces a novel epigenetic program that is conserved in intestinal adenomas and in colorectal cancer. *Cancer Res*, 2015, 75(10), 2120-30.
- [198] Lawrence, T. The nuclear factor NF- κ B pathway in inflammation. *Cold Spring Harb Perspect Biol*, 2009, 1(6), a001651.
- [199] Valinluck, V; Sowers, LC. Inflammation-mediated cytosine damage: a mechanistic link between inflammation and the epigenetic alterations in human cancers. *Cancer Res*, 2007, 67(12), 5583-6.
- [200] Hodge, DR; Peng, B; Cherry, JC; Hurt, EM; Fox, SD; Kelley, JA; et al. Interleukin 6 supports the maintenance of p53 tumor suppressor gene promoter methylation. *Cancer Res*, 2005, 65(11), 4673-82.
- [201] Wehbe, H; Henson, R; Meng, F; Mize-Berge, J; Patel, T. Interleukin-6 contributes to growth in cholangiocarcinoma cells by aberrant promoter methylation and gene expression. *Cancer Res*, 2006, 66(21), 10517-24.
- [202] Isomoto, H; Mott, JL; Kobayashi, S; Werneburg, NW; Bronk, SF; Haan, S, et al. Sustained IL-6/STAT-3 signaling in cholangiocarcinoma cells due to SOCS-3 epigenetic silencing. *Gastroenterology*, 2007, 132(1), 384-96.
- [203] Gray, SG; The, BT. Histone acetylation/deacetylation and cancer: an "open" and "shut" case? *Curr Mol Med*, 2001, 1(4), 401-29.

- [204] Huang, L. Targeting histone deacetylases for the treatment of cancer and inflammatory diseases. *J Cell Physiol*, 2006, 209(3), 611-6.
- [205] Santos-Rosa, H; Caldas, C. Chromatin modifier enzymes, the histone code and cancer. *Eur J Cancer*, 2005, 41(16), 2381-402.
- [206] Yang, SH; Sharrocks, AD. SUMO promotes HDAC-mediated transcriptional repression. *Mol Cell*, 2004, 13(4), 611-7.
- [207] Xu, J; Zhang, X; Monestier, M; Esmon, NL; Esmon, CT. Extracellular histones are mediators of death through TLR2 and TLR4 in mouse fatal liver injury. *J Immunol*, 2011, 187(5), 2626-31.
- [208] Takagi, T; Naito, Y; Mizushima, K; Hirata, I; Yagi, N; Tomatsuri, N; et al. Increased expression of microRNA in the inflamed colonic mucosa of patients with active ulcerative colitis. *J Gastroenterol Hepatol*, 2010, 25 Suppl 1(S129-33).
- [209] Meng, F; Henson, R; Wehbe-Janek, H; Smith, H; Ueno, Y; Patel, T. The MicroRNA let-7a modulates interleukin-6-dependent STAT-3 survival signaling in malignant human cholangiocytes. *J Biol Chem*, 2007, 282(11), 8256-64.
- [210] Meng, F; Wehbe-Janek, H; Henson, R; Smith, H; Patel, T. Epigenetic regulation of microRNA-370 by interleukin-6 in malignant human cholangiocytes. *Oncogene*, 2008, 27(3), 378-86.

Chapter 66

A NEW MOLECULAR APPROACH TO OBESITY: EPIGENETIC PERSPECTIVE

Burcu Bayoglu*, PhD and Mujgan Cengiz, PhD

Istanbul University Cerrahpasa Medical Faculty,
Department of Medical Biology, Istanbul, Turkey

ABSTRACT

Obesity is a public health problem leading to morbidity and mortality throughout the world. It arises from the interactions between genetics and environmental factors. In recent years, susceptibility to obesity has been also linked to epigenetic factors. Epigenetics, is the study of heritable changes in gene expression which do not involve in the underlying DNA sequence. The epigenetic mechanisms include DNA methylation, covalent histone modifications, chromatin folding, miRNAs, and polycomb group repressive complexes. Both dietary factors and individual behaviors affect obesity development via epigenetic mechanisms. Epigenetic mechanisms are also linked to programmed changes in gene expression as a result of early environmental exposures during pregnancy which alter offspring growth and development. There is evidence that nutrient and environmental exposures during pregnancy may affect fetal/newborn development and result in offspring obesity or metabolic syndrome which is a cluster of metabolic abnormalities. Obesity related genes, epiobesigenes, display methylation patterns playing important roles in the development of obesity which are potential future epigenetic biomarkers of obesity. The susceptibility genes have been reported as FGF2, PTEN, CDKN1A, and ESR1, functional in adipogenesis; SOCS1 and SOCS3, functional in inflammation, and COX7A1, LPL, CAV1, and IGFBP3 which are functional in fat metabolism and insulin signaling. It is important to prevent this era's epidemic, obesity, since it leads to chronic diseases including hypertension, atherosclerosis, insulin resistance, and moreover metabolic syndrome. It may be easier to prevent the progress of the disease by revealing the epigenetic mechanisms especially methylation profiles of the susceptibility genes.

Keywords: obesity, adipogenesis, epigenetics, epiobesigenic genes, DNA methylation, histon modifications

* Corresponding Author's Email: burcubayoglu@yahoo.com.

ABBREVIATIONS

AcH3	histone acetylation
ADIPOQ	adiponectin
Avy	agouti
BMI	body mass index
CASP8	caspase 8
CAV-1	caveolin-1
CDKN1A	cyclin-dependent kinase inhibitor 1A
CEBPA	CCAAT/enhancer binding protein (C/EBP) a
CG	Cytosine-Guanine
COX7A1	cytochrome c oxidase subunit VIIa polypeptide 1
DNA	deoxyribonucleic acid
DNMTs	DNA methyltransferases
ESR1	Oestrogen receptor-alpha
FABP4	fatty acid-binding protein 4, adipocyte
FASN	fatty acid synthase
FGF2	fibroblast growth factor-2
FTO	fat mass and obesity associated
GLUT4	insulin-responsive glucose transporter 4
H2K4me3	trimethylation of histone H3 at lysine 4
H3K4	histone lysine methylation
H3K9	histone H3 lysine 9
H3K9me2	dimethylation of histone H3 at lysine 9
HDAC	histone deacetylase
HIF1A	hypoxia inducible factor 1
HMT	histone methylase
IFNG	Interferon γ
IGF2	insulin-like growth factor 2
IGFBP3	insulin-like growth factor-binding protein 3
INS	insulin
INSR	insulin receptor
H3-K27	lysine 27 of histone
H3	LEP, leptin
LPL	lipoprotein lipase
MC4R	melanocortin 4 receptor
miRNA	microRNA
ncRNAs	non-coding RNAs
NR0B2	nuclear receptor small heterodimer partner fibroblast growth factor-2
NPY	neuropeptide Y
NR3C1	glucocorticoid receptor; nts, nucleotides
PBMC	peripheral blood mononuclear cell
PcG	polycomb group
PcG	polycomb group proteins
PEPCK	phosphoenolpyruvate carboxykinase

PIK3CG	phosphoinositide-3-kinase, catalytic, gamma pp
PLA2G4A	plasma secretory phospholipase A(2) type IIA
POMC	proopiomelanocortin
PPARA	peroxisome proliferator-activated receptor α
PPAR γ	peroxisome proliferator-activated receptor gamma
PPARGC1A	peroxisome proliferator-activated receptor gamma-coactivator-1 α PTEN phosphatase and tensin homologue
PWS	Prader Willi Syndrome
PRC2	polycomb repressive complex 2
RARA	retinoic acid receptor- α
SOCS1	suppressor of cytokine signalling 1
SOCS3	suppressor of cytokine signalling 3
SOD3	extracellular superoxide dismutase
SSTR2	somatostatin receptor 2
HSD11B2	11 beta-hydroxysteroid dehydrogenase 2
T2D	type 2 diabetes
TNF	tumor necrosis factor
UCP1	uncoupling protein 1

1. OBESITY AND ITS MOLECULAR PATHOGENESIS

Obesity is a major health problem leading to morbidity and mortality throughout the world [1]. Obesity has a multifactorial etiology which involves interactions between genetic background, hormones and environmental factors [2]. However, the etiology of obesity is not well understood. Obesity also plays a crucial role for developing metabolic syndrome which includes insulin resistance, hypertension, elevated triglycerides and atherosclerosis [3]. Metabolic syndrome is a growing cause of morbidity and mortality worldwide. It is characterized by the cluster of metabolic disturbances including obesity, hyperlipidemia, hypertension, and elevated fasting blood sugar. The risk for metabolic syndrome has largely been attributed to adult lifestyle factors such as poor nutrition, lack of exercise, and smoking, however; there is now strong evidence suggesting that predisposition to metabolic syndrome development begins in *utero* [4]. Obesity is related with both type 2 diabetes (T2D) and metabolic syndrome development. Other than individual life style choices, there is evidence that developing obesity is in part due to genetic disposition and epigenetic alterations [5].

Body mass index (BMI) is the commonly used measure of weight status. BMI is a calculation of the ratio of someone's height and weight ($BMI = \text{kg/m}^2$).

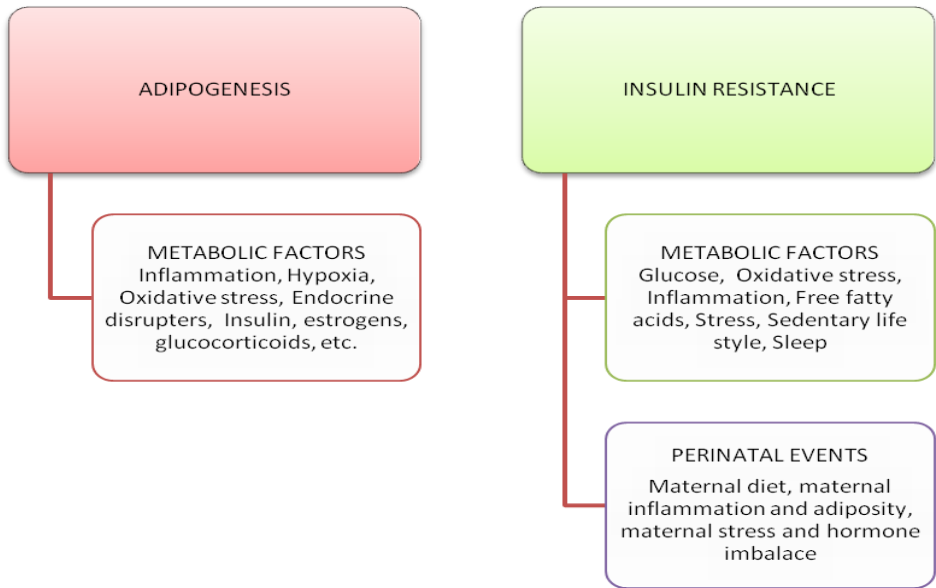
Research have shown that, BMI is a good estimate of fatness and correlates with chronic diseases including heart disease, diabetes, cancer and overall mortality. For adult men and women, a BMI between 18.5 and 24.9 is considered healthy. BMI between 25.0 and 29.9 is defined as overweight and a BMI of 30 or higher is considered obese [6].

When energy intake exceeds energy expenditure, the excess energy is often stored as body fat and cause obesity. In recent years, it is suggested that the availability of energy-dense food, environmental changes create an obesogenic environment. Since obesity is multifactorial, an individual can become obese in the absence of an obesogenic environment or on the contrary,

an individual does not become obese in an obesogenic environment. Various factors causing obesity can be endocrine disrupters, which are the chemicals that mimic the effect of hormones in the body, and also reproductive factors including intrauterine effects, epigenetics and maternal age [7]. An obese phenotype has occurred among people in a short period of time, just one or two generations.

This shows that environmental or epigenetic factors play a major role rather than genetic mechanisms in the etiology of obesity. Although, high fat diet with decreased physical activity seems to be the strong contributor of the prevalence of obesity, some studies support that obesity may have its origin in *utero* [8-12]. Obesity and the metabolic abnormalities related with it, are developmentally programmed since, although normal postnatal nutrition is carried out, obesity exist in offspring because of the suboptimal nutrition in *utero* [13]. Similarly to type 2 diabetes, hypertension, atherosclerosis and other metabolic disorders, predisposition to obesity, and weight loss have been associated with epigenetic pattern alterations. Nutritional factors and also different non-nutritional risk factors accompanying obesity are involved in the epigenetic modifications, and they affect adipogenesis, insulin sensitivity, especially hyperglycemia, inflammation, endocrine disruptors, hypoxia, and oxidative stress [14], (Figure 1).

In recent years, the epigenetic regulation of the gene expression in obesity seems to be a potential considerable contributor. Epigenetics is defined as the study of heritable gene expression changes that occur in the absence of a change in the DNA sequence itself [15]. In other words, epigenetics is the study of gene expression's molecular control that is not caused by DNA sequence.



Modified from 14.

Figure 1. Metabolic and perinatal factors related to adipogenesis, insulin resistance.

Epigenetic modifications include methylation, covalent modifications of histones, DNA packaging around nucleosomes, chromatin folding and chromatin attachment to the nuclear matrix [16].

Epigenetics may help to understand genetic fetal programming, monozygotic twin differences, and chronic disease onset in adults, which interact with dietary intake and nutritional processes [14]. For instance, DNA sequence methylation usually reduces gene expression. Epigenetic factors may be passed on mitotically, through cell division, or meiotically, which is called transgenerational inheritance.

Thus, epigenetic programming of parental or maternal gametes, the fetus, or early postnatal development may be influenced by environmental factors [7] (Figure 2).

2. EPIGENETICS: BASIC MECHANISMS

DNA methylation, histone-tail acetylation, poly-ADP-ribosylation and ATP-dependent chromatin-remodeling are all included in chromatin-remodeling mechanisms [16]. DNA methylation is the covalent attachment of a methyl group onto the Cytosine residues' C5 position on CpG islands. CpG islands, are the genomic regions containing high number of Cytosine-Guanine (CG) dinucleotides. In the methylation of CpG islands, cytosine is converted to 5-methylcytosine and since CpG islands are mostly located on the promoter region, methylation process is often associated with gene silencing [17] (Figure 3). DNA methyltransferases (DNMTs) catalyze the addition of the methyl group to the DNA. Three major enzymes function in maintaining DNA methylation patterns.

- 1- DNA methyltransferases 3A (DNMT3A)
- 2- DNA methyltransferases 3B (DNMT3B)
- 3- DNA methyltransferases 1 (DNMT1)

DNMT3A and 3B are *de novo* methyltransferases, DNMT1 controls the methylation patterns they are copied in each cell division [18]. DNMTs interact with histone deacetylases, histone methyltransferases and methyl-cytosine-binding proteins to establish and maintain DNA methylation patterns in a complex regulatory pathway in the cell. In histone modifications, methylation of histone H3 lysine 9 (H3K9) means condensed and inactive chromatin whereas acetylation is related with active gene expression [17]. DNA methylation typically occurs during DNA replication, so that methylation patterns maintain in cell division via hemimethylated DNA. However during replication and embryogenesis, *de novo* DNA methylation can also occur [18].

To date four current models have been discussed for how DNA methylation can mediate gene silencing [19]. In the first model, DNA methylation can prevent the binding of transcriptional activator to the target DNA sequence and this directly inhibits transactivation [20]. The second model explains that the DNMT protein may link to histone deacetylase (HDAC) and histone methylase (HMT) proteins which allows the coupling of enzymatic activities [21].

Next, DNA methylation within the gene exhibits an effect of repression on transcriptional elongation. Methylation occurs generally in the promoter region, however it can occur downstream the gene too [22]. Finally, methylated DNA can be recognized directly by methyl-CpG-binding proteins. And these proteins can recruit transcriptional repressors to silence the surrounding chromatin [23, 24].

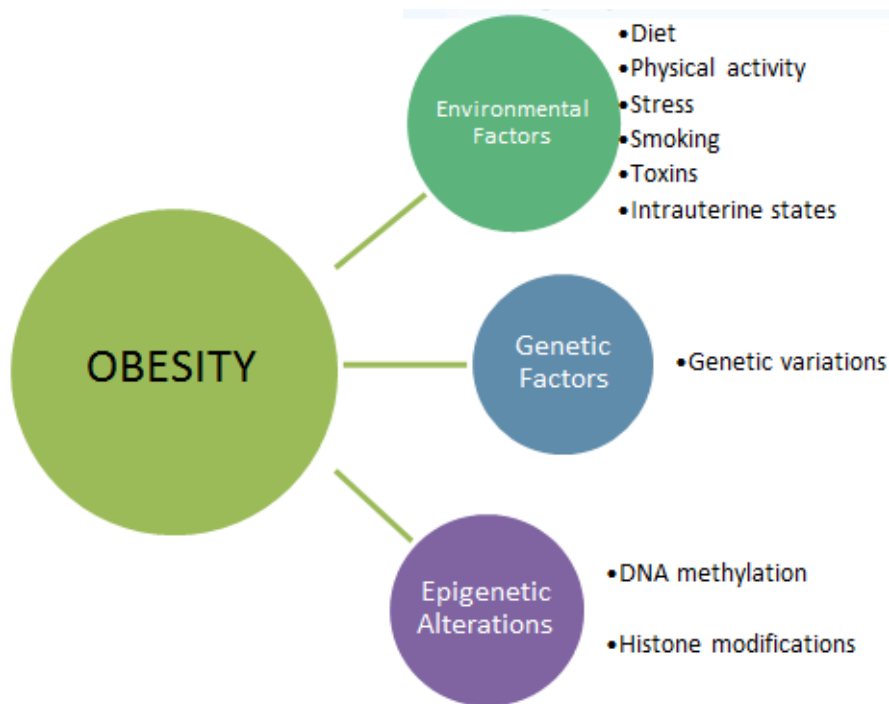


Figure 2. The interplay between environmental, genetic and epigenetic factors in the etiology of obesity.

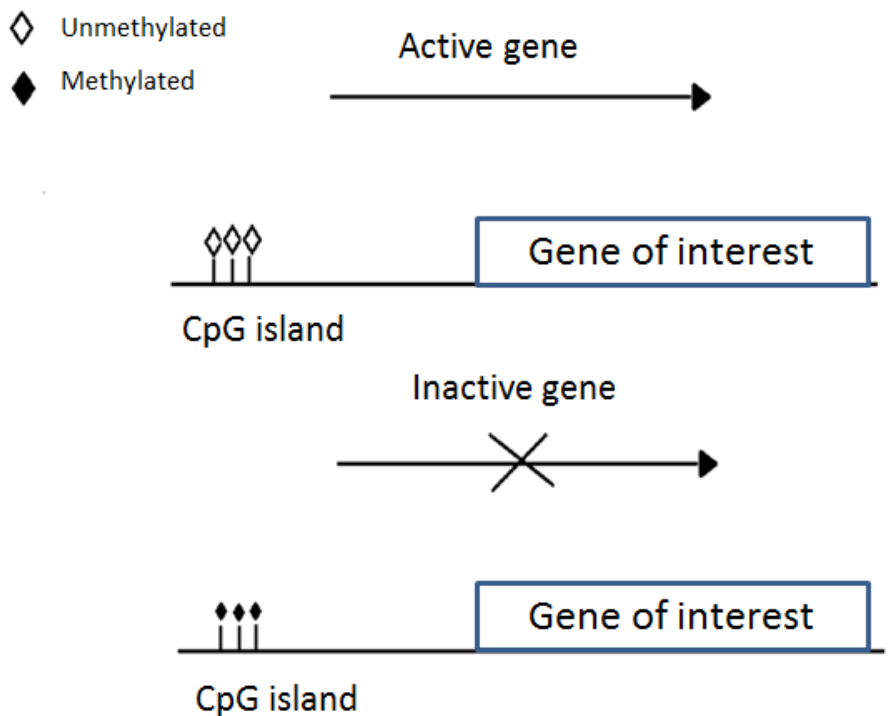


Figure 3. DNA methylation of CpG islands.

There are also other mechanisms related with epigenetic pathways. One of them is chromatin packaging into open and closed states. The open state of the chromatin is called the euchromatin and the closed state is called the heterochromatin. DNA is packaged around histones and forms the chromatin. Histone tails are translationally modified with methylation/demethylation and acetylation/deacetylation mechanisms by histone modifying enzymes. Histone methylation may activate or repress transcription, depending on which lysine residue is methylated. For instance, trimethylation of histone H3 at lysine 4 (H3K4me3) means gene transcription is active, however dimethylation of histone H3 at lysine 9 (H3K9me2) means transcriptional silencing. Unlike methylation, acetylation activates, deacetylation represses gene expression [13, 25, 26].

An additional mechanism for the regulation of gene expression and silencing is polycomb group (PcG) complexes. PcG complexes are associated with remodeling of chromatin; for example they get involved in histone H3K27 methylation or H2AK119 ubiquitination. In this mechanism, chromatin is epigenetically repressed by these protein complexes. They are recruited to the promoter region of gene of interest and prevent transcription factors to bind the promoter of DNA sequences [27].

In the epigenetic regulation of gene expression, non-coding RNAs (ncRNAs) also take place. ncRNAs are transcribed from DNA, but they are not translated to proteins. Their main function is regulation in gene expression at the transcriptional and post-transcriptional states. ncRNAs are classified into short ncRNAs (<30 nts) and the long ncRNAs (>200 nts). Three major short ncRNAs participate in gene silencing; microRNAs (miRNAs), short inhibitory RNAs, and piwi-interacting RNAs [28, 29]. Long ncRNAs play a regulatory role during development [30]. They are expressed cell type-specific [31, 32]. Long ncRNAs are also found to be associated with adipogenesis [33, 34]. Recently, it is suggested that ncRNAs do not just regulate gene expression in translational level but also they take part in DNA methylation and histone modifications, and thus regulating the transcription of their target genes [35, 36]. Epigenetic gene silencing is also related with the microRNAs (miRNAs). miRNAs are small ncRNA genes composed of 22-nucleotides. They are specific gene regulators and control the gene expression [37]. miRNAs are transcriptionally regulated by DNA methylation and histone modifications. And so, they have significant roles in targeting key enzymes which control chromatin structure and histone modifications [17].

Epigenome is the term covering overall epigenetic state of a cell. During gametogenesis and between fertilization and blastocyst formation, epigenome may be susceptible to dysregulate [38], however during gestation, lactation and throughout life course DNA methylation patterns are maintained [39]. DNA methylation patterns change during aging process, thus a good working epigenetic homeostatic mechanism must be activated to prevent epigenetic abnormalities [40]. Otherwise, epigenetic lesions would accumulate to develop nutrition-related disorders like obesity [39]. Epigenomics is the study of genome-wide epigenetic modifications. Epigenetics is the study trying to understand relations between genes and environmental factors such as diet, inactivity, smoking, and etc. to generate a phenotype [41]. Epigenome is dynamic, since it is changed in response to nutrition, physical exercise, weight loss and aging [42-45]. In *utero* nutrition, environmental exposures, maternal or fetal stress may change the fetus's gene expression via epigenetic mechanisms, thus change the cells' and organs' structure and function leading to metabolic anomalies [13].

3. OBESITY: FROM AN EPIGENETIC PERSPECTIVE

Obesity is a multifactorial disorder involving complex interactions among genetic factors, hormones, and some environmental factors including sedentary life style and inadequate dietary habits [2]. Obesity has been associated with lifestyle and accordingly, studies have investigated thermoenergetic, life style and balance between energy intake and expenditure. Nowadays, genetic studies and molecular pathways are being investigated in obesity pathogenesis. These studies include gene mutations and polymorphisms, candidate gene approach etc [46]. Also, in recent years epigenetic regulation of gene expression has been related with obesity pathogenesis as a potential significant contributor.

There are largely animal and limited human studies reporting the association between epigenetic mechanisms and obesity. Tissue gene expression can be changed via epigenetic regulation in accordance with environmental conditions throughout life. Nutritional and environmental factors may interact with the genotype and modulate epigenetic marks in somatic cells and affect the phenotype. This is shown in monozygous twins and in *agouti* mouse.

In monozygous twins, the offspring have divergent DNA methylation and histone acetylation patterns. Lymphocyte and muscle biopsy samples were taken from 3-74 years old Spanish monozygotic twins. They are analyzed for large-scaled and locus-specific DNA methylation and histone acetylation. Older twins showed differences in their epigenetic marks and gene expression profiles; however younger twins showed no differences, suggesting that there is environmental effect on gene expression regulation, despite a similar genotype [44].

Animal obesity model of *agouti* mouse with the *agouti* viable yellow (*Avy*) mutation is used for obesity studies and this model can be modulated through an epigenetic mechanism [47]. In *agouti* mouse, the diet affects methylation status and the animal's coat color [48, 49]. In *agouti* mice, when the *agouti* gene is normally methylated, the mice are thin and have brown color. However, when the *agouti* gene completely unmethylated, the mice are obese and diabetic and also they have yellow coat color. This mutation causes the ectopic expression of the *agouti* protein. As a result of this mutation, the *agouti* protein binds to the melanocortin 4 receptor in the hypothalamus antagonistically and leads to hypothalamic dysfunction and induces hyperphagic obesity. It is known that, methylation controls gene expression, therefore, obesity phenotype depends on methylation [47]. Body weight can be affected by epigenetic regulation with dietary factors. In *agouti* mice, maternal diet with methyl supplementation affects the methylation of the *Avy* allele and offspring obesity [49]. The epigenetic regulation of hedonic reward pathways and metabolic regulation of energy balance may be changed with maternal and post-weaning high fat diet in mice and results in altering methylation of the leptin promoter in rats [50-53]. Yellow fat mice and brown thin mice have the same genotype but they have different phenotypes as a result of epigenetic alterations. These methylation patterns can also be transmitted across generations [48, 49]. Interestingly, when yellow fat mice were fed with methyl-rich diet during pregnancy, they produced offspring with brown coat without obesity. This result showed the interaction between nutrition and epigenetic modifications influencing the phenotype [54].

Epigenetic modulation of imprinted genes is also well known, for example the impact of insulin-like growth factor 2 (*IGF2*) on body weight and obesity is well studied [55]. By several autocrine, paracrine and endocrine hormonal factors, adipose tissue growth can be induced. For

the increase in adipose tissue, IGF2 expression is increased methylation of differentially methylated regions in the H19 imprinting control region. Therefore, studies on the IGF2/H19 locus showed that hypomethylated CpG islands in the 5' region of the H19 gene are critical binding sites for the 11-zinc finger protein CTCF [56]. Activation of IGF2 gene transcription is regulated via maternal diets such as low dietary protein and folate through increased methylation of ICR/H19 DMR [57]. Also, another mechanism for elevated IGF2 expression is acetylation of histone 4 (H4) at lysines 16 and 8 [58]. In this mechanism, from polycomb group proteins (PcG), polycomb repressive complex 2 (PRC2) is recruited through the interaction of CTCF with Suz12 PRC2 subunit and leads to allele-specific methylation at lysine 27 of histone H3 (H3-K27) and represses the maternal IGF2 promoters [59].

Epigenetic alterations may also occur in nonimprinted genes involved in energy metabolism. Genes functioning in adipogenesis, glucose homeostasis, inflammation and/or insulin signaling are regulated by epigenetic mechanisms, for example hormone encoding genes like, leptin; nuclear receptors like adipogenic and lipogenic transcription factors, PPAR γ and PPAR α , respectively; gluconeogenic enzymes, phosphoenolpyruvate carboxykinase (PEPCK); and transmembrane proteins, uncoupling protein 1 [53, 60-67]. For example, leptin plasma levels can be modulated by diet with epigenetic mechanisms. It is also suggested that leptin methylation affects leptin gene expression in rats since high-fat diet in rats increased the leptin gene promoter methylation in retroperitoneal adipocytes and this was related with lower circulating leptin levels [53]. Also, adipocyte differentiation and the induction of adipogenic transcription factors via epigenetic mechanisms, such as methylation of PPAR γ and C/EBP α , may drive adipogenesis. Adipocyte differentiation is also regulated by both histone lysine methylation (H3K4) and acetylation (AcH3). Increased histone trimethylation (H3K4me3) and acetylation (AcH3K9/K14) are also consistent with the PPAR γ and C/EBP α upregulation during terminal adipocyte differentiation. When H3K4me3 and AcH3K9/K14 increases, adipogenic genes may be induced contributing to early adipocyte differentiation and obesity [68-70].

Hepatic PEPCK is the rate-limiting enzyme in the metabolic pathway of gluconeogenesis. High-calorie diet causes hypomethylation of the *PEPCK* gene in rats. And this was reported to be associated with the increased *PEPCK* mRNA levels, which shows increased liver gluconeogenesis [71]. Methylation alterations were found to be associated with changes in target gene expression and glucose and lipid metabolism [13].

Recently, DNA methylation and specific histone methylation (H3K4 and H3K9) and some miRNAs were found to be related with BMI [72-74]. In a study performed with genome wide methylation analysis, human obesity was reported to be associated with methylation changes in blood leukocyte DNA [72]. Another study related with epigenetic modifications was performed in overweight individuals with and without type 2 diabetes. Histone methylation of K4 and K9 in primary human adipocytes of overweight individuals emerged 40% lower levels of K4 dimethylation in overweight nondiabetic individuals. Also, K4 trimethylation was reported 40% higher in adipocytes of overweight diabetic subjects when compared with normal-weight and overweight non-diabetic subjects. In the same study, obese and lean individuals showed varying DNA methylation levels in specific genes. Also, in cultured preadipocytes and mature adipocytes miRNA expression profiles showed different expressions [73]. Similarly, in human subcutaneous adipose tissue biopsies, differential and dysregulated miRNA expression was observed in obese and lean individuals of genome wide miRNA

profiling. In human adipose tissue, miRNA expression pattern was associated with obesity and glucose metabolism parameters [74].

In the development of obesity, animal studies were also performed to reveal the role of the epigenome. But this relation was more specific to histone modifications and miRNAs rather than DNA methylation [65, 75-78]. For example, the main adipogenic transcription factor PPAR and its target genes regulation have been associated with histone modifications (methylation and acetylation) and miRNAs, and leads to obesity, hyperphagia, and hyperlipidemia [75, 79-81]. In another animal study, maternal obesity, high-fat diet in mice were shown to increase the key transcription factor, Zfp423, which is responsible for adipogenic lineage commitment during fetal development. The repressive histone methylation, H3K27me3, was observed lower in Zfp423 promoter of fetal tissues from maternal obese mice [82].

Increased mRNA expression and CpG site/ island DNA methylation changes of pro-adipogenic factors (Zfp423 and C/EBP- β) have been shown in 3 week old rat offspring from obese pregnancies [83]. On the other hand, mice with genetically increased or decreased levels of DNA methyltransferases, DNMT1 and DNMT3b, do not gain weight or have increased adiposity. In the same study, *de novo* DNMT3a was more than doubled in white adipose tissue of obese mice, and a transgenic mouse having threefold increased *DNMT3a* mRNA levels in adipose tissue, did not show increased body weight or adiposity [84]. In that study, it was argued that, perhaps it is histone deacetylases which was functional in tissue specific expression rather than DNA methyltransferases. However, specific histone methyltransferases play a significant role in adipogenesis; mice with mutations in the histone H3, lysine 4 methyltransferase *MLL3* show decreased adipogenesis and they are protected from high-fat diet induced obesity [85]. In contrast, mice with loss of function of the H3K9-specific demethylase JmjC domain with histone demethylase 2A are obese and hyperlipidemic [75, 76]. Thus, it has been shown that mutations in epigenetic-modifiers may lead to obesity [13].

3.1. Programmed Obesity and the Epigenome

During, gametogenesis and early embryogenesis, epigenetic modifications are not completely erased. And some memory of the epigenetic state transfers to the offspring [86, 87]. This is called the transgenerational epigenetic inheritance [48]. Also, mother's nutrition or perinatal lifestyle choices in *utero* or lactation periods, may change the fetus's developmental program and may lead to obesity development in the future [88]. However the role of the nutrition on epigenetic pattern modifications during adulthood and its transfer to the gametes is still unknown [89]. So the effects of the dietary and environmental factors on DNA methylation patterns can lead to the susceptibility to develop obesity and related metabolic diseases [39].

Environmental exposures during early life may induce offspring's epigenome changes. This may lead to increased risk of obesity and metabolic syndrome in later life. Studies showed that programmed obesity in humans occurs via DNA methylation changes. Histone modifications and chromatin structure changes has not been shown yet in programmed obesity. Exposure to maternal famine or obesity decreases the DNA methylation of the imprinted *IGF2* gene in the offspring. 60-year-old adults, prenatally exposed to famine, in the Dutch Famine (1944–1945) cohort, exhibit hypomethylation of whole blood *IGF2* gene, and

hypermethylation of two obesity-related non-imprinted genes, *TNF* and leptin, compared to their unexposed, same-sex siblings. This association showed early developmental period is crucial for establishing and maintaining epigenetic marks [90, 91]. Recently, paternal obesity was found to be associated with *IGF2* hypomethylation in umbilical cord blood leukocytes of newborns [92, 93]. Also, specific maternal features like gestational weight gain and gestational diabetes have been related with signature DNA methylation in cord blood and elevated placental leptin gene methylation, respectively [42, 94]. Also, in early pregnancy maternal carbohydrate intake was related with the umbilical cord methylation levels of retinoid X receptor- α which is a nuclear receptor gene and make heterodimers with adipogenic transcription factor PPAR γ , and also related with adiposity in later childhood [95]. Animal studies show that both DNA methylation and histone modification changes are related with developmental programming of obesity in rodent models of maternal under or overnutrition; for example, high-fat diet, maternal obesity [41, 96-99]. Genes that regulate growth factors [100], adipogenesis [101], brain appetite and satiety/reward pathways [51, 102-104] and glucose homeostasis [105] have been detected to be regulated by epigenetic changes. Human and animal studies provide evidence of programmed obesity and metabolic syndrome resulting from early nutritional exposures and suggesting epigenetic modification of nonimprinted genes may be a major contributing factor [13].

3.2. Epiobesigenic Genes

In the onset, development and therapy of type 2 diabetes and obesity, epigenetic research became to have advances [16]. Some gene promoters are susceptible to epigenetic regulation and may have potential role in the development of obesity. These genes are called epiobesigenic genes [39]. Some human genes regulated by epigenetic mechanisms related with obesity are shown in Table 1.

In human obesity, many genes have been identified. These gene effects' may be modified by epigenetic mechanisms [144]. The most known example of obesity is the Prader Willi Syndrome (PWS) caused by an epigenetic mechanism in human. PWS occurs due to a lack of expression from parental chromosome 15q11-q13. This lack of expression arise from the epigenetic silencing on the maternal and paternal copy of the genes, instead of just the maternal copy, in 25% of cases. PWS is characterized with hyperphagia, infertility, and obesity [145], however obese phenotype is not passed on, due to infertility [7].

Leptin and proopiomelanocortin gene mutations also cause human obesity, and these genes have both CpG islands in which their expression can be controlled through methylation [114]. Some of these susceptibility genes are associated with various metabolic diseases like hypertension (*HSD11B2*) [134], atherosclerosis (*PPARG*) [61], diabetes (*PPARGC1*) and endotoxin tolerance [66].

Also, some of these genes may play a role in obesity since they affect adipogenesis process [61, 107-116, 134], inflammation [66, 114, 123-126, 128] and insulin signaling [66, 131-135]. When unmethylated CpG islands are hypermethylated, this means that transcriptional activation is repressed. The presence of CpG islands in the promoter region of a gene shows that, the gene might be partly regulated by CpG methylation [134]. Normal and pathologic gene expressions are being studied with methylation pattern mapping of CpG islands.

Table 1. Some human genes regulated by epigenetic mechanisms (Epiobesigenes) involved in obesity pathogenesis (Modified from 14, 39, 106)

Gene symbol	Gene name	Role in obesity	Epigenetic mechanism	Ref.
<i>PPARGC1A</i>	Peroxisome proliferator-activated receptor gamma-coactivator-1 alpha	Adipogenesis	Important in human islet insulin secretion	[107]
<i>NR0B2</i>	Nuclear receptor small heterodimer partner Fibroblast growth factor-2	Adipogenesis	Tumour suppressor methylation related	[108]
<i>FGF2</i>	Fibroblast growth factor-2	Adipogenesis	Homocysteine disrupts endothelial cells through altered promoter DNA methylation	[109]
<i>PPARG</i>	Peroxisome proliferator-activated receptor gamma	Adipogenesis	Changes in DNA methylation during cellular ageing and atherosclerosis	[61]
<i>PTEN</i>	Phosphatase and tensin homologue	Adipogenesis	Epigenetic role in colorectal cancer and gliomas	[110]
<i>RARA</i>	Retinoic acid receptor-alpha	Adipogenesis	Promoter hypermethylation associated with prostate and mammary cancer	[111]
<i>CDKN1A</i>	Cyclin-dependent kinase inhibitor 1A (p21, Cip1)	Adipogenesis and cell cycle	Aberrant promoter methylations related to cancer	[112]
<i>LEP</i>	Leptin	Adipogenesis and inflammation, appetite regulation	Post-zygotic development, adipocyte maturation and cellular ageing, DNA methylation	[53, 61, 113, 114]
<i>ESR1</i>	Oestrogen receptor-alpha	Adipogenesis	Prognostic value of Oestrogen receptor hypermethylation	[115]
<i>NR3C1</i>	Glucocorticoid receptor	Adipogenesis, stress	Methylation status is sensitive to prenatal maternal mood, Histone acetylation	[116]
<i>CEBPA</i>	CCAAT/enhancer binding protein (C/EBP) a	Adipogenesis	Histone acetylation and methylation	[117]

Gene symbol	Gene name	Role in obesity	Epigenetic mechanism	Ref.
<i>PPARA</i>	Peroxisome proliferator-activated receptor α	Adipogenesis	DNA methylation	[118]
<i>Avy</i>	Agouti	Appetite control and adipogenesis	DNA methylation	[119, 120]
<i>MC4R</i>	Melanocortin 4 receptor	Appetite regulation	DNA methylation	[121]
<i>NPY</i>	Neuropeptide Y	Appetite regulation	DNA methylation	[122]
<i>POMC</i>	Proopiomelanocortin	Appetite regulation	DNA methylation and histone acetylation and methylation	[122]
<i>TNF</i>	Tumor necrosis factor	Inflammation and insulin resistance	Epigenetic silencing during endotoxin tolerance and myeloid differentiation, DNA methylation, Methylation status is sensitive to prenatal maternal mood	[66]
<i>PLA2G4A</i>	Plasma secretory phospholipase A(2) type IIA	Inflammation	Malignant prostate cells	[123]
<i>SOD3</i>	Extracellular superoxide dismutase	Inflammation	Development of foam cells	[124]
<i>SOCS1</i>	Suppressor of cytokine signalling 1	Inflammation	Severity of liver fibrosis and hepatocarcinoma	[125]
<i>SOCS3</i>	Suppressor of cytokine signalling 3	Inflammation	Role in cellular growth and migration and melanomas	[126]
<i>IFNG</i>	Interferon γ	Inflammation	DNA methylation	[127]
<i>CASP8</i>	Caspase 8	Inflammation and apoptosis	Hypermethylation in neuroblastomas and medulloblastomas	[128]
<i>COX7A1</i>	Cytochrome c oxidase subunit VIIa polypeptide 1 (muscle)	Energy metabolism	Age influences DNA methylation	[129]
<i>LPL</i>	Lipoprotein lipase	Fat metabolism	Changes in DNA methylation during cellular ageing	[61]

Table 1. (Continued)

Gene symbol	Gene name	Role in obesity	Epigenetic mechanism	Ref.
<i>FABP4</i>	Fatty acid-binding protein 4, adipocyte	Fat metabolism	Changes in DNA methylation during cellular ageing	[61]
<i>FASN</i>	Fatty acid synthase	Lipid storage	DNA methylation	[130]
<i>CAV1</i>	Caveolin-1	Insulin signalling	Aberrant methylation is associated with hepatocellular carcinoma	[131]
<i>PIK3CG</i>	Phosphoinositide-3-kinase, catalytic, gamma pp	Insulin signalling	CpG hypermethylation is associated with progression of colorectal cancer	[132]
<i>SSTR2</i>	Somatostatin receptor 2	Insulin resistance	Tissue-specific related	[133]
<i>HSD11B2</i>	11 beta-hydroxysteroid dehydrogenase 2	Insulin resistance and adiposity	<i>In vivo</i> epigenetic repression and relation to hypertension	[134]
<i>IGFBP3</i>	Insulin-like growth factor-binding protein 3	Insulin resistance	Hypermethylation is associated with non-small cell lung cancer	[135]
<i>IGF2</i>	Insulin-like growth factor 2	Glucose homeostasis	DNA methylation and histone acetylation	[136]
<i>ADIPOQ</i>	Adiponectin	Glucose homeostasis	DNA methylation and histone acetylation	[137]
<i>GLUT4</i>	Insulin-responsive glucose transporter 4	Glucose homeostasis	DNA methylation and histone acetylation	[117]
<i>INS</i>	Insulin	Glucose homeostasis	DNA methylation and histone acetylation	[138, 139]
<i>INSR</i>	Insulin receptor	Adipogenesis, Insulin sensitivity disruption	DNA methylation	[140]
<i>FTO</i>	Fat mass and obesity associated	Body weight homeostasis	DNA methylation	[141]
<i>HIF1A</i>	Hypoxia inducible factor 1	Hypoxia	DNA methylation and histone acetylation and methylation	[142]
<i>UCP1</i>	Uncoupling protein 1	Thermogenesis	DNA methylation	[143]

Recently, primers are being designed for studying methylation patterns of genes regulated by methylation process. CpG islands of susceptible epiobesigenes are analyzed to emerge their epigenetic regulation [146]. With these analysis, it has been found that, *FGF-2*, *PTEN*, *CDKN1A*, and *ESR1* genes have enough CpG islands in their promoter region to be potential susceptibility genes for obesity and adipogenesis [147, 148-150]. In leptin regulation, *SOCS1* and *SOCS3* (suppressor of cytokine signaling 1 and 3) take place and cytokine signaling affected in obese state is attenuated [151-153]. Also, in epigenetic regulation of obesity, energy homeostasis genes, *COX7a*, *LPL*, and insulin-related genes like *CAV1* and *IGFBP3* can be potential targets [2, 154-156]. There are also other genes which do not have CpG risk sequences and have roles in body weight control [66].

For example, Glucocorticoid receptor and *TNF- α* genes are under epigenetic control but don't have remarkable CpG islands although they have significant roles in obesity pathogenesis [66, 157, 158].

CpG methylation in somatic cells may be changed by some environmental factors in postnatal life. These environmental factors may cause variation or reversibility in the DNA methylation patterns. One of these environmental factors is aging. Aging affects the DNA methylation patterns in tissue-specific way [159]. Methylation of such genes like hepatic glucokinase increases with age [160]. This shows that, DNA methylation can play a role in age-related susceptibility to hepatic insulin resistance and diabetes. Other factors influencing DNA methylation is inflammation [161], oxidative stress [162], and hypoxia [163]. In adipose tissue of obese subjects, these factors are all exacerbated [147].

In recent years, the relationship between obesity and epigenetic pathways are reported. Methylation of many genes is reversible since it can be altered by exercise or diet throughout life. Alternatively, methylation of several genes at birth are associated with obesity later in life since epigenetic regulation of inheritance may be more stable in human [164]. It has been shown that, hypocaloric diet may be associated with *TNF- α* promoter in PBMC cells [39]. Recently, studies have also been demonstrated the effects of different dietary compounds on epigenetic marks and, on gene expression regulation in disease states. Some of the nutritional factors influencing the epigenetic marks on some disease states are shown in Table 2 [14, 165].

Studies are going on to investigate the relationship between obesity, obesity related diseases and epigenetics. In a recent study, increased DNA methylation has been associated with T2D in a Swedish population [175]. Also, global DNA hypermethylation has been related with diabetic retinopathy in T2D [176].

In another study, *BCL11A* gene polymorphism has been established as a risk for T2D, and a recent study suggested a possible male gender-specific association between *BCL11A* gene methylation and T2D [177]. Besides, cholesterol and the different types of fatty acids in the diet have been shown to affect on genome-wide DNA methylation patterns [178]. Recently, a GWAS on 459 European origin individuals was performed to investigate the relationship between DNA methylation and BMI, and the analysis reported that increased BMI is associated with the elevation of methylation at the *HIF3A* locus in blood cells and in adipose tissue [179]. This study generated a support for the relationship between the epigenome and the development of obesity in genetically susceptible individuals.

Table 2. Some dietary factors influencing epigenetic marks and have beneficial effects on obesity (modified from 14, 165)

Dietary factor	Related Metabolic Disease	Epigenetic mechanism	Ref.
Methyl donors			
Methionine	Insulin resistance, obesity	Histone and DNA methylation	[166]
Folate	Insulin resistance, adiposity	DNA methylation	[166]
Vitamin B-12	Insulin resistance, obesity	DNA methylation	[166]
Serine, Glycine and Histidine	Amino acid metabolism	Histone and DNA methylation	[98]
Phytochemicals			
Curcumin	Inflammation, obesity	Histone acetylation, DNA methylation, and microRNA	[167]
Epigallocatechin 3-gallate	Obesity, insulin resistance	Histone acetylation and DNA methylation	[168]
Genistein	Obesity	Histone acetylation and DNA methylation	[169]
Sulforaphane	Adipocyte differentiation	Histone acetylation	[170]
Resveratrol	Obesity	Histone acetylation	[171, 172]
Fatty acids			
Eicosapentaenoic acid	n-3 Polyunsaturated fatty acid metabolism	DNA methylation	[173]
Docosahexaenoic acid	n-3 Polyunsaturated fatty acid metabolism	DNA methylation	[174]

CONCLUSION

In human obesity, many genes have been identified. The effects of these genes may be modified by epigenetic mechanisms. There is preliminary but compelling evidence from human and animal studies that changes in the epigenetic states may contribute to this era's epidemic, obesity. However, there are still questions remaining. Does the epigenetic analysis of a specific tissue mirror the epigenome of another tissue?, and are these epigenetic changes reversible? Further studies in epigenetics will enlighten the obesity etiology and related metabolic disease phenotypes and will provide therapeutic approaches to improve obesity prognoses.

REFERENCES

- [1] James P. T., Leach R., Kalamara E., Shayeghi M. The worldwide obesity epidemic. *Obes. Res.*, 2001; 9: 228S–233S.

- [2] Martinez J. A., Parra M. D., Santos J. L., Moreno-Aliaga M. J., Marti A., Martinez-Gonzalez M. A. Genotype-dependent response to energy-restricted diets in obese subjects: towards personalized nutrition. *Asia Pac. J. Clin. Nutr.*, 2008; 17(Suppl. 1): 119–122.
- [3] Reilly M. P., Rader D. J. The metabolic syndrome: more than the sum of its parts? *Circulation*, 2003; 108: 1546–1551.
- [4] Smith C. J., Ryckman K. K. Epigenetic and developmental influences on the risk of obesity, diabetes, and metabolic syndrome. *Diabetes Metab. Syndr. Obes.*, 2015 Jun. 29; 8: 295–302.
- [5] Mau T., Yung R. Potential of epigenetic therapies in non-cancerous conditions. *Front. Genet.*, 2014 Dec. 19; 5: 438.
- [6] <http://www.hsph.harvard.edu/obesity-prevention-source/obesity-definition/> (Last accessed: 01.08.2015).
- [7] Dhurandhar E. J., Keith S. W. The aetiology of obesity beyond eating more and exercising less. *Best Pract. Res. Clin. Gastroenterol.*, 2014 Aug.; 28(4): 533–44.
- [8] de Rooij S. R., Painter R. C., Holleman F., Bossuyt P. M., Roseboom T. J. The metabolic syndrome in adults prenatally exposed to the Dutch famine. *Am. J. Clin. Nutr.*, 2007; 86: 1219–1224.
- [9] Desai M., Gayle D., Babu J., Ross M. G. Programmed obesity in intrauterine growthrestricted newborns: modulation by newborn nutrition. *Am. J. Physiol. Regul. Integr. Comp. Physiol.*, 2005; 288: R91–R96.
- [10] Desai M., Jellyman J. K., Han G., Beall M., Lane R. H., Ross M. G. Maternal obesity and high-fat diet program offspring metabolic syndrome. *Am. J. Obstet. Gynecol.*, 2014; 211: 237.e1–237.e13.
- [11] Vickers M. H., Krechowec S. O., Breier B. H. Is later obesity programmed in utero? *Curr. Drug Targets*, 2007; 8: 923–934.
- [12] Remacle C., Bieswal F., Reusens B. Programming of obesity and cardiovascular disease. *Int. J. Obes. Relat. Metab. Disord.*, 2004; 28: S46–S53.
- [13] Desai M., Jellyman J. K., Ross M. G. Epigenomics, gestational programming and risk of metabolic syndrome. *Int. J. Obes.*, (Lond) 2015 Apr.; 39(4): 633–41.
- [14] Martínez J. A., Milagro F. I., Claycombe K. J., Schalinske K. L. Epigenetics in adipose tissue, obesity, weight loss, and diabetes. *Adv. Nutr.*, 2014 Jan. 1; 5(1): 71–81.
- [15] Junien C., Nathanielsz P. Report on the IASO Stock Conference 2006: early and lifelong environmental epigenomic programming of metabolic syndrome, obesity and type II diabetes. *Obes. Rev.*, 2007; 8: 487–502.
- [16] Dolinoy D. C., Jirtle R. L. Environmental epigenomics in human health and disease. *Environ. Mol. Mutagen.*, 2008; 49: 4–8.
- [17] Chuang J. C., Jones P. A. Epigenetics and microRNAs. *Pediatr. Res.*, 2007; 61: 24–29.
- [18] Klose R. J., and Bird A. P. Genomic DNA methylation: the mark and its mediators. *Trends Biochem. Sci.*, 2006; 31: 89–97.
- [19] Feng J., and Fan G. The role of DNA methylation in the central nervous system and neuropsychiatric disorders. *Int. Rev. Neurobiol.*, 2009; 89: 67–84.
- [20] Watt F. and Molloy P. L. Cytosine methylation prevents binding to DNA of a HeLa cell transcription factor required for optimal expression of the adenovirus major late promoter. *Genes Dev.*, 1988; 2: 1136–1143.

- [21] Fuks F., Burgers W. A., Brehm A., Hughes-Davies L., Kouzarides T. DNA methyltransferases Dnmt1 associates with histone deacetylase activity. *Nat. Genet.*, 2000; 24: 88-91.
- [22] Lorincz M. C., Dickerson D. R., Schmitt M., Groudine M. Intra- genic DNA methylation alters chromatin structure and elongation efficiency in mammalian cells. *Nat. Struct. Mol. Biol.*, 2004; 11: 1068-1075.
- [23] Nan X., Campoy F. J., Bird A. MeCP2 is a transcriptional repressor with abundant binding sites in genomic chromatin. *Cell*, 1997; 88: 471-481.
- [24] Nan X., Ng H. H., Johnson C. A., Laherty C. D., Turner B. M., Eisenman R. N., et al. Transcriptional repression by the methyl-CpG-binding protein MeCP2 involves a histone deacetylase complex. *Nature*, 1998; 393: 386-389.
- [25] Jenuwein T., Allis C. D. Translating the histone code. *Science*, 2001; 293: 1074–1080.
- [26] Cosgrove M. S. Histone proteomics and the epigenetic regulation of nucleosome mobility. *Expert Rev. Proteomics*, 2007; 4: 465–478.
- [27] Schwartz Y. B., Pirrotta V. Polycomb complexes and epigenetic states. *Curr. Opin. Cell Biol.*, 2008; 20: 266–273.
- [28] Higgs P. G., Lehman N. The RNA World: molecular cooperation at the origins of life. *Nat. Rev. Genet.*, 2015; 16: 7–17.
- [29] Taft R. J., Pang K. C., Mercer T. R., Dinger M., Mattick J. S. Non-coding RNAs: regulators of disease. *J. Pathol.*, 2010; 220: 126–139.
- [30] Rinn J. L., Kertesz M., Wang J. K., Squazzo S. L., Xu X., Bruggmann S. A. et al. Functional demarcation of active and silent chromatin domains in human HOX loci by noncoding RNAs. *Cell*, 2007; 129: 1311–1323.
- [31] Mercer T. R., Dinger M. E., Sunkin S. M., Mehler M. F., Mattick J. S. Specific expression of long noncoding RNAs in the mouse brain. *Proc. Natl. Acad. Sci. USA*, 2008; 105: 716–721.
- [32] Clemson C. M., Hutchinson J. N., Sara S. A., Ensminger A. W., Fox A. H., Chess A. et al. An architectural role for a nuclear noncoding RNA: NEAT1 RNA is essential for the structure of paraspeckles. *Mol. Cell*, 2009; 33: 717–726.
- [33] Sun L., Goff L. A., Trapnell C., Alexander R., Lo K. A., Hacisuleyman E. et al. Long noncoding RNAs regulate adipogenesis. *Proc. Natl. Acad. Sci. USA*, 2013; 110: 3387–3392.
- [34] Zhao X. Y., Li S., Wang G. X., Yu Q., Lin J. D. A long noncoding RNA transcriptional regulatory circuit drives thermogenic adipocyte differentiation. *Mol. Cell*, 2014; 55: 372–382.
- [35] Chen K., Rajewsky N. The evolution of gene regulation by transcription factors and microRNAs. *Nat. Rev. Genet.*, 2007; 8: 93–103.
- [36] Ding X. C., Weiler J., Grosshans H. Regulating the regulators: mechanisms controlling the maturation of microRNAs. *Trends Biotechnol.*, 2009; 27: 27–36.
- [37] Saetrom P., Snøve O. Jr., Rossi J. J. Epigenetics and microRNAs. *Pediatr. Res.*, 2007 May; 61(5 Pt 2): 17-23.
- [38] Nafee T. M., Farrell W. E., Carroll W. D., Fryer A. A., Ismail K. M. Epigenetic control of fetal gene expression. *BJOG*, 2008; 115: 158–168.
- [39] Campión J., Milagro F. I., Goyenechea E., Martínez J. A. TNF-alpha promoter methylation as a predictive biomarker for weight-loss response. *Obesity*, (Silver Spring). 2009 Jun.; 17(6): 1293-7.

- [40] Zhu J. DNA methylation and hepatocellular carcinoma. *J. Hepatobiliary Pancreat. Surg.*, 2006; 13: 265–273.
- [41] Campión J., Milagro F., Martínez J. A. Epigenetics and obesity. *Prog. Mol. Biol. Transl. Sci.*, 2010; 94: 291–347.
- [42] Bouchard L., Rabasa-Lhoret R., Faraj M., Lavoie M. E., Mill J., Pérusse L. et al. Differential epigenomic and transcriptomic responses in subcutaneous adipose tissue between low and high responders to caloric restriction. *Am. J. Clin. Nutr.*, 2010; 91: 309–320.
- [43] McGee S. L., Hargreaves M. Exercise and skeletal muscle glucose transporter 4 expression: molecular mechanisms. *Clin. Exp. Pharmacol. Physiol.*, 2006; 33: 395–399.
- [44] Fraga M. F., Ballestar E., Paz M. F., Ropero S., Setien F., Ballestar M. L. et al. Epigenetic differences arise during the lifetime of monozygotic twins. *Proc. Natl. Acad. Sci. USA*, 2005; 102: 10604–10609.
- [45] Moleres A., Campión J., Milagro F. I., Marcos A., Campoy C., Garagorri J. M. et al. Differential DNA methylation patterns between high and low responders to a weight loss intervention in overweight or obese adolescents: the EVASYON study. *FASEB J.*, 2013; 27: 2504–2512.
- [46] Marti A., Martinez-Gonzalez M. A., Martinez J. A. Interaction between genes and lifestyle factors on obesity. *Proc. Nutr. Soc.*, 2008; 67: 1–8.
- [47] Wolff G. L., Roberts D. W., Mountjoy K. G. Physiological consequences of ectopic agouti gene expression: the yellow obese mouse syndrome. *Physiol. Genomics*, 1999; 1: 151–63.
- [48] Morgan H. D., Sutherland H. G., Martin D. I., Whitelaw E. Epigenetic inheritance at the agouti locus in the mouse. *Nat. Genet.*, 1999; 23: 314–318.
- [49] Waterland R. A., Jirtle R. L. Transposable elements: targets for early nutritional effects on epigenetic gene regulation. *Mol. Cell Biol.*, 2003; 23: 5293–5300.
- [50] Strakovsky R. S., Zhang X., Zhou D., Pan Y. X. Gestational high fat diet programs hepatic phosphoenolpyruvate carboxykinase gene expression and histone modification in neonatal offspring rats. *J. Physiol.*, 2011; 589: 2707–17.
- [51] Vucetic Z., Kimmel J., Totoki K., Hollenbeck E., Reyes T. M. Maternal high-fat diet alters methylation and gene expression of dopamine and opioid-related genes. *Endocrinology*, 2010; 151: 4756–64.
- [52] Vucetic Z., Kimmel J., Reyes T. M. Chronic high-fat diet drives postnatal epigenetic regulation of mu-opioid receptor in the brain. *Neuropsychopharmacol e off Publ. Am. Coll. Neuropsychopharmacol.*, 2011; 36: 1199–206.
- [53] Milagro F. I., Campion J., Garcia-Diaz D. F., Goyenechea E., Paternain L., Martinez J. A. High fat diet-induced obesity modifies the methylation pattern of leptin promoter in rats. *J. Physiol. Biochem.*, 2009; 65: 1–9.
- [54] Waterland R. A., Travisano M., Tahiliani K. G. Diet-induced hypermethylation at agouti viable yellow is not inherited transgenerationally through the female. *FASEB J.*, 2007; 21: 3380–3385.
- [55] Weinstein L. S., Xie T., Qasem A., Wang J., Chen M. The role of GNAS and other imprinted genes in the development of obesity. *Int. J. Obes.*, (Lond). 2010 Jan.; 34(1):6–17.
- [56] Pant V., Mariano P., Kanduri C., Mattsson A., Lobanenko V., Heuchel R., Ohlsson R. The nucleotides responsible for the direct physical contact between the chromatin

- insulator protein CTCF and the H19 imprinting control region manifest parent of origin-specific long-distance insulation and methylation-free domains. *Genes Dev.*, 2003; 17: 586–90.
- [57] Gao H., Sathishkumar K. R., Yallampalli U., Balakrishnan M., Li X., Wu G., Yallampalli C. Maternal protein restriction regulates IGF2 system in placental labyrinth. *Front. Biosci.*, (Elite Ed) 2012; 4: 1434–50.
- [58] Grandjean V., O'Neill L., Sado T., Turner B., Ferguson-Smith A. Relationship between DNA methylation, histone H4 acetylation and gene expression in the mouse imprinted Igf2–H19 domain. *FEBS Lett.*, 2001; 488: 165–9.
- [59] Li T., Hu J. F., Qiu X., Ling J., Chen H., Wang S., Hou A., Vu T. H., Hoffman A. R. CTCF regulates allelic expression of Igf2 by orchestrating a promoter- polycomb repressive complex 2 intrachromosomal loop. *Mol. Cell Biol.*, 2008; 28: 6473–82.
- [60] Yang B. T., Dayeh T. A., Kirkpatrick C. L., Taneera J., Kumar R., Groop L. et al. Insulin promoter DNA methylation correlates negatively with insulin gene expression and positively with HbA(1c) levels in human pancreatic islets. *Diabetologia*, 2011; 54: 360–367.
- [61] Noer A., Boquest A. C., Collas P. Dynamics of adipogenic promoter DNA methylation during clonal culture of human adipose stem cells to senescence. *BMC Cell Biol.*, 2007; 8: 18.
- [62] Stepanow S., Reichwald K., Huse K., Gausmann U., Nebel A., Rosenstiel P. et al. Allele-specific, age-dependent and BMI-associated DNA methylation of human MCHR1. *PLoS One*, 2011; 6: 17711.
- [63] Yokomori N., Tawata M., Onaya T. DNA demethylation during the differentiation of 3T3-L1 cells affects the expression of the mouse GLUT4 gene. *Diabetes*, 1999; 48: 685–690.
- [64] Fujiki K., Kano F., Shiota K., Murata M. Expression of the peroxisome proliferator activated receptor gamma gene is repressed by DNA methylation in visceral adipose tissue of mouse models of diabetes. *BMC Biol.*, 2009; 7: 38.
- [65] Musri M. M., Corominola H., Casamitjana R., Gomis R., Párrizas M. Histone H3 lysine 4 dimethylation signals the transcriptional competence of the adiponectin promoter in preadipocytes. *J. Biol. Chem.*, 2006; 281: 17180–17188.
- [66] Sullivan K. E., Reddy A. B., Dietzmann K., Suriano A. R., Kocieda V. P., Stewart M. et al. Epigenetic regulation of tumor necrosis factor alpha. *Mol. Cell Biol.*, 2007; 27: 5147–5160.
- [67] Kilpeläinen T. O., Zillikens M. C., Stančáková A., Finucane F. M., Ried J. S., Langenberg C. et al. Genetic variation near IRS1 associates with reduced adiposity and an impaired metabolic profile. *Nat. Genet.*, 2011; 43: 753–760.
- [68] Musri M. M., Párrizas M. Epigenetic regulation of adipogenesis. *Curr. Opin. Clin. Nutr. Metab. Care*, 2012; 15: 342–349.
- [69] Ge K. Epigenetic regulation of adipogenesis by histone methylation. *Biochim. Biophys. Acta*, 2012; 1819: 727–732.
- [70] Li H. X., Xiao L., Wang C., Gao J. L., Zhai Y. G. Review: Epigenetic regulation of adipocyte differentiation and adipogenesis. *J. Zhejiang Univ. Sci. B*, 2010; 11: 784–791.
- [71] Burdge G. C., Hoile S. P., Uller T., Thomas N. A., Gluckman P. D., Hanson M. A. et al. Progressive, transgenerational changes in offspring phenotype and epigenotype following nutritional transition. *PLoS One*, 2011; 6: e28282.

- [72] Wang X., Zhu H., Snieder H., Su S., Munn D., Harshfield G. et al. Obesity related methylation changes in DNA of peripheral blood leukocytes. *BMC Med.*, 2010; 8: 87.
- [73] Jufvas A., Sjödin S., Lundqvist K., Amin R., Vener A. V., Strålfors P. Global differences in specific histone H3 methylation are associated with overweight and type 2 diabetes. *Clin. Epigenetics*, 2013; 5: 15.
- [74] Ortega F. J., Moreno-Navarrete J. M., Pardo G., Sabater M., Hummel M., Ferrer A. et al. MiRNA expression profile of human subcutaneous adipose and during adipocyte differentiation. *PLoS One*, 2010; 5: e9022.
- [75] Tateishi K., Okada Y., Kallin E. M., Zhang Y. Role of Jhdm2a in regulating metabolic gene expression and obesity resistance. *Nature*, 2009; 458: 757–761.
- [76] Inagaki T., Tachibana M., Magoori K., Kudo H., Tanaka T., Okamura M. et al. Obesity and metabolic syndrome in histone demethylase JHDM2a-deficient mice. *Genes Cells*, 2009; 14: 991–1001.
- [77] Lu Y. L., Jing W., Feng L. S., Zhang L., Xu J. F., You T. J. et al. Effects of hypoxic exercise training on microRNA expression and lipid metabolism in obese rat livers. *J. Zhejiang Univ. Sci. B*, 2014; 15: 820–829.
- [78] Son Y. H., Ka S., Kim A. Y., Kim J. B. Regulation of Adipocyte Differentiation via MicroRNAs. *Endocrinol. Metab.*, (Seoul) 2014; 29: 122–135.
- [79] Vinnikov I. A., Hajdukiewicz K., Reymann J., Beneke J., Czajkowski R., Roth L. C. et al. Hypothalamic miR-103 protects from hyperphagic obesity in mice. *J. Neurosci.*, 2014; 34: 10659–10674.
- [80] Funato H., Oda S., Yokofujita J., Igarashi H., Kuroda M. Fasting and high-fat diet alter histone deacetylase expression in the medial hypothalamus. *PLoS One*, 2011; 6: e18950.
- [81] McGregor R. A., Choi M. S. microRNAs in the regulation of adipogenesis and obesity. *Curr. Mol. Med.*, 2011; 11: 304–316.
- [82] Yang Q. Y., Liang J. F., Rogers C. J., Zhao J. X., Zhu M. J., Du M. Maternal obesity induces epigenetic modifications to facilitate Zfp423 expression and enhance adipogenic differentiation in fetal mice. *Diabetes*, 2013; 62: 3727–3735.
- [83] Borengasser S. J., Zhong Y., Kang P., Lindsey F., Ronis M. J., Badger T. M. et al. Maternal obesity enhances white adipose tissue differentiation and alters genome-scale DNA methylation in male rat offspring. *Endocrinology*, 2013; 154: 4113–4125.
- [84] Kamei Y., Suganami T., Ehara T., Kanai S., Hayashi K., Yamamoto Y. et al. Increased expression of DNA methyltransferase 3a in obese adipose tissue: studies with transgenic mice. *Obesity*, (Silver Spring) 2010; 18: 314–321.
- [85] Lee J., Saha P. K., Yang Q. H., Lee S., Park J. Y., Suh Y. et al. Targeted inactivation of MLL3 histone H3-Lys-4 methyltransferase activity in the mouse reveals vital roles for MLL3 in adipogenesis. *Proc. Natl. Acad. Sci. USA*, 2008; 105: 19229–19234.
- [86] Morgan D. K., Whitelaw E. The case for transgenerational epigenetic inheritance in humans. *Mamm Genome*, 2008; 9: R182.
- [87] Rakyan V. K., Chong S., Champ M. E., Cuthbert P. C., Morgan H. D., Luu K. V., Whitelaw E. Transgenerational inheritance of epigenetic states at the murine Axin(Fu) allele occurs after maternal and paternal transmission. *Proc. Natl. Acad. Sci. U S A*, 2003; 100: 2538–2543.
- [88] Vickers M. H. Developmental programming and adult obesity: the role of leptin. *Curr. Opin. Endocrinol. Diabetes Obes.*, 2007; 14: 17–22.
- [89] Cobiac L. Epigenomics and nutrition. *Forum Nutr.*, 2007; 60: 31–41.

- [90] Heijmans B. T., Tobi E. W., Stein A. D., Putter H., Blauw G. J., Susser E. S. et al. Persistent epigenetic differences associated with prenatal exposure to famine in humans. *Proc. Natl. Acad. Sci. USA*, 2008; 105: 17046–17049.
- [91] Tobi E. W., Lumey L. H., Talens R. P., Kremer D., Putter H., Stein A. D. et al. DNA methylation differences after exposure to prenatal famine are common and timing- and sex-specific. *Hum. Mol. Genet.*, 2009; 18: 4046–4053.
- [92] Soubry A., Murphy S. K., Wang F., Huang Z., Vidal A. C., Fuemmeler B. F., Kurtzberg J., Murtha A., Jirtle R. L., Schildkraut J. M., Hoyo C. Newborns of obese parents have altered DNA methylation patterns at imprinted genes. *Int. J. Obes.*, (Lond). 2015 Apr.; 39(4):650-7.
- [93] Soubry A., Schildkraut J. M., Murtha A., Wang F., Huang Z., Bernal A. et al. Paternal obesity is associated with IGF2 hypomethylation in newborns: results from a Newborn Epigenetics Study (NEST) cohort. *BMC Med.*, 2013; 11: 29.
- [94] Morales E., Groom A., Lawlor D. A., Relton C. L. DNA methylation signatures in cord blood associated with maternal gestational weight gain: results from the ALSPAC cohort. *BMC Res. Notes*, 2014; 7: 278.
- [95] Godfrey K. M., Sheppard A., Gluckman P. D., Lillycrop K. A., Burdge G. C., McLean C. et al. Epigenetic gene promoter methylation at birth is associated with child's later adiposity. *Diabetes*, 2011; 60: 1528–1534.
- [96] Lavebratt C., Almgren M., Ekström T. J. Epigenetic regulation in obesity. *Int. J. Obes.*, (Lond) 2012; 36: 757–765.
- [97] Lillycrop K. A., Burdge G. C. Epigenetic changes in early life and future risk of obesity. *Int. J. Obes.*, (Lond) 2011; 35: 72–83.
- [98] Wang J., Wu Z., Li D., Li N., Dindot S. V., Satterfield M. C., Bazer F. W., Wu G. Nutrition, epigenetics, and metabolic syndrome. *Antioxid. Redox. Signal*, 2012; 17: 282–301.
- [99] Ganu R. S., Harris R. A., Collins K., Aagaard K. M. Maternal diet: a modulator for epigenomic regulation during development in nonhuman primates and humans. *Int. J. Obes. Suppl.*, 2012; 2: S14–S18.
- [100] Tosh D. N., Fu Q., Callaway C. W., McKnight R. A., McMillen I. C., Ross M. G. et al. Epigenetics of programmed obesity: alteration in IUGR rat hepatic IGF1 mRNA expression and histone structure in rapid vs delayed postnatal catch-up growth. *Am. J. Physiol. Gastrointest. Liver Physiol.*, 2010; 299: G1023–G1029.
- [101] Desai M., Jellyman J. K., Han G., Beall M. H., Ross M. G. Mechanism of programmed obesity in intrauterine growth restricted newborns: epigenetic mediated early induction of adipocyte differentiation contributes to enhanced adipogenesis. *J. Dev. Orig. Health Dis.*, 2013; 4(Suppl. 2): S138.
- [102] Cho C. E., Sánchez-Hernández D., Reza-López S. A., Huot P. S., Kim Y. I., Anderson G. H. High folate gestational and post-weaning diets alter hypothalamic feeding pathways by DNA methylation in Wistar rat offspring. *Epigenetics*, 2013; 8: 710–719.
- [103] Stevens A., Begum G., Cook A., Connor K., Rumball C., Oliver M. et al. Epigenetic changes in the hypothalamic proopiomelanocortin and glucocorticoid receptor genes in the ovine fetus after periconceptional undernutrition. *Endocrinology*, 2010; 151: 3652–3664.
- [104] Plagemann A., Harder T., Brunn M., Harder A., Roepke K., Wittrock-Staar M. et al. Hypothalamic proopiomelanocortin promoter methylation becomes altered by early

- overfeeding: an epigenetic model of obesity and the metabolic syndrome. *J. Physiol.*, 2009; 587: 4963–4976.
- [105] Raychaudhuri N., Raychaudhuri S., Thamotharan M., Devaskar S. U. Histone code modifications repress glucose transporter 4 expression in the intrauterine growth-restricted offspring. *J. Biol. Chem.*, 2008; 283: 13611–13626.
- [106] Drummond E. M., Gibney E. R. Epigenetic regulation in obesity. *Curr. Opin. Clin. Nutr. Metab. Care*, 2013 Jul.; 16(4): 392-7.
- [107] Ling C., Del Guerra S., Lupi R., Ronn T., Granhall C., Luthman H., Masiello P., Marchetti P., Groop L., Del Prato S. Epigenetic regulation of PPARGC1A in human type 2 diabetic islets and effect on insulin secretion. *Diabetologia*, 2008; 51: 615–622.
- [108] He N., Park K., Zhang Y., Huang J., Lu S., Wang L. Epigenetic inhibition of nuclear receptor small heterodimer partner is associated with and regulates hepatocellular carcinoma growth. *Gastroenterology*, 2008; 134: 793–802.
- [109] Chang P. Y., Lu S. C., Lee C. M., Chen Y. J., Dugan T. A., Huang W. H., Chang S. F., Liao W. S., Chen C. H., Lee Y. T. Homocysteine inhibits arterial endothelial cell growth through transcriptional downregulation of fibroblast growth factor-2 involving G protein and DNA methylation. *Circ. Res.*, 2008; 102: 933–941.
- [110] Goel A., Arnold C. N., Niedzwiecki D., Carethers J. M., Dowell J. M., Wasserman L., Compton C., Mayer R. J., Bertagnolli M. M., Boland C. R. Frequent inactivation of PTEN by promoter hypermethylation in microsatellite instability-high sporadic colorectal cancers. *Cancer Res.*, 2004; 64: 3014–3021.
- [111] Roupert M., Hupertan V., Catto J. W., Yates D. R., Rehman I., Proctor L. M., Phillips J., Meuth M., Cussenot O., Hamdy F. C. Promoter hypermethylation in circulating blood cells identifies prostate cancer progression. *Int. J. Cancer*, 2008; 122: 952–956.
- [112] Majid S., Kikuno N., Nelles J., Noonan E., Tanaka Y., Kawamoto K., Hirata H., Li L. C., Zhao H., Okino S. T., Place R. F., Pookot D., Dahiya R. Genistein induces the p21WAF1/CIP1 and p16INK4a tumor suppressor genes in prostate cancer cells by epigenetic mechanisms involving active chromatin modification. *Cancer Res.*, 2008; 68: 2736–2744.
- [113] Stoger R. In vivo methylation patterns of the leptin promoter in human and mouse. *Epigenetics*, 2006; 1: 155–162.
- [114] Melzner I., Scott V., Dorsch K., Fischer P., Wabitsch M., Bruderlein S., Hasel C., Moller P. Leptin gene expression in human preadipocytes is switched on by maturation-induced demethylation of distinct CpGs in its proximal promoter. *J. Biol. Chem.*, 2002; 277: 45420–45427.
- [115] Lai J. C., Cheng Y. W., Chiou H. L., Wu M. F., Chen C. Y., Lee H. Gender difference in estrogen receptor- α promoter hypermethylation and its prognostic value in non-small cell lung cancer. *Int. J. Cancer*, 2005; 117: 974–980.
- [116] Oberlander T. F., Weinberg J., Papsdorf M., Grunau R., Misri S., Devlin A. M. Prenatal exposure to maternal depression, neonatal methylation of human glucocorticoid receptor gene (NR3C1) and infant cortisol stress responses. *Epigenetics*, 2008; 3: 97–106.
- [117] Zheng S., Rollet M., Pan Y. X. Maternal protein restriction during pregnancy induces CCAAT/enhancer-binding protein (C/EBP β) expression through the regulation of histone modification at its promoter region in female offspring rat skeletal muscle. *Epigenetics*, 2011;6: 161–70.

- [118] Lillycrop K. A., Phillips E. S., Torrens C., Hanson M. A., Jackson A. A., Burdge G. C. Feeding pregnant rats a protein-restricted diet persistently alters the methylation of specific cytosines in the hepatic PPAR alpha promoter of the offspring. *Br. J. Nutr.*, 2008; 100: 278–82.
- [119] Ebrahim S. Epigenetics: the next big thing. *Int. J. Epidemiol.*, 2012; 41: 1–3.
- [120] Delaney C., Garg S. K., Fernandes C., et al. Maternal Diet Supplemented with Methyl-Donors Protects against Atherosclerosis in F1 ApoE(–/–) Mice. *Plos One*, 2013; 8:e56253.
- [121] Widiker S., Karst S., Wagener A., Brockmann G. A. High-fat diet leads to a decreased methylation of the Mc4r gene in the obese BFMi and the lean B6 mouse lines. *J. Appl. Genet.*, 2010; 51: 193–7.
- [122] Crujeiras A. B., Campion J., Díaz-Lagares A., Milagro F. I., Goyenechea E., Abete I., Casanueva F. F., Martínez J. A. Association of weight regain with specific methylation levels in the NPY and POMC promoters in leukocytes of obese men: a translational study. *Regul. Pept.*, 2013; 186: 1–6.
- [123] Menschikowski M., Hagelgans A., Gussakovsky E., Kostka H., Paley E. L., Siegert G. Differential expression of secretory phospholipases A2 in normal and malignant prostate cell lines: regulation by cytokines, cell signaling pathways, and epigenetic mechanisms. *Neoplasia*, 2008; 10: 279–286.
- [124] Jiang Y., Jiang J., Xiong J., Cao J., Li N., Li G., Wang S. Homocysteine-induced extracellular superoxide dismutase and its epigenetic mechanisms in monocytes. *J. Exp. Biol.*, 2008; 211: 911–920.
- [125] Okochi O., Hibi K., Sakai M., Inoue S., Takeda S., Kaneko T., Nakao A. Methylation-mediated silencing of SOCS-1 gene in hepatocellular carcinoma derived from cirrhosis. *Clin. Cancer Res.*, 2003; 9: 5295–5298.
- [126] Tokita T., Maesawa C., Kimura T., Kotani K., Takahashi K., Akasaka T., Masuda T. Methylation status of the SOCS3 gene in human malignant melanomas. *Int. J. Oncol.*, 2007; 30: 689–694.
- [127] Bollati V., Baccarelli A., Sartori S., Tarantini L., Motta V., Rota F., Costa G. Epigenetic effects of shiftwork on blood DNA methylation. *Chronobiol. Int.*, 2010; 27: 1093–104.
- [128] Lazcoz P., Munoz J., Nistal M., Pestana A., Encio I., Castresana J. S. Frequent promoter hypermethylation of RASSF1A and CASP8 in neuroblastoma. *BMC Cancer*, 2006; 6: 254.
- [129] Ronn T., Poulsen P., Hansson O., Holmkvist J., Almgren P., Nilsson P., Tuomi T., Isomaa B., Groop L., Vaag A., Ling C. Age influences DNA methylation and gene expression of COX7A1 in human skeletal muscle. *Diabetologia*, 2008; 51: 1159–1168.
- [130] Lomba A., Martinez J. A., Garcia-Diaz D. F., Paternain L., Marti A., Campion J., Milagro F. I. Weight gain induced by an isocaloric pair-fed high fat diet: a nutrieigenetic study on FAS and NDUFB6 gene promoters. *Mol. Genet. Metab.*, 2010; 101: 273–8.
- [131] Hirasawa Y., Arai M., Imazeki F., Tada M., Mikata R., Fukai K., Miyazaki M., Ochiai T., Saisho H., Yokosuka O. Methylation status of genes upregulated by demethylating agent 5-aza-2'- deoxycytidine in hepatocellular carcinoma. *Oncology*, 2006; 71: 77–85.
- [132] Semba S., Itoh N., Ito M., Youssef E. M., Harada M., Moriya T., Kimura W., Yamakawa M. Down-regulation of PIK3CG, a catalytic subunit of phosphatidylinositol 3-OH kinase, by CpG hypermethylation in human colorectal carcinoma. *Clin. Cancer Res.*, 2002; 8: 3824–3831.

- [133] Torrisani J., Hanoun N., Laurell H., Lopez F., Maoret J. J., Souque A., Susini C., Cordelier P., Buscail L. Identification of an upstream promoter of the human somatostatin receptor, hSSTR2, which is controlled by epigenetic modifications. *Endocrinology*, 2008; 149: 3137–3147.
- [134] Alikhani-Koopaei R., Fouladkou F., Frey F. J., Frey B. M. Epigenetic regulation of 11 beta-hydroxysteroid dehydrogenase type 2 expression. *J. Clin. Invest.*, 2004; 114: 1146–1157.
- [135] Chang Y. S., Wang L., Liu D., Mao L., Hong W. K., Khuri F. R., Lee H. Y. Correlation between insulin-like growth factor-binding protein 3 promoter methylation and prognosis of patients with stage I non-small cell lung cancer. *Clin. Cancer Res.*, 2002; 8: 3669–3675.
- [136] Gong L., Pan Y. X., Chen H. Gestational low protein diet in the rat mediates Igf2 gene expression in male offspring via altered hepatic DNA methylation. *Epigenetics*, 2010; 5:619–626.
- [137] Bouchard L., Hivert M. F., Guay S. P., St-Pierre J., Perron P., Brisson D. Placental adiponectin gene DNA methylation levels are associated with mothers' blood glucose concentration. *Diabetes*, 2012; 61: 1272–80.
- [138] Kuroda A., Rauch T. A., Todorov I., et al. Insulin gene expression is regulated by dna methylation. *Plos One*, 2009; 4: e6953.
- [139] Mosley A. L., Ozcan S. Glucose regulates insulin gene transcription by hyperacetylation of histone H4. *J. Biol. Chem.*, 2003; 278: 19660–19666.
- [140] Plagemann A., Roepke K., Harder T., et al. Epigenetic malprogramming of the insulin receptor promoter due to developmental overfeeding. *J. Perinat. Med.*, 2010; 38: 393–400.
- [141] Toperoff G., Aran D., Kark J. D., Rosenberg M., Dubnikov T., Nissan B., Wainstein J., Friedlander Y., Levy-Lahad E., Glaser B., et al. Genome-wide survey reveals predisposing diabetes type 2-related DNA methylation variations in human peripheral blood. *Hum. Mol. Genet.*, 2012;21:371–83.
- [142] Brigati C., Banelli B., di Vinci A., Casciano I., Allemanni G., Forlani A., Borzì L., Romani M. Inflammation, HIF-1, and the epigenetics that follows. *Mediators Inflamm.*, 2010; 2010: 263914.
- [143] Shore A., Karamitri A., Kemp P., Speakman J. R., Lomax M. A. Role of Ucp1 enhancer methylation and chromatin remodelling in the control of Ucp1 expression in murine adipose tissue. *Diabetologia*, 2010;53: 1164–73.
- [144] Rankinen T., Zuberi A., Chagnon Y. C., Weisnagel S. J., Argyropoulos G., Walts B., et al. The human obesity gene map: the 2005 update. *Obes. e Silver Spring*, 2006; 14: 529–644.
- [145] Goldstone A. P. Prader-Willi syndrome: advances in genetics, pathophysiology and treatment. *Trends Endocrinol. Metab. e Tem.*, 2004; 15: 12–20.
- [146] Li L. C., Dahiya R. MethPrimer: designing primers for methylation PCRs. *Bioinformatics*, 2002; 18: 1427–1431.
- [147] Campion J., Milagro F. I., Fernandez D., Martinez J. A. Differential gene expression and adiposity reduction induced by ascorbic acid supplementation in a cafeteria model of obesity. *J. Physiol. Biochem.*, 2006; 62: 71–80.
- [148] Patel N. G., Kumar S., Eggo M. C. Essential role of fibroblast growth factor signaling in preadipocyte differentiation. *J. Clin. Endocrinol. Metab.*, 2005; 90: 1226–1232.

- [149] Sasaoka T., Wada T., Tsuneki H. Lipid phosphatases as a possible therapeutic target in cases of type 2 diabetes and obesity. *Pharmacol. Ther.*, 2006; 112: 799–809.
- [150] Shi H., Maier S., Nimmrich I., Yan P. S., Caldwell C. W., Olek A., Huang T. H. Oligonucleotide-based microarray for DNA methylation analysis: principles and applications. *J. Cell Biochem.*, 2003; 88: 138–143.
- [151] Alexander W. S., Hilton D. J. The role of suppressors of cytokine signaling (SOCS) proteins in regulation of the immune response. *Annu. Rev. Immunol.*, 2004; 22: 503–529.
- [152] Howard J. K., Flier J. S. Attenuation of leptin and insulin signaling by SOCS proteins. *Trends Endocrinol. Metab.*, 2006; 17: 365–371.
- [153] Guilherme A., Virbasius J. V., Puri V., Czech M. P. Adipocyte dysfunctions linking obesity to insulin resistance and type 2 diabetes. *Nat. Rev. Mol. Cell Biol.*, 2008; 9: 367–377.
- [154] Huttemann M., Muhlenbein N., Schmidt T. R., Grossman L. I., Kadenbach B. Isolation and sequence of the human cytochrome c oxidase subunit VIIaL gene. *Biochim. Biophys. Acta*, 2000; 1492: 252–258.
- [155] Lopez I. P., Milagro F. I., Marti A., Moreno-Aliaga M. J., Martinez J. A., De Miguel C. Gene expression changes in rat white adipose tissue after a high-fat diet determined by differential display. *Biochem. Biophys. Res. Commun.*, 2004; 318: 234–239.
- [156] Pellegrini M., Marzocca G. Obesity, insulin-like growth factors and colorectal cancer. *Panminerva Med.*, 2006; 48: 143–144.
- [157] Lillycrop K. A., Slater-Jefferies J. L., Hanson M. A., Godfrey K. M., Jackson A. A., Burdge G. C. Induction of altered epigenetic regulation of the hepatic glucocorticoid receptor in the offspring of rats fed a protein-restricted diet during pregnancy suggests that reduced DNA methyltransferase-1 expression is involved in impaired DNA methylation and changes in histone modifications. *Br. J. Nutr.*, 2007; 97: 1064–1073.
- [158] Achard V., Boullu-Ciocca S., Desbriere R., Grino M. Perinatal programming of central obesity and the metabolic syndrome: role of glucocorticoids. *Metab. Syndr. Relat. Disord.*, 2006; 4: 129–137.
- [159] Richardson B. Impact of aging on DNA methylation. *Ageing Res. Rev.*, 2003; 2: 245–261.
- [160] Jiang M. H., Fei J., Lan M. S., Lu Z. P., Liu M., Fan W. W., Gao X., Lu D. R. Hypermethylation of hepatic Gck promoter in ageing rats contributes to diabetogenic potential. *Diabetologia*, 2008; 51: 1525–1533.
- [161] Stenvinkel P., Karimi M., Johansson S., Axelsson J., Suliman M., Lindholm B., Heimbürger O., Barany P., Alvestrand A., Nordfors L., Qureshi A. R., Ekström T. J., Schalling M. Impact of inflammation on epigenetic DNA methylation – a novel risk factor for cardiovascular disease? *J. Intern. Med.*, 2007; 261: 488–499.
- [162] Franco R., Schoneveld O., Georgakilas A. G., Panayiotidis M. I. Oxidative stress, DNA methylation and carcinogenesis. *Cancer Lett.*, 2008; 266: 6–11.
- [163] Shahrzad S., Bertrand K., Minhas K., Coomber B. L. Induction of DNA hypomethylation by tumor hypoxia. *Epigenetics*, 2007; 2: 119–125.
- [164] van Dijk S. J., Molloy P. L., Varinli H., Morrison J. L., Muhlhausler B. S. Epigenetics and human obesity. *Int. J. Obes. e Lond.*, 2014.
- [165] Milagro F. I., Mansego M. L., De Miguel C., Martínez J. A. Dietary factors, epigenetic modifications and obesity outcomes: progresses and perspectives. *Mol. Aspects Med.*, 2013 Jul.-Aug.; 34(4): 782–812.

- [166] Sinclair, K. D., Allegrucci, C., Singh, R., Gardner, D. S., Sebastian, S., Bispham, J., Thurston, A., Huntley, J. F., Rees, W. D., Maloney, C. A., Lea, R. G., Craigon, J., McEvoy, T. G., Young, L. E., 2007. DNA methylation, insulin resistance, and blood pressure in offspring determined by maternal periconceptional B vitamin and methionine status. *Proc. Nat. Acad. Sci. U.S.A.*, 104, 19351–19356.
- [167] Shao W., Yu Z., Chiang Y., Yang Y., Chai T., Foltz W., Lu H., Fantus I. G., Jin T. Curcumin prevents high fat diet induced insulin resistance and obesity via attenuating lipogenesis in liver and inflammatory pathway in adipocytes. *PLoS ONE*, 2012;7:e28784.
- [168] Yun J. M., Jialal I., Devaraj S. Effects of epigallocatechin gallate on regulatory T cell number and function in obese v. lean volunteers. *Br. J. Nutr.*, 2010;103:1771–7.
- [169] Dolinoy D. C., Weidman J. R., Waterland R. A., Jirtle R. L. Maternal genistein alters coat color and protects Avy mouse offspring from obesity by modifying the fetal epigenome. *Environ. Health Perspect.*, 2006; 114: 567–72.
- [170] Choi K. M., Lee Y. S., Sin D. M., Lee S., Lee M. K., Lee Y. M., Hong J. T., Yun Y. P., Yoo H. S. Sulforaphane inhibits mitotic clonal expansion during adipogenesis through cell cycle arrest. *Obesity*, (Silver Spring). 2012;20: 1365–71.
- [171] Bujanda L., Hijona E., Larzabal M., Beraza M., Aldazabal P., García-Urkia N., Sarasqueta C., Cosme A., Irastorza B., González A., et al. Resveratrol inhibits nonalcoholic fatty liver disease in rats. *BMC Gastroenterol.*, 2008;8:40.
- [172] Baile C. A., Yang J. Y., Rayalam S., Hartzell D. L., Lai C. Y., Andersen C., Della-Fera M. A. Effect of resveratrol on fat mobilization. *Ann. NY Acad. Sci.*, 2011; 1215: 40–47.
- [173] Ceccarelli V., Racanicchi S., Martelli M. P., Nocentini G., Fettucciari K., Riccardi C., Marconi P., Di Nardo P., Grignani F., Binaglia L., Vecchini A. Eicosapentaenoic acid demethylates a single CpG that mediates expression of tumor suppressor CCAAT/enhancer-binding protein delta in U937 leukemia cells. *J. Biol. Chem.*, 2011; 286: 27092–27102.
- [174] Kulkarni A., Dangat K., Kale A., Sable P., Chavan-Gautam P., Joshi S. Effects of altered maternal folic acid, vitamin B12 and docosahexaenoic acid on placental global DNA methylation patterns in Wistar rats. *PLoS One*, 2011; 6: e17706.
- [175] Gu T., Gu H. F., Hilding A., Sjöholm L. K., Ostenson C. G., Ekström T. J., Brismar K. Increased DNA methylation levels of the insulin-like growth factor binding protein 1 gene are associated with type 2 diabetes in Swedish men. *Clin. Epigenetics*, 2013 Nov. 19; 5(1): 21.
- [176] Maghbooli Z., Hossein-nezhad A., Larijani B., Amini M., Keshtkar A. Global DNA methylation as a possible biomarker for diabetic retinopathy. *Diabetes Metab. Res. Rev.*, 2015 Feb.; 31(2): 183-9.
- [177] Tang L., Wang L., Ye H., Xu X., Hong Q., Wang H., Xu L., Bu S., Zhang L., Cheng J., Liu P., Ye M., Mai Y., Duan S. BCL11A gene DNA methylation contributes to the risk of type 2 diabetes in males. *Exp. Ther. Med.*, 2014 Aug.; 8(2): 459-463.
- [178] Voisin S., Almén M. S., Moschonis G., Chrousos G. P., Manios Y., Schiöth H. B. Dietary fat quality impacts genome-wide DNA methylation patterns in a cross-sectional study of Greek preadolescents. *Eur. J. Hum. Genet.*, 2015 May; 23(5): 654-62.

- [179] Dick K. J., Nelson C. P., Tsaprouni L., Sandling J. K., Aïssi D., Wahl S., Meduri E., Morange P. E., Gagnon F., Grallert H., Waldenberger M., Peters A., Erdmann J., Hengstenberg C., Cambien F., Goodall A. H., Ouwehand W. H., Schunkert H., Thompson J. R., Spector T. D., Gieger C., Trégouët D. A., Deloukas P., Samani N. J. DNA methylation and body-mass index: a genome-wide analysis. *Lancet*, 2014 Jun. 7; 383 (9933): 1990-8.

Chapter 67

GENETICS AND EVOLUTION OF ENTOMOPATHOGENIC FUNGI

**A. Garcia², M. Michel¹, S. Villarreal¹, F. Castillo³, E. Osorio⁴,
M. T. García-Ruiz⁶, F. Veana⁵, A. C. Flores¹ and R. Rodríguez^{1,*}**

¹Food Research Department, School of Chemical Sciences,
Universidad Autónoma de Coahuila, Saltillo, Coahuila, Mexico

²Department of Agricultural Parasitology,
Universidad Autónoma Agraria Antonio Narro,
Buenavista, Saltillo, Coahuila, México

³INIFAP-C. E. Saltillo, Saltillo, Coah, Mexico

⁴School of Engineering and Sciences,
Universidad Autónoma de Tamaulipas, Victoria Tamaulipas, México

⁵Proteomics and Molecular Biomedicine Laboratory,
Molecular Biology Division-IPICYT. San Luis Potosi. Mexico

⁶Universidad Autónoma Chapingo. Texcoco, Edo, de Mexico, Mexico

ABSTRACT

Entomopathogenic fungi have been mentioned as one of the best alternatives for insect pest control. These fungi cause insects death. More than 750 fungal species have been described infecting insects, some of the most utilized for insect control are: *Beauveria bassiana* and *Metarhizium anisopliae*. This is evidence of cosmopolitan distribution of entomopathogenic fungi and its evolutionary success, this type of fungi is related to a serie of interactions among fungi, plants, and insects. In this chapter, the main objective is to review and discuss the most recent information on genetics and evolution of entomopathogenic fungi. In this chapter are covered the following themes: Entomopathogens role in nature, Entomopathogenic fungi and their interactions with the insect immune system, Isolation and identification of entomopathogenic fungi, Genetic diversity among strains of entomopathogenic fungus, Genes involved in virulence of entomopathogenic fungi, Molecular phylogeny of entomopathogenic fungi and their

* Corresponding Author's Email: raul.rodriguez@uadec.edu.mx.

biogeographic implications, Evolution of entomopathogenicity in fungi, Genetic improvement of entomopathogenic fungi for insect biocontrol, Future trends and Conclusions.

Keywords: isolation, identification, diversity, virulence, pathogenicity, breeding, evolution, phylogeny

ENTOMOPATHOGENS ROLE IN NATURE

Entomopathogenic are employed in different fields, to control various insect species. These fungi are widely distributed throughout the fungi kingdom. Some insect pathogenic fungi have restricted host range, while others have a wide host range, with individual isolates to be more specific (Barreto et al., 2004, Hesketh, et al., 2009). Entomopathogenic fungi have been isolated from almost all regions of the earth and almost all types of soil. This is evidence of cosmopolitan distribution and its evolutionary success, involving a series of interactions between fungi, plants, insects and other sources nutrients in their environments. Entomopathogenic fungi are distinguished from other fungi and they transmitted directly by contact with susceptible guests, rather than having to be ingested to initiate infection (Cory and Ericsson, 2010). Early studies with entomopathogenic fungi occurred in early 1800 and focused on developing ways of diseases management that were ravaging the silkworm industry in France (Vega et al., 2009).

Biological Control

Entomopathogenic fungi are a large group of microorganisms which provide a great variety of services to agro ecological systems (Téllez et al., 2009). Among these are: ability to regulate pests and also to keep them at adequate levels. Entomopathogenic fungi have great potential as control agents (Hesketh, et al., 2009), forming a group with more than 750 species, scattered in the environment and causing fungal infections with arthropod populations; among the most important genera of entomophthogen fungi are: *Metarhizium*, *Beauveria*, *Aschersonia*, *Entomophthora*, *Zoophthora*, *Erynia*, *Eryniopsis*, *Akanthomyces*, *Fusarium*, *Hirsutella*, *Hymenostilbe*, *Paecilomyces* and *Verticillium* (López-Llorca and Jansson, 2001), while for FAO (2003), the most important genera are *Metarhizium*, *Beauveria*, *Paecilomyces*, *Verticillium*, *Rhizopus* and *Fusarium*. However, Oliveira et al., 2013 mentioned that entomopathogenic fungi group includes natural enemies of insect pests and the most important genera are: *Beauveria bassiana* (Bals.) Vuill., *B. brongniartii* (Sacc.) Petch, *Metarhizium anisopliae* (Metsch.) Sorokin, *Isaria farinosa* Holms. and *I. fumosorosea* Wize, which have a wide range of insect hosts. These fungi are particularly common in soils, where often cause epizootic disease in their hosts.

The entomopathogenic fungus *Verticillium lecanii* naturally infects a wide range of sucking pests such as thrips, whiteflies, aphids and mites; these are economically important pests of major horticultural crops. By increasing the incidence level of this entomopathogenic fungus could improve the control of target pests (Visalakshy et al., 2005). Efforts to develop methods of biological control of insect pest have been pursued; one of the best entomopathogen fungi is

M. anisopliae which uses a combination of enzymes and mechanical force to penetrate the host cuticle and hemolymph, this fungus produces chitinases for invasion and as one of the host-killing components (Barreto et al., 2004). Production of enzymes, mechanical mechanisms of invasion, improved mycelia growth and formation of conidia increase the infectivity of entomopathogenic fungi as biopesticide (Visalakshy et al., 2005). Several unexpected roles have been reported by fungi entomopathogens, among them their presence as endophytes, plant disease antagonists, colonizers and rhizosphere plant growth promoting fungi (Vega et al., 2009). Entomopathogenic fungi are very important ecological components since they perform insect control functions, which is very important because mishandling of pesticides by humans has promoted that insects pest become uncontrollable and some have develop resistance to pesticides.

Table 1. Entomopathogenic fungi used in biological control

Genus	Species
<i>Metarhizium</i>	<i>M. anisopliae</i>
	<i>M. flavoviridae</i>
<i>Paecilomyces</i>	<i>P. fumosoroseus</i>
<i>Rhizopus</i>	<i>Rhizopus spp.</i>
<i>Cordyceps</i>	<i>Cordyceps spp.</i>
<i>Culicinomyces</i>	<i>Culicinomyces spp.</i>
<i>Lagenidium</i>	<i>Lagenidium giganteum</i>
<i>Nomuraea</i>	<i>N. rileyi</i>
<i>Gliocladium</i>	<i>Gliocladium spp.</i>
<i>Lecanicillium</i>	<i>L. lecanii</i>
(<i>verticillium</i>)	<i>L. longispoum</i>
	<i>L. muscarium (Verticillium lecanii)</i>
<i>Beauveria</i>	<i>B. bassiana</i>
	<i>B. brongniartii</i>

Motta-Delgado and Murcia-Ordoñez, 2011.

Plants Manipulate Fungal Entomopathogens

There is scientific evidence that plants can influence behavior of certain groups of natural enemies, particularly parasitoids to increase herbivorous suppression and presumably increase the plant fitness. Plants can also influence entomopathogens in a similar manner. Entomopathogens fungi perhaps offer the best opportunities to become bodyguards of plant (Cory and Hoover, 2006). Any plant traits that improves entomopathogen fungal infection and shows a variation of genetic base could theoretically be selected as an adaptive response. For entomopathogenic fungi act as bodyguards' plant they must have benefits for plant processes (Cory and Ericsson, 2010).

The Ecology of Fungal Entomopathogens in the Rhizosphere

The ecology of fungal entomopathogens in the rhizosphere is an understudied area of insect pathology. Entomopathogenicity is a lifestyle that has emerged and lost several times in many lines of fungi (Bruck, 2010). The abundance of entomopathogenic fungi in the environment is correlated with the host abundance of species, and is facilitated by epizootic events (Cory and Ericsson, 2010).

ENTOMOPATHOGENIC FUNGI AND THEIR INTERACTIONS WITH THE INSECT IMMUNE SYSTEM

Entomopathogenic fungi are microscopic creatures living in the environment at expense of other organisms from where they get food; they are morphologically hyphae, which are cylindrical filiform structures 2 to 10 micrometers in diameter and several centimeters long, which together form a mycelium.

Classification

Two divisions are proposed: Myxomycota, those who are plasmodia, and Eumycota, which do not form plasmodia and have mycelia. Entomopathogenic fungi are located in the Eumycota division within five subdivisions; Mastigomycotina (form zoospores, or oospores), Zygomycotina (form zygospores), Ascomycotina (form ascospores), Basidiomycotina (form basidiospores) and Deuteromycotina (Ainsworth, 1973).

Important Species

According to Tellez et al., (1999), more than 750 species of entomopathogenic fungi in 100 genera have been described, the most important ones are *Metarhizium*, *Beauveria*, *Aschersonia*, *Entomophthora*, *Zoophtora*, *Erynia*, *Eryniopsis*, *Fusarium*, *Hirsutella*, *Hymenostilbe*, *Paecilomyces* (*Isaria*) and *Verticillium*; the more widely used species worldwide for biological control of insects are, *Metarhizium anisopliae* (33.9%), *Beauveria bassiana* (33.9%), *Isaria fumosorosea* (formerly known as *Paecilomyces fumosoroseus*) (5.8%) and *Beauveria brongniartii* (4.1%) (De Faria and Wraight, 2007).

Mode of Action

Entomopathogenic fungi act by contact through the cuticle, as they are able to enter and invade the insect completely, thus causing death by infection (Figure 1). According to different authors, the development cycle can be divided in two or three phases respectively; Parasitic and saprophytic (Elosegui, 2006) and infection, growth and reproduction (Boomsma et al., 2014);

adhesion and germination, penetration into the haemocoel and development of the fungus (Tanada and Kaya, 1993); adhesion, penetration and replication (Tellez et al., 2009).

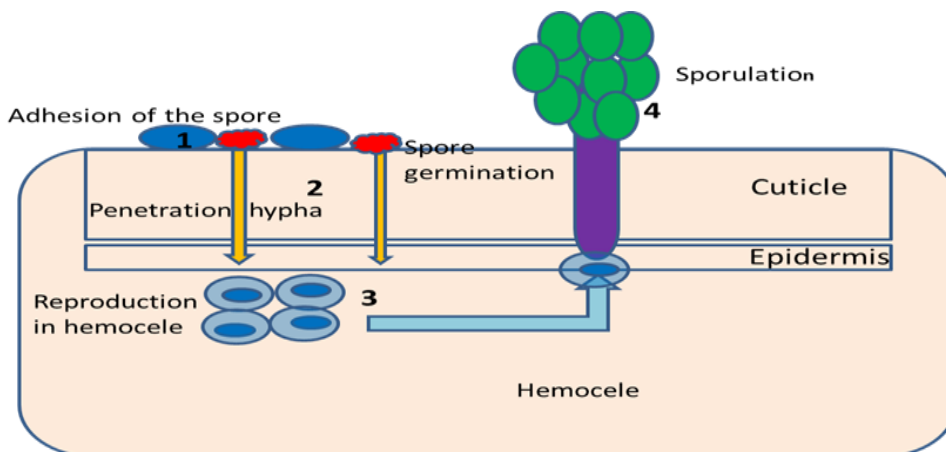


Figure 1. Scheme of entomopathogenic fungal infection process to a susceptible insect. 1. Adhesion and germination. 2. Penetration of the cuticle and epidermis. 3. Reproduction. 4. Sporulation.

Adhesion and Germination

The adhesin is a molecule produced by the fungus with adherent property and comes into operation for spores adhesion to insect cuticle, this molecule is recognized by specific receptors of glycoprotein nature. Once bound to the epicuticle, spore germination phase starts in which the spore emits one or more germinating appresoria tubes. This phase is not determined entirely by the percentage of spore germination, but rather by the way of germination, type of spore, aggressiveness of the fungus and host susceptibility.

Penetration

This process is achieved through the penetration tube (hyphae) by enzymatic degradation of the cuticle involving enzymes such as proteases, aminopeptidases, lipases esterases, chitinases and mechanical pressure haustoria which deforms the cuticular layer and subsequently sclerotic that breaks the membrane areas of the cuticle. This phase is determined by cuticle hardness, thickness and presence of antifungal substances or nutrients. Even at this stage, the insect has a last defense against foreign body that is within the insect, physiological reactions such as phagocytosis, cell encapsulation and formation of antimicrobial compounds.

Reproduction and Replication

When the fungus enters the haemocoel, evades the insect immune defense through a transformation of mycelium cells to yeast, which are called blastospores, accelerating the process of reproduction and dispersal, resulting in septicemia (Pérez, 2004; Tellez et al., 2009).

Among the physiological symptoms produced by fungal infection in insects are seizures, lack of coordination in their movements, altered or abnormal behavior (stops feeding, reduces movement) and total paralysis. Death is the end result of infection due to fluid loss, physical injury and toxicosis (Bustillo, 2001).

With the insect death, the fungus parasitic phase ends, and the saprofit phase begins, where large amounts of mycelia masses emerge to the outside through natural openings like mouth, spiracles and anus as well as the intersegmental abdomen and thorax regions thereof; fruiting on the body and generating inoculum to infect other insects (Cañedo and Ames, 2004).

Each of the different stages is influenced by internal and external factors. Although most insects have defense mechanisms against entomopathogens, they somehow manage to evade and penetrate these barriers; insect defenses are considered physical, cellular, humoral and even social behavior (Marmaras et al., 1996). Some factors influence or determine behavior of the spores, when they achieved and adhere to the epicuticle host insect, such as water, nutrients or debris on the surface of the exoskeleton, ions, fatty acids and even the developmental stage of insect (larva, nymph, pupa or adult) (Hassan et al., 1989). In this matter, the insect epicuticle is well known for its hydrophobic feature, and sets the pattern as an excellent reservoir for adhesion and spore germination; however, a combination of external factors such as temperature, humidity, radiation and insect's habits determine the full or no adhesion of the fungal spore. The main ways of fungal transmission or infection that an insect is exposed are by direct contact on the ground, scattered spores in the air, contact with bodies of other infested insects and the infection routes are the oral cavity and spiracles; the entomopathogenic fungi liquid solutions that are used in modern agriculture for biological control of insects, naturally contributed to the regulation of a wide range of pests affecting world agriculture.

ISOLATION OF ENTOMOPATHOGENIC FUNGI

The fungi species of major interest for its potential as insect pathogens are: *Beauveria bassiana*, *Metarhizium anisopliae*, *M. flavoviride*, *Nomuraea rileyi*, *Lecanicillium lecanii*, *Hirsutella thompsonii*, *Aschersonia aleyrodis*, *Paecilomyces* spp., and the fungi belonging to the Entomophthorales order (Zoophthora, Entomophthora and Entomophaga). For isolation of entomopathogenic fungi, some artificial media are used such as: sabouraud dextrose agar, potato dextrose agar (PDA) and malt extract agar (MEA) (Elósegui et al., 2006). In addition, some specific media to isolate entomopathogenic fungi have been developed (Inglis et al., 2012). Different methodologies for entomopathogen fungi isolation have been developed.

Isolation by Serial Dilution

This method involves placing a mycosed insect in a vessel containing 10 ml of Tween 80 (0.01%) in sterile distilled water. The resulting suspension has to be stirred for 1 minute to release conidia insect body. As a result, there is obtained a concentrated suspension or stock solution of inoculums with other particles. The stock solution can be used to prepared serial dilutions. The first dilution (10^{-1}) can be obtained by mixing 1 ml of the stock solution in a tube containing 9 ml of Tween 80 (0.01%) in sterile distilled water. This operation is repeated

several times until a dilution series (10^{-1} to 10^{-6}) are obtained. For inoculating and obtaining fungus, the latest dilutions (10^{-4} , 10^{-5} , 10^{-6}) are used. Subsequently, they are inoculated in Petri dishes with 100 μ L of PDA medium and incubated at $25 \pm 1^\circ\text{C}$ for 5 to 7 days. Once the fungi are grown, the next step is to identify them, especially some of interest (primarily *Metarhizium anisopliae*, *Beauveria bassiana* and *Entomophthora* spp.). Identification can be accomplished using a compound microscope and dichotomous keys based on morphological characteristics of reproductive structures, such as conidiophores, conidia and phialides.

Direct Isolation of Entomopathogenic Fungi from Insects

For direct isolation, it is necessary to monitor the culture of interest. For this, dead insects, suspected of fungal infections with mummification symptoms, or sporulation of fungal nature in the cuticular surface have to be collected (Diaz et al., 2008). Samples are handled with soft tweezers and placed into microcentrifuge tubes (depending on the size of the insect). In aseptic conditions, insects are disinfested in a solution of 2% sodium hypochlorite for 3 minutes and rinsed three times with sterile distilled water. Such type of isolation can be done in two ways: first, by scraping with a bacteriological loop all fungal particles present in a disinfested insect, and then passing them in Petri dishes with PDA culture medium through the technique of rifling; second, with a dry, sterile forceps, take the sporulated insect previously disinfested and shake it with vertical and horizontal movements on the surface of Petri dishes with PDA medium and incubated at $25 \pm 1^\circ\text{C}$ for 5 to 7 days. Once fungal growth is observed, proceed to the characterization.

Isolation of Entomopathogenic Fungi from Soil

Insect trap technique which use in most cases *Galleria mellonella* (Lepidoptera: Pyralidae) and *Tenebrio molitor* (Coleoptera: Tenebrionidae) (Zimmermann 1986); these insects are reproduced under lab conditions besides they are susceptible to fungi. Soil samples are collected from depth of 0-20 cm and placed in properly labeled plastic bags. At each site a subsample is taken, in order to form a composite sample of 1 kg. From each sample of homogenized soil, 500 g are taken and placed in a plastic container of $14 \times 18 \times 6$ cm with seven larvae of *G. mellonella* or *T. molitor* of last stage and incubated at $25 \pm 1^\circ\text{C}$ for 7 days. Subsequently, the dead larvae with symptoms of fungal infection are placed in a moist chamber. Isolation and purification of the fungi is performed in a laminar flow chamber for direct transfer of conidia and/or mycelia in Petri dishes containing PDA medium with 100 μ g streptomycin sulfate and 10 μ g of tetracycline hydrochloride (antibiotics) (Khudhair et al., 2014). After incubation, observation of fungal growth is performed to identify them with keys based on morphological characters of reproductive structures, such as conidiophores, conidia and phialides.

MORPHOLOGICAL CHARACTERISTICS OF ENTOMOPATHOGENIC FUNGI

Beauveria bassiana

Colonies on PDA at the beginning are white and fluffy, become beige slightly and dusty and eventually sporulate. On the reverse side, they are slightly yellow. This fungus has abundant conidiophores with dense groups of phialidic cells, conidiogenous and globose cells, both bottle shaped with intermediate forms, $1.5\text{--}3.0$ ($2.7\text{ }\mu\text{m}$) $\times$ $2.0\text{--}3.0$ ($2.2\text{ }\mu\text{m}$) and hyaline conidia, sometimes with a very definite globose to ellipsoid apex $1.5\text{--}3.0$ ($2.3\text{ }\mu\text{m}$) $\times$ $1.5\text{--}3.0$ ($2.1\text{ }\mu\text{m}$).

Metarhizium anisopliae

This fungus has white and cottony colonies on PDA, in the beginning turn greenish yellow and finally dark olive green and crusted with areas with abundant white aerial mycelium. The reverse of the Petri dish is intense yellow. Also it has conidiophore typical pattern with 2-3 whorled branches each, with dark olive green tones but, clarifies the apex. Others characteristics of this fungus are conidiogenous cells $6.0\text{--}10.0$ ($8.1\text{ }\mu\text{m}$) $\times$ $2.0\text{--}2.5$ ($2.1\text{ }\mu\text{m}$), subhyaline to slightly green, and cylindrical to slightly ellipsoidal conidia, $5.0\text{--}8.5$ ($6.6\text{ }\mu\text{m}$) $\times$ $2.0\text{--}3.0$ ($2.4\text{ }\mu\text{m}$).

Paecilomyces lilacinus

Colonies of this fungus on PDA at the beginning are downy white, then they turn in gray to purple depending on age and shown an irregular aerial mycelium. On Petri dish reverse, it is observed a red color. Alone or in groups, conidiophores form synnemata with whorls of up to six phialides, which with age become wrinkled. Typical conidiogenous cells, $7.0\text{--}13.0$ ($9.5\text{ }\mu\text{m}$) $\times$ $1.5\text{--}3.0$ ($2.8\text{ }\mu\text{m}$) have catenulate conidia, light gray-violet, ellipsoidal to fusiform, some with an apex, $1.5\text{--}3.0$ ($2.2\text{ }\mu\text{m}$) $\times$ $1.5\text{--}3.0$ ($2.4\text{ }\mu\text{m}$).

***Nomurae rileyi* (*Metarhizium rileyi*)**

This fungal species grows slowly in culture medium (Sabouraud-maltose agar) showing initially a pale green colony, changing its conidia into malachite green or olive green as they mature. Vegetative hyphae are septate, smooth, hyaline or slightly pigmented wall; its reproductive structure is a septate conidiophore born from hyphae forming dense clusters with groups of 2 or 3 compacted phialides around the conidiophore, the phialides are cylindrical with occasionally flared base and very short neck or absent of $4.7\text{--}6.5 \times 2.3\text{--}3\text{ }\mu\text{m}$. Conidia in chain are ellipsoid or sometimes cylindrical of $3.5\text{--}4.5 \times 2\text{--}3.1\text{ }\mu\text{m}$ (Samson, 1974). Mainly limited to the Lepidoptera order: *Spodoptera frugiperda* and *Helicoverpa armigera* (Adsure

and Mohite, 2015), and at least two species of beetles: *Hypera punctata* (Fabricius) and *Leptinotarsa decemlineata* (Kroatz).

Lecanicillium (=Verticillium) lecanii

In potato-carrot agar, the colonies of this fungus are fluffy and white. Solitary or whorled and prostrated conidiophores, carrying a hyaline apical mass of subglobose, oval, falcate, fusiform and unicellular subcolumnar conidia, non-adhesive and do not exhibit latent structures (Zare et al., 2001). Conidia with dimensions between 6.2 to 11.3 μm diameter, and phialides from 11.2 to 30 $\times$ 1.7-2.5 μm , the conidia are 3.0 to 4.0 μm width and 5.8 to 10.5 μm length (Sugimoto et al., 2002). This fungus has a wide host range: Hemiptera insects (*Toxoptera aurantii*, *Aphis gossypii*, *Myzus persicae* and *Bemisia tabaci*), Coleoptera, Diptera and Lepidoptera orders and mites of *Tetranychidae* family.

Isaria fumosorosea

This fungal specie presents yellowish hyaline, septate hyphae with thin walls. Most *I. fumosorosea* isolates have whorled or irregularly branched, carrying in its terminal part at each branch phialides groups, which can also be lonely. Its phialides have a cylindrical or swollen basal portion, often abruptly tapering to form a very noticeable neck. Conidiophores carry conidia chains; these are hyaline, unicellular and ovoid (Bustillo, 2001). This entomopathogenic fungus is characterized by bright colors like white, yellow, pale green, pink, red or purple. It has hyaline to yellowish, septate hyphae with smooth walls. The conidiogen structure is a synnema or monosynnema consisting of compacted hyphae, irregular and whorled conidiophores with terminal branches, which clusters widened bottle-shaped with a distinct neck where the conidia are born, which grow as chain in a basipetal form by a cell, rarely two, hyaline or slightly pigmented with smooth, echinulate walls or more forms (Berlanga, 1997; Hernandez, 1997). *I. fumosorosea* is used to control mainly *Bemisia tabaci* and *Diaphorina citri* (Flores et al., 2013).

Hirsutella thompsonii

This fungal specie presents septate, fine and hyaline mycelium with a diameter from 1.7 to 3.3 μm . Hyphae usually develops in a simple form except in species producing synnema, sporulates moderately, its conidiogenous cell is a gray phialides, from 10 to 15 μm long, originating from the sides of hyphae. The conidia are spherical from 2 to 4 μm diameter with verrucose surface (McCoy, 1981). Growth in malt agar medium, has an abundant mycelial mass and subsequently acquires light gray shades between gray and bluish gray with presence of hyaline droplets of liquid and white synnemata over 10 cm in length (Cabrera and Dominguez, 1987). *H. thompsonii* is specific to mites, mainly Eriophyidae and Tetranychidae, and it has been reported on plowman of citrus *P. oleivora* in the United States, China and Cuba (Samson et al., 1980); *E. gerreronis* in coconut in Cuba and Mexico and other citrus mites and cherry,

in laboratory *Tetranychus cinnabarinus*, *Eutetranychus orientalis* and *Tetranychus turkestan* are susceptible (Gerson et al., 1979).

Aschersonia aleyrodis

Colonies of this fungal specie on PDA grow slowly reaching 35 mm in diameter in three weeks at 23°C, filamentous white to yellowish white hyphae, sometimes with a grayish-yellow color with peripheral circle. Conidia masses of variable color, light, intense or reddish orange. Most conidia measure $10\text{--}16 \times 1.0\text{--}1.5 \mu\text{m}$, thick-walled hyphae, or in a whorl of 2-6 from the end of the thick-walled hyphae; conidiophores with multiple ramifications of 35-65 μm , some elongated phialides that reach paraphyses (Liu et al., 2006). *Aschersonia* species form a simple small stroma which consists of a dense interweaving hyphal of thick wall forming a structure or a fruiting body called pycnidium, in which conidia are produced. Different species of this fungus can be distinguished by the color of the stroma and conidia, which presents a red, brown, orange or yellow tonality. The conidia have little variation in shape, being generally fusoid or oval with pointed ends (Mains, 1959). *A. aleyrodis* is reported by infect midge benches of agricultural importance highlighting *Dialeurodes citrii*, *D. citrifolii*, *Bemisia tabaci* (Genn), *Trialeurodes floridensis* (Quaintance) *T. vaporariorum* (Westw) *Aleurocanthus woglumi* (Ashby) and *Aleurothrixus floccosus* (Mask) (Hernandez et al., 2005).

Entomophthora muscae

The species are subspherical, ellipsoidal, kidney-shaped or rounded bodies hyphae measuring $24\text{--}54 \times 17\text{--}41 \mu\text{m}$ with 10 to 24 nuclei (average 15-19). Conidiophores are unbranched and contain 10 to 27 nuclei (average 15 to 20). Primary conidia average measured is $27 \text{ to } 31 \times 20 \text{ to } 24 \mu\text{m}$ with a length/diameter ratio of 1.20-1.32 and 10 to 27 nuclei (average 15-20) relationship. The secondary conidia average measured is $19\text{--}24 \times 15\text{--}19 \mu\text{m}$ with a length/diameter ratio of 1.20-1.33 (Keller et al., 1999). *Bemisia tabaci*, *Aphis fabae*, *Acyrtosiphon pisum*, *Delia radicum*, *Pollenia angustigena*, *Coenosia tigrina*, *Musca domestica*, *Scatophaga stercoraria*, Syrphidae and Anthomyiidae. (MacLeod et al., 1976; Jensen et al., 2006) are controlled by *E. muscae*.

MOLECULAR IDENTIFICATION OF ENTOMOPATHOGENIC FUNGI

There are several molecular techniques for DNA analysis and search for information to identify entomopathogenic fungi. Among the most important techniques are RAPD's (Random Amplification of Polymorphic DNA) or AFLP's (Amplification Length Polymorphism Fragments). For these techniques, combinations of primers (oligonucleotides) that produce different banding patterns can differentiate DNA from different individuals (Bulat et al., 2000). On the market are available kits with universal oligonucleotides for microorganisms and plants, which can be used for fungi. It is also possible to design PCR protocols to amplify the internal

transcribed spacer regions (ITS) which can be cut using enzymes. It is known that ITS are highly conserved in all organisms, so small changes in them can be used for classification. There are PCR-based primers used for characterization of specific entomopathogenic fungi such as *B. brongniartii* (Enkerli et al., 2001), *B. bassiana* (Rehner and Buckley, 2003), *M. anisopliae* (Enkerli et al., 2005), as well as *Isaria fumosorosea* (Gauthier et al., 2007).

Ribosomal RNA (rRNA) and its template ribosomal DNA (rDNA) sequence comparisons are useful for estimating both close and distant relationships among fungi. Small subunit rRNA gene (18S rDNA and 5.8S) sequence divergence has contributed in phylogenetic analysis among distantly relaxed taxa; however, those sequences are generally too conserved and attempts have been made to utilize the more variable domains of the large subunit (26 or 28S) for species identification and detection. Since 2012, the internal transcribed spacers of the nrDNA (ITS) are accepted as the official DNA barcode for fungi (Schoch et al., 2012). The high number of copies of the ITS per cell, in particular, makes it an attractive target for diagnostics and it can be detected with great sensitivity. Nevertheless, a secondary identification marker is sometimes needed and calmodulin (CaM), β -tubulin (benA) and the RNA polymerase II second largest subunit (rpb2) have been suggested for this purpose (Samson et al., 2014). The rpb2 gene is not easy to amplify, rendering its use as secondary identification marker frustrating. In contrast, benA is easy to amplify, but has been reported to vary in the number of introns and PCR sometimes results in the amplification of paralogous genes (Peterson, 2008; Hubka and Kolarik, 2012). In addition, the CaM sequence database is almost complete for all accepted species.

GENETIC DIVERSITY AMONG ENTOMOPATHOGENIC FUNGAL STRAINS

Diversity of entomopathogenic fungi is determined in an ecological community by ecological indices based on morphological characteristics. However these indices are not enough to determine the genetic diversity within communities (Garrido-Jurado et al., 2015). Genetic diversity and population structure are required to understand and direct the use of fungi as biological control agent. The overall genetic diversity in entomopathogenic fungi resulted from the genetic variation within geographical populations. This genetic diversity correlates with presence or absence of recombination.

Molecular tools can be used to identify the genetic differences between and among fungal species (Garrido-Jurado et al., 2015) which can allow understanding population structure and gene flow within communities. The genetic variability has been studied using several molecular markers such as RFLPs (length polymorphisms of restriction fragments), RAPD (DNA polymorphisms amplified random), minisatellites, AFLP (Amplified Fragment Length Polymorphism), SCAR (sequence-characterized amplified region), SSR (simple sequence repeats or microsatellites), ISSR (inter-simple sequence repeat), SSCP (single-strand conformation polymorphism), TGGE (temperature gradient gel electrophoresis) and DGGE (denaturing gradient gel electrophoresis) (Enkerli and Widmer, 2010; Guo, 2010). Specific DNA sequences are used as genetic markers and allow to discriminate within taxonomic groups a community structures and the phylogenetic relationship, or characterization and monitoring of strains in different environments can be studied (Enkerli and Widmer, 2010). Genetic

variability in several species of *Beauveria* sp, *Metarhizium* sp, *Paecilomyces* sp. and other entomopathogenic fungi has been evaluated by several molecular techniques.

Garrido-Jurado et al., (2015) analyzed the diversity of entomopathogenic fungi during four seasons in olive orchards, holm oak reforestation, holm oak dehesa and sunflower plantation. In this study were identified entomopathogenic fungi species such as: *Beauveria amorpha*, *B. bassiana*, *B. pseudobassiana*, *B. varroae*, *Metarhizium brunneum*, *M. guizhoense*, *M. robertsii*, *Paecilomyces marquandii* and *lilacinum* using gene EF-1 α sequences and ISSR. The Shannon–Wiener index was used to measure diversity and abundance in the eco-system. The ISSR and ITS analysis revealed a high degree of polymorphism among the different isolates with 80% of polymorphic bands particularly in all species of *B. bassiana*; which was dominant in all cropping system and all seasons. In some studies, it has been found differences among communities of entomopathogenic fungi in different types of soils such as *Beauveria bassiana* in forest soil and *Metarhizium anisopliae* in cultivated soil; these differences reflected a high level of DNA polymorphism (Garrido-Jurado et al., 2015).

***Beauveria* spp**

In this fungal specie were used molecular markers such as RAPD to determine the genetic diversity among isolates from different geographical regions of Chile. The RAPD analysis indicated high genetic diversity with 43% similarity between isolates, indicating that genetic diversity was not associated with the geographical origin of the isolates (Becerra et al., 2007). Fernandes et al., (2009) evaluated the genetic diversity among 49 species of *Beauveria bassiana* from Brazil and 4 species of *Beauveria* spp. from USA. The isolates were analyzed by ITS (including the 5.8S gene) and AFLP and revealed considerable genetic variability among *B. bassiana* isolated from different geographical regions of Brazil, and differences between Brazilian and USA *B. bassiana* isolates. Other study reported the genetic diversity in *Beauveria*, where were used 82 polymorphic AFLP loci; several alleles were found in the *Beauveria* strains as alleles 2 and 5 in the locus MDH in all strain, allele 3 in the loci PGM and 6PGD in 96% of *B. bassiana*, allele 2 in the locus G2D in 90%, allele 2 and 3 in the loci FUM and PK in the 88% and allele 1 in IDH to 86%. The geographical distance between entomopathogenic fungi demonstrated notable genotypic variation but no correlation was observed according to the host (Fernandes et al., 2009).

SSR markers are often highly polymorphic and can be used for genotyping of different fungal strain and monitor entomopathogenic fungus in different environments. Reineke et al., (2014) monitored *Beauveria bassiana* in different environments using SSR markers and found that SSR were important tools to monitor fungal strain. The microsatellite primers used were Ba01, Ba02, Ba08, Ba12 and Ba13. *B. bassiana* strains from different origin shared the same alleles at 5 SSR loci, among them 18 isolated strains, 3 strain from India and 2 from the USDA culture collection.

Meyling et al., (2009) identified *B. bassiana*, *B. brongniartii* and a monophyletic group nearby from *B. bassiana*. The entomopathogenic fungi were collected from insects sampled in cultivated field and soils in Denmark; for the phylogenetic analysis were used some markers identified from *B. bassiana* such as Ba06, Ba08, Ba12, Ba13, Ba15, Ba18, Ba21, Ba26, Ba27 and markers isolated from *B. brongniartii* such as Bb1F4, Bb2A3, Bb2F8, Bb4H9, Bb5F4 and Bb8D6. The same authors determined multilocus microsatellites for five phylogenetic

species of *B. bassiana*; these species co-exist on the sites of collect. The five population showed multiple allelic differences of microsatellite loci by EF-1 α phylogenetic analysis; the Eu_3, Eu_4, Eu_5 and Eu_6 populations had low allelic variation, only show 1, 2, 5 and 2 polymorphic loci respectively, while the population Eu_1 had the greatest genotypic variability with 12 polymorphic loci. This variability could be attributed to recombination; also they found that *Beauveria* diversity was highest in the semi natural habitat because of great abundance and diversity of insect hosts. The semi-natural habit had increased humidity and environmental stability compared with the agricultural field.

The virulence of 21 isolates of *B. brongniartii* and 2 isolates of *B. bassiana* was evaluated against adults and larvae of *Schizonycha affinis* and *Hypopholis sommeri*, respectively. The isolates of *B. brongniartii* showed low level of genetic diversity using microsatellite markers. In this study, it was found that 21 isolates of *B. brongniartii* represent 17 haplotypes and that were genetically close. The isolates of *B. brongniartii* may vary in their virulence with a mortality of 50.1–95% to *S. affinis* and 39–74% of mortality to *Tenebrio molitor* (Goble et al., 2015).

***Metarhizium* spp**

The fungus is usually isolated from soil, but the distribution of species in soil is unknown. The phylogenetic relationship and genetic diversity of *Metarhizium* spp. isolated from *Schiodtella formosana* were analyzed by ISSR. The 86% of the generated bands were polymorphic and the polymorphic loci P was from 70 to 100%. The *Metarhizium* populations were differentiated into 2 clades, reflecting a proportion of 79.2% of gene differentiation within the populations. The isolates of *Metarhizium* spp. were identified based on ITS sequences in *M. anisopliae* var. *anisopliae* and *M. robertsii* (Luan et al., 2012).

In diverse Asia and European countries was determined the genetic diversity of *Metarhizium anisopliae* by microsatellite markers (SSR) and ITS rDNA regions (ITS-1, ITS-2 and 5.8S gene regions). Using SSR was found that existed a low level of polymorphism in *M. anisopliae*, gene diversity was 0.37 and the result of ITS-rDNA sequence confirmed that SSR had minimal genetic variation in all isolates with a genetic distance of 0.34%, the closely related genotypes of *M. anisopliae* were dominant in different ecosystems in regions of China (Freed et al., 2011).

Metarhizium spp. were obtained from bulked soil samples of agricultural field and surrounding hedgerow in single agro ecosystem in Denmark. The genotypic diversity of *Metarhizium* isolates was analyzed by SSR and it was found co-occurrence of four species in the soil of the smaem agro ecosystem. To determine the different genotypes were used 18 SSR markers: Ma145, Ma325, Ma307, Ma2049, Ma2054, Ma2055, Ma2056, Ma2057, Ma2060, Ma2063, Ma2069, Ma2070, Ma2077, Ma2089, Ma2283, Ma2287, Ma2292, and Ma2296. *Metarhizium brunneum* was most frequent (78.8%) followed by *M. robertsii* (14.6%), while *M. majus* and *M. flavoviride* were infrequent (3.3% each). Based on SSR analysis were identified 5 genotype of *M. brunneum* and 6 genotypes of *M. robertsii* indicating particular adaptations to these soil environments (Steinwender et al., 2104).

SSR markers were used to characterize polymorphisms in one collection of 65 strain of *Metarhizium*. These strain included *M. anisopliae*, *M. brunneum*, *M. guizhouense*, *M. lepidiotae*, *M. majus*, *M. pingshaense* and *M. robertsii*, represent more of 75% of the strain

isolated and were amplified using 21 markers (15 polymorphic) based on EF1 α sequence (Mayerhofer et al., 2015). Bischoff et al., (2009) evaluated phylogenetic relationships within *M. anisopliae*, *M. taii*, *M. pingshaense* and *M. guizhouense* by EF-1 α , RPB1, RPB2, IGS and β -tubulin gene regions.

***Paecilomyces* spp**

Genetic relationship of several isolates of *Paecilomyces fumosoroseus* from various host species and geographical locations was investigated; the isolates of *P. fumosoroseus* from *Bremisi tabaci* indicated the association of genotypes with this insect. The analysis of allelic variability by microsatellite loci allowed group in two lineages, one lineage included genotypes related to *B. tabaci* in America and other lineage distributed in Asia (Gauthier et al., 2007). Other specie of *Paecilomyces* in which genetic diversity was evaluated by SSR markers was *Paecilomyces variotii*. The isolates were obtained from pistachio gardens and were identified 22 alleles; this allowed grouped all the isolates into 8 groups with high genetic variability but not demonstrated the relationship between isolates and the geographical regions of collect; the SSR primers PfrBtB04 and PfrBtD05 amplified alleles at numerous loci (Ebrahimi et al., 2015). In addition, genetic diversity of *P. variotii* was examined by SSR analysis using the primers PfrBtD11a, PfrBtD11b, and PfrBtB04 and RFLP analysis. The polymorphism information content and marker index were calculated; in this study was evaluated the amount of detectable polymorphism in the isolates, RFLP produced 20 loci con 90% of polymorphism and SSR produced 32 loci with 37.5% of polymorphism (Rostami et al., 2015).

Paecilomyces liliacinus is other specie in which was analyzed the degree of polymorphisms and haplotype diversity. Li et al., (2013) analyzed 80 specimens of *P. liliacinus* and observed diversity among six gene fragments and the haplotype h18 which was the most common in 20 isolates. Besides, it was found an extensive gene flow among strains. The phylogenetic trees based on the six gene fragments allowed grouping the strains in four clades. Evaluation of composition and dynamic of entomopathogenic fungi populations may improve the effects of entomopathogenic fungi for biocontrol. In addition, resolution of genetic markers allows description of population structures and help to understand local migration and dissemination patterns.

GENES INVOLVED IN VIRULENCE OF ENTOMOPATHOGENIC FUNGI

Entomopathogenic fungi comprising a group of approximately 100 genera, however, a low percentage is used for pest control. It is important to know that genes involved in the virulence of entomopathogenic fungi, with the objective of genetic manipulation and decrease of agricultural losses (Sansores-Lara et al., 2014). Most fungal species used for biological control are *Beauveria bassiana*, *Metarhizium brunneum*, and *M. anisopliae*.

The virulence of entomopathogen fungi involves adhesion, germination, differentiation and penetration steps. Particularly, genes involved in virulence from *Beauveria bassiana* are distributed in the seven step of infection, as follow: host adhesion, germination, cuticle

degradation, growth as blastospores, host colonization and killing, immune response interactions and hyphal extrusion and conidiation. During this procedure, 53 genes are expressed and correspond to hydrophobin, mitogen-activated protein kinase (MAPK), protein kinase, subtilisin-like protease, chitinase mannitol-1-phosphate dehydrogenase, regulates calcium sensor (acidification regulator), beauvericin toxin, bassianolide toxin, bassiacridin, among others (Valero-Jiménez et al., 2016).

In cuticle degradation and hyphal extrusion and conidiation, a chitinase gene (Bbchit1) is related. (Fang et al., 2005). The Bbchit1 from *B. bassiana* has a length of 1047 bp and encode for a protein of 348 amino acids, where the 28 amino acid residues corresponding to signal peptide. Later, a hybrid protease was development by fusion of a chitin binding domain BmChBD from *Bombyx mori* chitinase to the C-terminal of CDEP-1, a subtilisin-like protease from *B. bassiana*. The hybrid was able to increased fungal virulence compared with the wild-type native protease (Fan et al., 2010). Since glycosidase hydrolase family 18 of the *M. anisopliae* has been studied for importance in virulence. Family 18 and 19 are responsible of hydrolysis of chitin present in the insects, and have been identified 24 genes belonging to family 18, including chit1, chi2 and chi3 (Junges et al., 2014).

Other gene that regulates fungal conidiation, virulence and stress control in *B. bassiana* is Ras homologs (Ras1, Ras2 and Ras3). Ras1 and Ras2 genes have length of 648 and 708 bp, which encodes for 216 and 236 amino acid, with molecular mass of 24.3 and 26.2 kDa, respectively (Xie et al., 2013). The Ras3 GTPase gene is localized in plasma membrane and is involved in the HOG1 pathway necessary for osmoregulation. By the way, the Ras3 can be regulating conidiation, germination, multi-stress tolerance and virulence (Guan et al., 2015).

Two hydrophobins encoded by hyd1 and hyd2 genes, which have been detected in *B. bassiana*, can be involved in cell surface hydrophobicity, adhesion, virulence and constitute the spore coatrodlet layer in fungi (Zhang et al., 2011). The hydrophobin gene (*hyd1*) promoter (GenBank ID:GU936631) has a size of 1798 bp, however a truncated 1290 bp fragment (*Phyd1-t1*) was used for eGFP expression (enhanced green fluorescence protein) which was decreased. In addition, an insect midgut-specific toxin gene (*vip3Aa1*) was expressed in transgenic strains of *B. bassiana* (BbHV8, BbV28 and Bb2860), where BbHV8 strain produced nearly 10-fold more toxin molecules in conidia and thus killed *Spodoptera litura* larvae more rapidly and effectively irrespective of normal or heat-killed conidia ingested (Wang et al., 2012). On the other hand, in *Metarhizium brunneum*, three genes with relationship virulence (HYD1, HYD2 and HYD3) were reported, which codified for hydrofobines class I and class II (form rodlet structures at interfaces and do not, respectively). These genes were deleted in the fungi and conidia or blastospores of each mutant were applied versus *S. exigua* larvae. Reductions in virulence and delayed mortality were presented in comparison with wild-type strain (Sevim et al., 2012).

Adhesin genes (MAD1 and MAD2, adhesion-like protein) from *M. anisopliae* are responsible for anchor to insect cuticle and facilitate the colonization. Full length genes are 2151 bp and 306 bp that encode for 717 and 306 amino acids (Wang and St-Leger, 2007). Other entomopathogenic fungus enzyme related with cell-surface localization and host adhesion is glyceraldehyde-3-phosphate dehydrogenase (GAPDH) with 338 amino acids. However the gene (*gpdh1*) is a full length of 1230 bp (GenBank id: EF050456) (Broetto et al., 2010).

A mitogen-activated protein kinase (MAPK) regulates the fungal development, growth, and pathogenicity over insects (Luo et al., 2012). According to Koduru et al., (2014), protein osmosensor (MOS1), laccase, superoxide dismutase (Bb-SOD2), mutation in β tubulin genes

and adenylate cyclase, are related in abiotic stress and virulence of insect pathogenic fungi. A MAPK gene was identified in *B. bassiana* (Bbmkp1), which is essential for spore adhesion, formation of appressoria and ability to cross the insect cuticle both during infection and exit after insect death. Other gene with same action is an HOG (high osmolarity glycerol response) kinase. Both genes affect the transcript levels of hydrophobin encoding genes (Koduru et al., 2014). Bbslt2 gene (GenBank: JF932503.1) is required for full virulence in *B. bassiana*. This gene has a length of 1257 bp, which encode for a protein of 418 amino acids that correspond to Slt2 family MAPK (Luo et al., 2012).

Recently, a correlation between Pr1 and Pr2 gene associated to virulence of *M. anisopliae* strains has been established. These genes encode for proteases such as subtilisin-like enzyme (Pr1) and trypsin-like enzymes (Pr2), these enzymes are important for insect cuticle degradation. A bioassay of mortality was performed with four strains of *M. anisopliae* (6342, 6345, 6347 and 798), which possess 11 pr1 genes, versus adults of *Prosapia* sp. and larvae of *Spodoptera exigua*. The highest mortality (73%) on this insect was generated by the 6342 strain, which showed the maximum Pr2 enzymatic activity. However, subtilin and trypsin production are not indispensable for pathogenicity (Rosas-García et al., 2014). In addition, Ifa-Pr1A gene from *Isaria farinose* (GenBank id: JN014832) was detected using the RACE technique with a length of 960 bp, which encode for a protein of 320 amino acids. The Ifa-Pr1A transcripts exhibit 58 800-fold at 12 h post-induction. Therefore, the gene is considered as major cuticle-degrading proteases in the studied fungus. Besides, the Ifa-Pr1A gene is an important virulence factor for the development of biopesticides (Wang et al., 2013). An *in vivo* infection study of insects (*Galleria mellonella*, *Eupoecilia ambiguella* and *Cactoblastis cactorum*) infected by *B. bassiana* revealed the expression of three genes: subtilin (Pr1h), exocyst component (Sec15) and EC391425 with a length of 98, 160 and 120 bp (Galidevara et al., 2016). The same gene from *M. acridum* and *M. anisopliae* has been described as heavily involved in the initial steps of fungal invasion of arthropod-host cuticle (Leão et al., 2015; Golo et al., 2015).

During immune response interactions, a superoxide dismutase is expressed in *B. bassiana* 2860. The BbSod2 gene (630 bp) from *B. bassiana* 2860 encodes a 209-aa protein with a theoretical molecular mass of 23.1 kDa and isoelectric point of 7.14. (Xie et al., 2010). Secondary metabolites tenellin (BbtenS), beauvericin (BbbeaS) and bassianolide (BbbslS) were measured by quantitative reverse transcription real-time PCR (qRT-PCR) during infection of *Triatoma infestans* (Chagas disease) by the entomopathogenic fungus *B. bassiana*. BbtenS and BbbeaS were highly expressed at days 3 and 12 post-treatment. In blastospore-injected insects, BbtenS and BbbeaS expression peaked at 24h post-injection and were also highly expressed in insect cadavers (Lobo et al., 2015).

The methyltransferase (mtrA) have a role in fungal physiological process. The mtrA gene (921 bp) encoding a protein of 307 amino acids (molecular mass 36 kDa, GenBank accession number EJP69472) in *B. bassiana*. Insect bioassay using wild-type *B. bassiana*, Δ BbmtrA, Δ BbmtrA : : mtrA strains on larvae of the greater wax moth, *G. mellonella*, revealed that median lethal time (LT₅₀) for wild-type strain is the 7.2 days and 80% of mortality, these values is decreased in Δ BbmtrA : : mtrA strains with values of 10 days and 60% for LT₅₀ and mortality, respectively (Qin et al., 2014). Actually, an insect-toxic protein (Bb70p) from *B. bassiana* was assayed versus *G. mellonella* with the activation of phenol oxidase cascade. The Bb70p can be enhancing the fungi virulence (Khan et al., 2016).

Some fungal genes have been mentioned in host colonization and killing, one of them is calcium sensor-1 (Bbcsa1) from *B. bassiana* which is homologue with a neuronal calcium-

sensor. Bbcsa1 is associated in pre-penetration or early penetration events, but is dispensable once the insect cuticle has been breached (Fan et al., 2012). For conidiation, multi-stress tolerance and virulence in *B. bassiana*, a response regulator denominated Bbssk1 is necessary. A full-length sequence of Bbssk1 is 2355 bp, encoding a 784 amino acids protein with a molecular mass of 84.5 kDa (Wang et al., 2014). Other gene expressed during colonization of insects is alcohol dehydrogenase (Adh1) from *M. anisopliae*, which is important for full capability of fungus for penetrate/colonize the insect.

MOLECULAR PHYLOGENY OF ENTOMOPATHOGENIC FUNGI AND THEIR BIOGEOGRAPHIC IMPLICATIONS

Entomopathogenic fungi constitute an important biotic component in the natural regulation of arthropod population sizes (Meyling and Eilenberg, 2007). More than 750 entomopathogenic fungal species have been isolated and described worldwide (Nielsen et al., 2008) but species of the Deuteromycete fungal genera *Metarhizium* and *Beauveria* are virulent, generalist entomopathogens that are well established as infecting a wide variety of non-social insect hosts. They are principally associated with hosts in the soil environment, and thus soil-nesting social insects, such as most ants and termites. Although these taxa are probably the most intensively studied entomopathogenic fungi, relatively little is known about their diversity in the environment. While they are known to be highly diverse across species or between widely separated populations, very few studies have looked at within population diversity (Tigano-Milani et al., 1995; Driver et al., 2000; Pantou et al., 2003).

Numerous commercial products using biological control agents (BCAs) have been developed. Most of them have used *Metarhizium* spp. (Zimmermann, 2007; Copping, 2009). Strains from the *Metarhizium* genus infect a range of insects comprising agronomically important pests such as locusts, grasshoppers, termites, noctuids, scarabid beetle larvae, spittlebugs and other hemipterans (Zimmermann, 2007). First, the classification of *Metarhizium* was based on morphological characters and was reviewed by Tulloch (1976), who only accepted *M. avoviride* and *M. anisopliae*, the latter with the short spored var. *anisopliae* and the long-spored var. *majus*. The taxonomy of *Metarhizium*, and particularly, the *M. anisopliae* morphospecies has been investigated by applying morphological, biochemical as well as genetic characteristics (Tulloch, 1976; Driver et al., 2000; Pantou et al., 2003). However, the proposed taxonomies have been incomplete and partially inconsistent. Traditional identification of *Beauveria* species is based on conidial morphology but also molecular phylogenies have revealed that the genus includes cryptic species (Rehner and Buckley, 2005).

Differentiation between inter- and intra-specific groups of entomopathogenic fungi has been possible because of molecular techniques using DNA sequences. Techniques involving polymerase chain reaction (PCR) amplification of DNA, such as RAPD (random amplified polymorphic DNA) or RFLP (restriction fragment length polymorphism), analyses of nuclear or mitochondrial DNA have been particularly useful in resolving taxonomic and evolutionary problems in fungi (Rosewich and McDonald, 1994). These tools have also led to a new phylogenetic classification of the fungi that has challenged many assumptions about the

relationships among entomopathogenic and other fungi (Blackwell et al. 2006; Hibbett et al. 2007).

The globally distributed *B. bassiana* consists of members placed in three separate clades which have been proposed to be considered separate species (Rehner and Buckley, 2005; Ghikas et al., 2010): Clade A, corresponds to *B. bassiana* and is the sister clade of Clade B, which comprises *Beauveria brongniartii* (Rehner and Buckley, 2005). The third clade, Clade C, is distantly related and includes members being morphologically indistinguishable from those of Clade A (Rehner and Buckley, 2005). Therefore, identification of *Beauveria* isolates of Clade C is only possible with the use of molecular markers. In addition, Clade A consists of an assemblage of cryptic species for which an identification system has been established designating phylogenetic species by continent and in the order of their discovery using terminology from Rehner et al., (2006) and Meyling et al., (2009). In case of Clade C, it has been reported that occurs at relatively low frequencies in a *Beauveria* community within a hedgerow habitat in Denmark (Meyling et al., 2009). Phylogenetic assignment of *Beauveria* isolates to specific clades can be done by sequencing the genomic DNA regions of the internal transcribed spacer (ITS) region of the ribosomal RNA (rRNA) gene cluster (White et al., 1990), the 5' end of Elongation Factor 1- α (EF1- α) (Bischoff et al., 2009) and the intergenic Bloc region (Rehner et al., 2006). Efficiency of these regions has been evaluated using a global collection of *Beauveria* spp, resulting that the entire EF1 region provided more phylogenetic resolution than the ITS region (Rehner and Buckley, 2005). Moreover, sequencing both EF1 α and the Bloc region of *Beauveria* isolates from a single community in an agro ecosystem in Denmark has resolved isolates among Clade A, B and C as well as identified five phylogenetic species within Clade A isolates (Meyling et al., 2009). Further molecular characterization of isolates has become possible by using molecular markers such as simple sequence repeat (SSR) in case of Clade A (Rehner and Buckley, 2003; Meyling et al., 2009) and Clade B (Enkerli et al., 2001) to assess genetic diversity of isolates from a single host species. Meyling et al., (2012) studied the genetic diversity among 32 *Beauveria* strains isolated from pollen beetles *Meligethes aeneus* collected in oilseed rape fields at different sites in Switzerland, in this study it was possible to identify three clades based on sequences of the ITS region and de 5' end of EF1- α . About half of the *Beauveria* isolates belonged to Clade C (18 of 32) indicating that it may be widespread in the investigated Swiss farmland. In fact, studies based on *Beauveria* isolates obtained from culture collections and representing worldwide distributions have revealed frequencies of *Beauveria* Clade C of 17% (Rehner and Buckley, 2005) and 22% (Ghikas et al., 2010), respectively. According to Rehner and Buckley, (2005), *Beauveria* Clade C isolates have been shown to originate from five separate insect host orders indicating a wide host range. In addition, the intergenic region Bloc provided best resolution of the clades A and B. Further genotypic diversity was revealed by simple sequence repeats (SSR) characterization of isolates within the clades except for Clade A (*B. bassiana*) which appeared to be clonal. However, the individual SSR markers were differentially amplified from isolates of the different clades, suggesting that not all available SSR markers are suitable for reliable characterization of diversity within *Beauveria* Clade C.

Phylogeny has been also assessed in *Beauveria bassiana* through analysis of genes encoding secondary metabolites that can mediate the interaction of a fungal pathogen with an insect host and help the fungus compete with other microbes. Major classes of secondary metabolites include no ribosomal peptides, alkaloids, terpenes, and polyketides, whose production appear to be controlled by various genetic and cellular regulatory mechanisms

(Punya et al., 2015). Polyketides, in particular, have been recognized as a structurally diverse family of natural products with various biological activities and pharmacological properties, which are synthesized by enzymes called polyketide synthases (PKSs). There are three types of PKSs, but fungi possess only type I PKSs and the presence of β -keto reductase (KR), dehydratase (DH) and enoyl reductase (ER) domains has led to classification of fungal PKSs into three subgroups (Cox, 2007; Chooi and Tang, 2012). Amnuaykanjanasin et al., (2009) conducted PCR employing PKS-specific degenerate primers to identify PKS genes from several fungi isolated from different habitats (insect cadavers, soil, ocean, lichen, and wood), founding a group of insect-specific PKSs classified into clade III. However, Punya et al., (2015) screened insect-specific PKS of 23 isolates of entomopathogenic fungi amplifying 72 PKS gene fragments in order to develop a more comprehensive phylogeny. Subsequent phylogenetic analysis, identified three insect-specific PKS groups in reducing clades IIa, IIb, and III which are highly conserved in entomopathogenic fungal species. PKS genes from insect-specific clades IIa and IIb were expressed only in insect-containing medium, while others were expressed only in PDB or in CYB, PDB and SDY. Also, three distinct reducing PKS clades V, VI, and VIII formed tight clusters of PKSs from various fungi with >94% NJ bootstrap support, supporting the notion that these are three new reducing PKS clades. As fungal polyketides from these entomopathogens represent valuable natural product resources, these PKSs phylogeny and expression data would be important for further studies of valuable polyketides in the future.

Metarhizium is a frequently isolated from soil environments. However, profound knowledge of natural occurrence and distribution, genetic diversity and community structure of the species is required to evaluate consequences of biocontrol initiatives (Meyling and Eilenberg, 2007). In 2009, Bischoff et al., provided a multilocus phylogeny of the *Metarhizium anisopliae* (Metschn.) Sorokin lineage and revised the taxonomy recognizing nine species within the *M. anisopliae* lineage, including several cryptic species. Given the cryptic diversity within the *M. anisopliae* lineage, discrimination of species cannot solely be based on morphology but requires the use of molecular methods for accurate identification. For species discrimination, 5' end of the elongation factor 1 α has been highlighted as a reliable marker (Bischoff et al., 2009). However, this region does not provide sufficient resolution for identification of genotypes within species. Thus, for genotyping and assessing within species diversity, simple sequence repeat (SSR) markers are currently among the most suitable markers (Enkerli and Widmer, 2010). Enkerli et al., (2005) and Oulevey et al., (2009) have developed SSR markers for strain-level genotyping within the *M. anisopliae* lineage, which have successfully been used to investigate molecular diversity of isolates collection (Velasquez et al., 2007; Oulevey et al., 2009). In 2014, Steinwender et al., evaluated *Metarhizium* community in soil from an agricultural field in Denmark using *Tenebrio molitor* as bait insect. By sequence analysis of 5' end of elongation factor 1 α and their genotypic diversity characterized by multilocus simple sequence repeat (SSR) typing, 123 isolates were identified. In this study, *Metarhizium brunneum* was most frequent (78.8%) followed by *M. robertsii* (14.6%), while *M. majus* and *M. flavoviride* were infrequent (3.3% each). It revealed co-occurrence of at four *Metarhizium* species in the soil of the same agro ecosystem with a single *M. brunneum* multilocus genotype being highly prevalent. Abundance of a single genotype could be due to variability among *Metarhizium* genotypes in response to biotic and abiotic factors prevalent in the ecosystem (Bidochka et al., 2005). Moreover, five genotypes of *M. brunneum* and six genotypes of *M. robertsii* were identified among the isolates based on SSR fragment length

analysis, demonstrating to be highly effective in detecting genotypic diversity beyond what could be found by sequencing 5' EF-1 α within the *M. anisopliae* lineage.

Additionally, the entomopathogenic hyphomycete *Paecilomyces fumosoroseus* is a promising candidate for biocontrol of the Silverleaf Whitefly *Bemisia tabaci-argentifolii* (Hemiptera: Aleyrodidae), which is considered as a major insect pest in field and greenhouse crops (Lacey et al., 1996). Since 1995, Tigano-Milani et al., demonstrated intraspecific genetic variation in this fungus by RAPD-PCR and tRNA-PCR. However, the analysis of 28S-rDNA sequences did not provide enough information for differentiation of *P. fumosoroseus* isolates. Fargues et al., (2002) investigated the genetic variability in 48 isolates of this fungus by analyzing the RFLPs and sequence data of the internal transcribed spacer sequences ribosomal RNA gene (rDNA-ITS). Digestion with six endonucleases (*AluI*, *HaeIII*, *Hin6I*, *HpaII*, *NdeII*, and *SmaI*) allowed their separation into three distinct groups. The group 1 was composed of strains isolated only from the host *B. tabaci-argentifolii*. By contrast, the group 3 included strains from various insect host and geographical origins. These data were strongly supported by phylogenetic analysis of rDNA-ITS sequence that recognized three monophyletic groups within the *P. fumosoroseus* complex. Thus, these molecular tools could be useful to assess genetic relatedness of these species into the monitoring of such biocontrol products. Genetic variations has been also suggested in *P. farinosus* according to the broad geographic and host origins and morphological variation correlated with one or more distinguishing adaptations. In the study of Chew et al., (1997) the genetic relatedness of twenty isolates of *P. farinosus* collected from seven insect species in eastern Canada was determined by RAPD analysis. All *P. farinosus* isolates were clearly distinguished from three other entomopathogenic fungi, including *P. fumosoroseus*; however, RAPD banding patterns did not correlate with ecological backgrounds or morphological phenotypes. These observations support the conclusion that *P. farinosus* from eastern Canada is not composed of strains which can be separated on the basis of the ecological or morphological criteria selected.

EVOLUTION OF PATHOGENICITY GENES IN ENTOMOPATHOGENS

Entomopathogenous fungi have developed production of different enzyme and metabolites to parasitize susceptible insect hosts, among them, hydrolytic, assimilatory, and/or detoxifying enzymes such as lipase/esterases, catalases, cytochrome P450s, proteases, and chitinases; and (b) secondary metabolites which facilitate infection (Ortiz-Urquiza and Keyhani, 2013). In *B. bassiana* bacterial-like toxins and effector-type proteins were reported (Xiao et al., 2012). Evolution of the genes codifying for these enzymes has been in a convergent way. Phylogenomic studies of different species in the *Cordyceps/Metarhizium* genera suggest that have evolved into insect pathogens independently of each other, and that their similar large secretomes and gene family expansions are due to convergent evolution (Zheng et al., 2011). Sequencing of *B. bassiana* genome confirmed that ascomycete entomopathogenicity is polyphyletic and convergent evolution to insect pathogenicity (Xiao et al., 2012).

Lipases are the first enzymes synthesized by the entomopathogenic fungi, in *Beauveria bassiana* and *Metarhizium robertsii* was reported a cytochrome P450 subfamily, these enzyme break down long-chain alkenes and fatty acids (Sanchez-Perez et al., 2014). Then, proteases

are synthesized. The catalytic function of proteases is to hydrolyze proteins releasing amino acids which could be used for fungal nutrition (Ventura-Sobrevilla et al., 2015). One of the protease produced are subtilisin type (Pr1) which are considered as virulence indicators. This type of enzyme are regulated by a signal transduction mechanism activated by the protein kinase A (PKA) mediated by AMPc (Sanchez-Perez et al., 2014). Other important group of enzyme produced by the entomopathogen fungi is chitinases which are used to break down the insect chitin.

Entomopathogens fungi descend from plant-associated fungi. Quiroz-Velasquez et al., (2014), studying the transcriptome of *Lagenidium giganteum* which is an oomycete entomopathogenic mentioned that alignments of the cellulose synthase sequence indicated that this fungus appears to be evolved from a phytopathogenic ancestor. In addition, these fungal species have retained genes indicative of plant associations, and may share similar cores of virulence factors. Members of the glycoside hydrolase family 5 subfamily 27 (GH5_27) have been proposed as putative virulence factors which may be active on the host insect cuticle, these genes are shared by different entomopathogenic fungi. These virulence factors may be very host specific with a very low risk of attacking non-target organisms or beneficial insects (Shahid et al., 2012). *Metarhizium* and *Beauveria* are also found as plant symbionts. Recent studies showed that these fungi are more closely related to grass endophytes and developed genes for insect pathogenesis while maintaining an endophytic lifestyle. Some genes for insect pathogenesis may have been co-opted from genes involved in endophytic colonization. In contrast, other genes may be multifunctional and serve in both lifestyles (Barelli et al., 2015).

Contraction and Expansion of Different Gene Families

Different entomopathogenic fungi such as *Ophiocordyceps polyrhachis-furcata*, *Beauveria bassiana*, *Metarhizium robertsii*, *Metarhizium acridum*, *Cordyceps militaris*, and *Ophiocordyceps sinensis* have similar genes implicated in pathogenicity and virulence. In *B. bassiana*, many species-specific virulence genes and gene family expansions and contractions correlate with host ranges and pathogenic strategies (Xiao et al., 2012). Contractions of some gene families in this type of fungi are implicated in narrow host-range insect species (specialists), some of these genes are cuticle-degrading genes and families of pathogen-insect interaction (PHI) genes. In some of the most specialized entomopathogens such as *O. polyrhachis-furcata*, for many genes-families has the least number of genes found (Wichadakul et al., 2015). Reduction in gene family sizes was reported in other entomopathogen (*Hirsutella thompsonii*) (Agrawal et al., 2015). The loss of different genes involved with pathogenicity result in a reduced capacity to exploit larger ranges of insect hosts and therefore in the different level of host specificity. Specialization is associated with retention of sexuality and rapid evolution of existing protein sequences (Hua et al., 2014). It is reported a co-evolution between entomopathogens and insect-host which in some case followed different patterns. Jensen et al., (2009) indicated that because of a high diversification over time among dipteran insects, the insect pathogenic fungi associated with these insects have also diversified.

On the other hand, expansions of genes involved in 1) the production of bacterial-like toxins, and 2) retrotransposable elements have been mentioned in *O. polyrhachis-furcata*, in comparison to other entomopathogenic fungi. Expansions of gene families suggest an adaptation to particular environments or sophisticated mechanisms underlying pathogenicity

through retrotransposons (Wichadakul et al., 2015). Similar pattern of evolutionary expansion of gene families such as chitinases, lipases, and proteases has been reported for *Beauveria bassiana*, *Cordyceps militaris*, and *Metarhizium anisopliae* which could be attributed to their insect killing strategies and host ranges (Agrawal et al., 2015). Generalist entomopathogens attack a wide range of insects; this condition has been associated with protein-family expansion, loss of genome-defense mechanisms, genome restructuring, horizontal gene transfer, and positive selection that accelerated after reinforcement of reproductive isolation. Generalists evolved from specialists via transitional species with intermediate host ranges and that this shift paralleled insect evolution (Hua et al., 2014). Species from the *Metarhizium* and *Beauveria* genera are recognized as generalists infecting a range of insects (Meyling et al., 2011).

Entomopathogen Fungi Pathogenicity versus Insect-Host Defense

Insects have evolved different mechanisms in response to pathogens attack, some of these mechanisms are: production of (epi) cuticular antimicrobial lipids, proteins, and metabolites; (b) shedding of the cuticle during development; and (c) behavioral-environmental adaptations (Ortiz-Urquiza et al., 2013). After 25th generations under constant selective pressure from *Beauveria bassiana*, the larvae of Greater wax moth, *Galleria mellonella*, exhibited significantly enhanced resistance, which was specific to this pathogen, and not to another insect pathogenic fungus, *Metarhizium anisopliae* (Dubovskiy et al., 2013). It was hypothesized by the same authors that insects developed a transgenerationally primed resistance which was achieved not by compromising life-history traits but rather by prioritizing and re-allocating pathogen-species-specific augmentations to integumental front-line defenses that are most likely to be encountered by invading fungi. However, there is a coevolution between entomopathogen virulence factors and host defense molecules of insects.

Virulence is thought to co evolve as a result of reciprocal selection between pathogens and their hosts. Because of shorter generation times and smaller genomes, microbes exhibit a high evolutionary adaptability in comparison with their hosts. In contrasts, insects can only compete with its pathogens if they develop mechanisms providing a comparable genetic plasticity (Vilcinskis, 2010). During *B. bassiana* infection, *Galleria mellonella* systemic immune defenses are suppressed in favor of a more limited but targeted repertoire of enhanced responses in the cuticle and epidermis of the integument (Dubovskiy et al., 2013). If a diversification of fungal proteinases for pathogenesis arose, an expansion of host proteinase inhibitors subsets contributing to insect innate immunity may occur. For example, the spectrum of proteolytic enzymes encompasses thermolysin-like metalloproteinases and this is associated with the pathogen-virulence. This spectrum putatively promoted the evolution of corresponding host inhibitors of these virulence factors (Vilcinskis, 2010). The same authors mentioned other molecular adaptations for host insect's defense such as sensing and feedback-loop regulation of microbial metalloproteinases. The genetic events behind the countermeasures in host defense effectors are gene or domain duplication and shuffling by recombination. The entomopathogens also develop different strategies in order to avoid the host insect's defenses. It has been proposed that *M. anisopliae* and *B. bassiana* survive to insect phagocytic haemocytes which are analogous to the mammalian macrophages as consequence of adaptations that have evolved in order to avoid predation by soil amoebae (Bidochka et al., 2010).

GENETIC IMPROVEMENT OF ENTOMOPATHOGENIC FUNGI FOR INSECT BIOCONTROL

Mycoinsecticides are being used for the control of many insect pests as an environmentally acceptable alternative to chemical insecticides uses (Leger et al., 1996; Hussain et al., 2014; Ortiz, et al., 2015). From 1960s, a substantial number of mycoinsecticides have been developed worldwide (Sardul et al. 2012). All groups of insects may be affected and over 700 species of fungi from around 90 genera are pathogenic to insects (Khachatourians and Sohail, 2008). Basically, the fungi pathogen activity depends on the ability of its enzymatic equipment, consisting of lipases, proteases and chitinases, which are in charge of breaking down the insect's integument (Sardul et al., 2012). Besides exoenzymes, the entomopathogenic fungi are reported to secrete toxin proteins and metabolites in vitro and sometime in vivo as well. There are a number of toxic compounds in the filtrate of entomopathogenic fungi such as small secondary metabolites, cyclic peptides and macromolecular proteins (Khan et al., 2012). To improve both commercial and technical efficiency of these entomopathogen fungi, a large number of studies had been conducting to improve their virulence which can be achieved by understanding the mechanisms of fungal pathogenesis and genetically modifying targeted genes.

In response, Fan et al., (2010) in order to accelerate penetration speed and have better target protein-chitin on the cuticle, they constructed a hybrid protease (CDEP-BmChBD) by fusion of a chitin binding domain BmChBD from *Bombyx mori* chitinase to the C-terminal of CDEP-1, a subtilisin-like protease from *B. bassiana*. After comparing studies, the hybrid protease was able to bind chitin and released greater amounts of peptides/proteins from insect cuticles. The insecticidal activity of *B. bassiana* was improved by including proteases, CDEP-1 or CDEP: BmChBD produced in *Pichia pastoris*, as an additive, however, the improved effect of CDEP: BmChBD was significantly higher than that of CDEP-1. Expression of the hybrid protease in *B. bassiana* also significantly increased fungal virulence compared to wild-type and strains overexpressing the native protease. These results demonstrated that rational design of virulence factor is a potential strategy for strain improvement by genetic engineering. Recently, a cytochrome P450 subfamily, referred as CYP52XI and MrCYP52 has been identified in *Beauveria bassiana* and *Metarhizium robertsii*, respectively (Sanchez-Perez et al., 2014). Entomopathogens such as *M. anisopliae* and *B. bassiana* are well characterized in respect to pathogenicity to several insects and have been used as myco-biocontrol agents for biological control of agriculture pests worldwide (Sardul et al., 2012).

Lenger et al., (2012) reported the development of a genetically improved entomopathogenic fungus because integration of copies from the gene encoding a regulated cuticle degrading protease (Prl) which were inserted into the genome of *M. anisopliae*. Prl was constitutively overproduced in the hemolymph of *Manduca sexta*, activating the prophenoloxidase system. The combined toxic effects of Prl and the reaction products of phenoloxidase caused larvae challenged with the engineered fungus to exhibit a 25% reduction in time of death and reduced food consumption by 40% compared to infections by the wild-type fungus. In addition, infected insects were rapidly melanized, and the resulting cadavers were poor substrates for fungal sporulation.

Fang et al., (2009) found that overexpression of a subtilisin-like protease (Pr1A) or a chitinase (Bbchit1) resulted in increased virulence of *M. anisopliae* and *B. bassiana*,

respectively. In this study, they found that a mixture of the *B. bassiana* Pr1A homolog (CDEP1) and Bbchit1 for degradation of insect cuticle were in vitro more efficient than either CDEP1 or Bbchit1 alone. Based on this, they constructed three plasmids; (1) Bbchit1, (2) CDEP1, and (3) a fusion gene of Bbchit1 linked to CDEP1 each under the control of the constitutive *gpd* promoter from *Aspergillus nidulans*. *B. bassiana* transformants secreting the fusion protein (CDEP1:Bbchit1) which penetrated the cuticle significantly faster than the wild type or transformants overexpressing either Bbchit1 or CDEP1. Compared to the wild type, the transformant overexpressing CDEP1 showed a 12.5% reduction in LT (50), without a reduction in LC (50). The LT (50) of the transformant expressing CDEP1:Bbchit1 was reduced by 24.9%. Strikingly, expression of CDEP1:Bbchit1 resulted in a 60.5% reduction in LC (50), more than twice the reduction obtained by overexpression of Bbchit1 (28.5%). This work represents a significant step towards the development of hypervirulent insect pathogens for effective pest control.

FUTURE TRENDS

It is very important to investigate more about the relation of entomopathogenic fungi to grass endophytes and how they developed genes for insect pathogenesis while maintaining an endophytic lifestyle. Although, there is known that some genes for insect pathogenesis may have been co-opted from genes involved in endophytic colonization, it is important to understand the protein changes that lead this adaptation. On the other hand, it has been mentioned that other genes may be multifunctional and serve in both lifestyles, but still there lack of knowledge about the genes modification and protein changes to have this dual activity. Advances in molecular tools for phylogeny analysis could lead to significant new insights that should allow us a better understand of the ecology of fungal entomopathogens as well as their additional roles in nature, including as plant endophytes, antagonists of plant pathogens, beneficial rhizosphere-associates and possibly even plant growth promoters.

Although some entomopathogen fungi are well known, some of the fungi of the Entomophthorales order such as *Entomophthora*, *Erynia* and *Pandora* species are poorly investigated because, the culture medium used to propagate entomopathogens such as *Beauveria*, *Metarhizium*, and *Lecanicillium*, are not the most adequate for these especial entomopathogens.

CONCLUSION

Different fungal entomopathogenic species such as *B. bassiana*, *M. acridum*, *M. anisopliae*, and *Metarhizium brunneum* have showed commercial potential for insect control, and are considered as friendly mycoinsecticide. The complete genome of some entomopathogen fungi such as *B. bassiana* has been sequenced, which has revealed multiples gene associated with virulence. In addition, many bacterial-like toxins and effector-type proteins were also discovered. Entomopathogenic fungi are able to adapt to different environments by activating well-define gene sets. During infection entomopathogenic fungi, many genes are involved in seven steps: host adhesion, germination, cuticle degradation,

growth as blastospores, host colonization and killing, immune response interactions and hyphal extrusion and conidiation.

ACKNOWLEDGMENTS

A. Garcia, M. Michel, and S. Villarreal want to thank to the National Council of Science and Technology of Mexico (CONACYT) for the financial support during their postgraduate studies.

REFERENCES

- [1] Adsure, S. P., and Mohite, P., B. 2015. Efficacy of entomopathogenic fungi against gram pod borer, *Helicoverpa armigera* (HUB.) on chickpea. *Journal of Global Biosciences*, 4 (8): 3154-3157.
- [2] Agrawal, Y., Khatri, I., Subramanian, S., Shenoy, B. D., 2015. Genome sequence, comparative analysis, and evolutionary insights into chitinases of entomopathogenic fungus *Hirsutella thompsonii*. *Genome Biol. Evol.* 916–930 DOI:10.1093/gbe/evv037.
- [3] Ainsworth, G., 1973. Introduction and keys to higher taxa. In: *The fungi: An advanced treatise*. Ainsworth, G., F. Sparrow and A. Sussman. Academic Press, New York (IVA): 1-7.
- [4] Amnuaykanjanasin, A., Ponghanphot, S., Sengpanich, N., Cheevadhanarak, S., Tanticharoen, M., 2009. Insect-specific polyketide synthases, potential PKS-NRPS hybrids, and novel PKS clades in tropical fungi. *Appl. Environ. Microbiol.* 75, 3721-3732.
- [5] Barelli, L., Moonjely, S., Behie, S. W., Bidochka, M. J., 2015. Fungi with multifunctional lifestyles: endophytic insect pathogenic fungi. *Plant Mol Biol.* DOI 10.1007/s11103-015-0413-z.
- [6] Barreto, C. C., Staats, C. C., Schrank, A., Vainstein, M. H., 2004. Distribution of chitinases in the entomopathogen *Metarhizium anisopliae* and effect of N-acetylglucosamine in protein secretion. *Current microbiology*, 48(2), 102-107.
- [7] Becerra, V. V., Paredes, M. C., Rojo, C. M., France, A. L., 2007. RAPD and ITS reveal molecular variation of Chilean populations of *Beauveria bassiana*. *Agricultura Técnica*. 67(2), 115-125.
- [8] Berlanga, P. A. M^a. 1997. Aislamiento, identificación y conservación de hongos entomopatógenos. *Memoria del II Curso Taller de Producción de Agentes de Control Biológico*. Tecoman, Colima. 18-30. [Isolation, identification and conservation of entomopathogenic fungi. *Report of the 2nd Course on the Production of Biological Control Agents*. Tecoman, Colima. 18-30]
- [9] Bidochka, M. J., Clark, D. C., Lewis, M. W., Keyhani, N. O., 2010. Could insect phagocytic avoidance by entomogenous fungi have evolved via selection against soil amoeboid predators? *Microbiology* 156: 2164–2171.

- [10] Bidochka, M. J., Small, C. L. N., Spironello, M., 2005. Recombination within sympatric cryptic species of the insect pathogenic fungus *Metarhizium anisopliae*. *Environ. Microbiol.* 7, 1361–1368.
- [11] Bischoff, J. F., Rehner, S. A., Humber, R. A., 2009. A multilocus phylogeny of the *Metarhizium anisopliae* lineage. *Mycologia*, 101(4), 512–530.
- [12] Blackwell, M., Hibbett, D. S., Taylor, J. W., Spatafora, J. W., 2006. Research coordination networks: a phylogeny of the kingdom Fungi (Deep Hypha). *Mycol.* 98, 829–837.
- [13] Boomsma, J. J.; Jensen, A. B; Meyling, N. V; Eilenberg, J., 2014. Evolutionary interaction networks of insect pathogenic fungi. *Annu. Rev. Entomol.* 59:467-85.
- [14] Broetto, L., Da Silva, W. O. B., Bailão, A. M., De Almeida Soares, C., Vainstein, M. H. and Schrank, A., 2010, Glyceraldehyde-3-phosphate dehydrogenase of the entomopathogenic fungus *Metarhizium anisopliae*: cell-surface localization and role in host adhesion. *FEMS Microbiology Letters*, 312: 101–109. DOI: 10.1111/j.1574-6968.2010.02103.x.
- [15] Bruck, D. J., 2010. Fungal entomopathogens in the rhizosphere. *BioControl*, 55(1), 103–112.
- [16] Bulat, S. A., Lubeck, M., Alekhina, I. A., Jensen, D. F., Knudsen, I. M. B., Lubeck, P. S., 2000. Identification of a universally primed-PCR-derived sequence characterized amplified region marker for an antagonistic strain of *Clonostachys rosea* and development of a strain-specific PCR detection assay. *Appl. Environ. Microbiol.*, 66:4758–4763.
- [17] Bustillo, A., 2001. Hongos en insectos y posibilidades de uso en el control biológico de plagas en Colombia. In: *seminario Uso de entomopatógenos en Colombia. Sociedad Colombiana de Entomología*. 30-53. [Fungi in insects and possibilities of use in the biological control of pests in Colombia. In: *seminar Use of entomopathogens in Colombia. Colombian Society of Entomology*. 30-53].
- [18] Cabrera, R. I., Domínguez, D. 1987. *Hirsutella nodulosa* e *Hirsutella kirchneri*; dos nuevos hongos patógenos del ácaro del moho *Phyllocoptruta oleivora*. *Ciencia y Técnica en la agricultura, Cítricos y otros Frutales*, 10 (2):135-138. [*Hirsutella nodulosa* and *Hirsutella kirchneri*; Two new pathogenic fungi of the mold mite *Phyllocoptruta oleivora*. *Science and Technology in agriculture, Citrus and other Fruit trees*, 10 (2): 135-138].
- [19] Cañedo, V and Ames, T., 2004. *Manual de laboratorio para el manejo de hongos entomopatógenos*. Primera edición. Centro Internacional de la Papa. [*Laboratory manual for the management of entomopathogenic fungi*. First edition. International Potato Center].
- [20] Chew, J. S. K., Strongman, D. B., MacKay, R. M., 1997. RFLP analysis of rRNA intergenic spacer region of 23 isolates of the entomopathogen *Paecilomyces farinosus*. *Can. J. Bot.* 75, 2038-2044.
- [21] Chooi, Y. H., Tang, Y., 2012. Navigating the Fungal Polyketide Chemical Space: From Genes to Molecules. *J Organ. Chem.* 77, 9933-9953.
- [22] Cory, J. S., and Ericsson, J. D., 2010. Fungal entomopathogens in a tritrophic context. *BioControl*, 55(1), 75-88.
- [23] Cory, J. S., and Hoover, K., 2006. Plant-mediated effects in insect–pathogen interactions. *Trends in ecology & evolution*, 21(5), 278-286.

- [24] Cox, R. J., 2007. Polyketides, proteins and genes in fungi: programmed nano-machines begin to reveal their secrets. *Organ. Biomol. Chem.* 5, 2010-2026.
- [25] De Faria, M., and S. Wraight., 2007. Mycoinsecticides and Mycoacaricides: A comprehensive list with worldwide coverage and international classification of formulation types. *Biological Control* 43:237256.
- [26] Díaz, B. M., López Lastra, C. C., Oggerin, M., Fereres, A., Rubio, V., 2008. Identificación de hongos entomopatógenos asociados a pulgones en cultivos hortícolas en la zona centro de la Península Ibérica. [Identification of entomopathogenic fungi associated with aphids in horticultural crops in the central area of the Iberian Peninsula]. *Bol. San. Veg. Plagas*, 34: 287-296.
- [27] Driver, F., Milner, R. J., Trueman, J. W. H., 2000. A taxonomic revision of *Metarhizium* based on a phylogenetic analysis of rDNA sequence data. *Mycol. Res.* 104, 134–150.
- [28] Dubovskiy I. M, Whitten MM, Yaroslavtseva ON, Greig C, Kryukov VY, Grizanov EV, Mukherjee K, Vilcinskas A, Glupov VV, Butt TM. 2013. Can insects develop resistance to insect pathogenic fungi?. *PLoS One* 8(4):e60248. doi: 10.1371/journal.pone.0060248. Epub 2013 Apr 1.
- [29] Ebrahimi, S. Sabbagh, S. K, Aminaei, M., Khosravimoghadam, F., Rostami, F., 2015. Genetic diversity of *Paecilomyces variotii* isolates by SSR marker in Kerman province, Iran. *Mycologia Iranica*. 2(1), 38 – 45.
- [30] Elósegui, C. O., Jiménez, R. J., Carr, P. A., 2006. Aislamiento, identificación y caracterización morfológica de aislados nativos de hongos mitosporicos con potencialidad para el control de especies de insectos plaga. [Isolation, identification and morphometric characterization of native isolates of mitosporic fungi with potential for the control of plague insect species]. *Fitosanidad*, 10 (4):265-272.
- [31] Elósegui, C. O. Primera edición. *Métodos artesanales de reproducción de plaguicidas a partir de hongos entomopatógenos y antagonistas*. Instituto de Investigaciones de Sanidad Vegetal. 2006. [First edition. *Artisan methods of reproduction of pesticides from entomopathogenic fungi and antagonists*. Plant Health Research Institute. 2006].
- [32] Enkerli, J., Kolliker, R., Keller, S., Widmer, F., 2005. Isolation and characterization of microsatellite markers from the entomopathogenic fungus *Metarhizium anisopliae*. *Molecular Ecology Notes*, 5:384-386.
- [33] Enkerli, J., Widmer, F., 2010. Molecular ecology of fungal entomopathogens: molecular genetic tools and their applications in population and fate studies. *Biocontrol* 55, 17–37.
- [34] Enkerli, J., Widmer, F., Gessler, C. Keller, S., 2001. Strain-specific microsatellite markers in the entomopathogenic fungus *Beauveria brongniartii*. *Mycological Research*, 105, 1079-1087.
- [35] Fan Y, Pei X, Guo S, Zhang Y, Luo Z, Liao X, Pei Y., 2010. Increased virulence using engineered protease-chitin binding domain hybrid expressed in the entomopathogenic fungus *Beauveria bassiana*. *Microb Pathog.* 49(6):376-80. DOI: 10.1016/j.micpath.2010.06.013.
- [36] Fan Y, Ortiz-Urquiza A, Kudia RA, Keyhani NO., 2012. A fungal homologue of neuronal calcium sensor-1, Bbcsa1, regulates extracellular acidification and contributes to virulence in the entomopathogenic fungus *Beauveria bassiana*. *Microbiology* 158(7):1843-51. DOI: 10.1099/mic.0.058867-0.
- [37] Fang W, Feng J, Fan Y, Zhang Y, Bidochka MJ, Leger RJ, Pei Y., 2009. Expressing a fusion protein with protease and chitinase activities increases the virulence of the insect

- pathogen *Beauveria bassiana*. *J Invertebr Pathol.*, 102(2):155-9. DOI: 10.1016/j.jip.2009.07.013.
- [38] Fang W, Leng B, Xiao Y, Jin K, Ma J, Fan Y, Feng J, Yang X, Zhang Y, Pei Y., 2005. Microbiology. All Rights Reserved. Cloning of *Beauveria bassiana* Chitinase Gene Bbchit1 and Its Application To Improve Fungal Strain Virulence. *Applied and Enviromental Microbiology* 71(1): 363-370. DOI:10.1128/AEM.71.1.363–370.
- [39] FAO. *Resistencia a los antiparasitarios: estado actual con énfasis en América Latina*. Roma: Dirección de producción y sanidad animal de la FAO, 2003. 33-35. [*Antiparasitic resistance: current status with emphasis on Latin America*. Rome: FAO Production and Animal Health Division, 2003. 33-35].
- [40] Fargues, J., Bon, M-C., Manguin, S., Couteaudier, Y., 2002. Genetic variability among *Paecilomyces fumosoroseus* isolates from various geographical and host insect origins based on the rDNA-ITS regions. *Mycol. Res.* 106, 1066-1074.
- [41] Fernandes, E. K. K., Moraes, A. M. L., Pacheco, R. S., Rangel, D. E. N., Miller, M. P., Bittencourt, V. R. E. P., Roberts, D. W., 2009. Genetic diversity among Brazilian isolates of *Beauveria bassiana*: comparisons with non-Brazilian isolates and other *Beauveria* species. *Journal of Applied Microbiology*. 107, 760-774.
- [42] Flores, M. A., Pucheta, D. M., Ramos-López, M. A., Rodríguez, N. S., Ramos, E. G., Juárez, R. D. 2013. Estudio del hongo entomopatógeno *Isaria fumosorosea* como control microbiológico de la mosquita blanca *Bemisia tabaci*. *Interciencia*, 38 (7):523-527. [Study of the entomopathogenic fungus *Isaria fumosorosea* as a microbiological control of the whitefly *Bemisia tabaci*. *Interscience*, 38 (7): 523-527].
- [43] Freed, S., Jin, F. L., Ren, S. X., 2011. Determination of genetic variability among the isolates of *Metarhizium anisopliae* var. *anisopliae* from different geographical origins. *World J Microbiol Biotechnol.* 27. 359–370.
- [44] Galidevara S, Reineke A, Koduru UD., 2016. In vivo expression of genes in the entomopathogenic fungus *Beauveria bassiana* during infection of lepidopteran larvae. *J Invertebr Pathol.* 2(136):32-34. DOI: 10.1016/j.jip.2016.03.002.
- [45] Garrido-Jurado, I., Fernandez-Bravo, M., Campos, C., Quesada-Moraga, E., 2015. Diversity of entomopathogenic Hypocreales in soil and phylloplanes of five Mediterranean cropping systems. *Journal of Invertebrate Pathology*. 130, 97–106.
- [46] Gauthier, N., Dalleau-Clouet, C., Fargues, J., 2007. Microsatellite variability in the entomopathogenic fungus *Paecilomyces fumosoroseus*: genetic diversity and population structure. *Mycologia*, 99:693-704.
- [47] Gerson, U., Kenneth, R., Muttath, T. I. 1979. *Hirsutella thompsonii* a fungal pathogen of mites. II Host-pathogen interactions. *Ann. Appl. Biol.* 91:29-40.
- [48] Ghikas, D. V., Kouvelis, V. N., Typas, M. A., 2010. Phylogenetic and biogeographical implications inferred by mitochondrial intergenic region analyses and ITS15. 8S-ITS2 of the entomopathogenic fungi *Beauveria bassiana* and *B. brongniartii*. *BMC Microbiol.* 10, 174.
- [49] Goble, T. A., Conlong, E. E., Hill, M. P., 2015. Virulence of *Beauveria brongniartii* and *B. bassiana* against *Schizonycha affinis* white grubs and adults (Coleoptera: Scarabaeidae). *J. Appl. Entomol.* 139, 134–145.
- [50] Golo PS, Santos HA, Perinotto WM, Quinelato S, Angelo IC, Camargo MG, Sá FA, Massard CL, Fernandes ÉK, Roberts DW, Bittencourt VR., 2015. The influence of

- conidial Pr1 protease on pathogenicity potential of *Metarhizium anisopliae* senso lato to ticks. *Parasitol Res.* 114(6):2309-15. DOI: 10.1007/s00436-015-4426-y.
- [51] Guan Y, Wang DY, Ying SH, Feng MG., 2015. A novel Ras GTPase (Ras3) regulates conidiation, multi-stress tolerance and virulence by acting upstream of Hog1 signaling pathway in *Beauveria bassiana*. *Fungal Genet Biol.* 82:85-94. DOI: 10.1016/j.fgb.2015.07.002.
- [52] Guo, L. D., 2010. Molecular diversity and identification of endophytic fungi. In Gherbawy, Y., Voigt, K. editors. *Molecular Identification of Fungi*. Springer Heidelberg Dordrecht London New York. Springer, 277-296.
- [53] Hassan, A. E. M., Dillion, R. J., and A. K. Charnley., 1989. *J Invert. Pathol*, 54, 227-279.
- [54] Hernández, M. A. A., Rosón, Á. C., Daquinta, R., C. Trujillo, M., R. 2005. Evaluación *in vitro* de la patogenicidad de *Aschersonia aleyrodis* webber sobre algunos insectos plaga de interés económico. [*In vitro* evaluation of the pathogenicity of *Aschersonia aleyrodis* webber on some plague insects of economic interest]. *Fitosanidad*, 9(4): 29-34.
- [55] Hernández, V., V. M., Berlanga P., A. M., Garza G. E., 1997. Detección de *Metarhizium flavoviridae* sobre *Schistocerca piceifrons* (Orthoptera: Acrididae) en las Islas Socorro, Archipiélago de Revillagigedo, [Detection of *Metarhizium flavoviridae* on *Schistocerca piceifrons* (Orthoptera: Acrididae) in the Socorro Islands, Revillagigedo Archipelago], *México. Vedralia* 4: 45-46.
- [56] Hesketh, H., Roy, H. E., Eilenberg, J., Pell, J. K., & Hails, R. S., 2009. Challenges in modelling complexity of fungal entomopathogens in semi-natural populations of insects. In *The Ecology of Fungal Entomopathogens* (pp. 55-73). Springer Netherlands.
- [57] Hibbett, D. S., Binder, M., Bischoff, J. F., Blackwell, M., Cannon, P. F., Eriksson, O., Huhndorff, S., James, T., Kirk, P. M., Lücking, R., Lumbsch, T., Lutzoni, F., Matheny, P. B., McLaughlin, D. J., Powell, M. J., Redhead, S., Schoch, C. L., Spatafora, J. W., Stalpers, J. A., Vilgalys, R., Aime, M. C., Aptroot, A., Bauer, R., Begerow, D., Benny, G. L., Castlebury, L. A., Crous, P. W., Dai, Y.-C., Gams, W., Geiser, D. M., Griffith, G. W., Gueidan, C., Hawksworth, D. L., Hestmark, G., Hosaka, K., Humber, R. A., Hyde, K., Köljalb, U., Kurtzman, C. P., Larsson, K.-H., Lichtwardt, R., Longcore, J., Mialikowska, J., Miller, A., Moncalvo, J.-M., Mozley-Standridge, S., Oberwinkler, F., Parmasto, R., Reeb, V., Rogers, J. D., Roux, C., Ryvarden, L., Sampaio, J. P., Schuessler, A., Sugiyama, J., Thorn, R. G., Tibell, L., Untereiner, W. A., Walker, C., Wang, Z., Weir, A., Weiss, M., White, M., Winka, K., Yao, Y.-J., Zhang, N., 2007. A higher-level phylogenetic classification of the fungi. *Mycol. Res.* 111, 509–547.
- [58] Hua, X., Xiao, G., Zhenga, P., Shanga, Y., Sub, Y., Zhangb, X., Liub, X., Zhana, S., Legerc R. J., Wanga, Ch., 2014. Trajectory and genomic determinants of fungal-pathogen speciation and host adaptation. *PNAS* 111(47):16796–16801.
- [59] Hubka, V., and Kolarik, M., 2012. β -tubulin paralogue tubC is frequently misidentified as the benA gene in *Aspergillus* section *Nigri* taxonomy: primer specificity testing and taxonomic consequences. *Persoonia* 29, 1–10. Hussain, A., Tian, MY., Ahmed S., Shahid M., 2012. Current status of entomopathogenic fungi as mycoinsecticides and their inexpensive development in liquid cultures. In García M. D., (Ed.), *Zoology InTech*, Available from: <http://www.intechopen.com/books/zoology/current-status-of-entomopathogenic-fungi-as-mycoinsecticides-andtheir-inexpensive-development-in-liq>.

- [60] Hussain, A., Rizwan-ul-Haq, M., Al-Ayedh, H., Al-Jabr, A. M. 2014. Mycoinsecticides: Potential and future perspective. *Recent Patents on Food, Nutrition & Agriculture*, 6:45-53.
- [61] Inglis, G. A., Enkerl, J., Goettel, M. S. Laboratory Techniques Used for Entomopathogenic Fungi: Hypocreales. pp. 189-253. In: L. Lacey (Ed.). *Manual of Techniques in Insect Pathology*. 2da. Edition Academic Press. 2012.
- [62] Jensen, A. B.; Eilenberg, J; Lastra, C. L., 2009: Differentially aphid and fly host driven divergence of obligate insect pathogenic *Entomophthora* species. *FEMS Microbiology Letters* 300: 180-187.
- [63] Jensen, A., B., Thomsen L., Eilenberg. J. 2006. Value of host range, morphological, and genetic characteristics within the *Entomophthora muscae* species complex. *Mycol., Res* 110: 941-950.
- [64] Junges Â, Boldo JT, Souza BK, Guedes RL, Sbaraini N, Kmetzsch L, Thompson CE, Staats CC, de Almeida LG, de Vasconcelos AT, Vainstein MH, Schrank A., 2014. Genomic analyses and transcriptional profiles of the glycoside hydrolase family 18 genes of the entomopathogenic fungus *Metarhizium anisopliae*. *PLoS One*. 18; 9(9):e107864. DOI: 10.1371/journal.pone.0107864.
- [65] Keller, S., Kalsbeek V., Eilenberg J., 1999. Redescription of *Entomophthora muscae* (Cohn) Fresenius. *Sydowia*, 51 (2): 197-209.
- [66] Khachatourians G. G., Sohail S. Q., 2008. Entomopathogenic fungi, In: Brakhage A. A., and Zipfel P. F. (eds.), *Biochemistry and molecular biology, human and animal relationships*, 2nd Edition. The Mycota VI, Springer-Verlag, Berlin, Heidelberg.
- [67] Khan S, Nadir S, Lihua G, Xu J, Holmes KA, Dewen Q., 2016. Identification and characterization of an insect toxin protein, Bb70p, from the entomopathogenic fungus, *Beauveria bassiana*, using *Galleria mellonella* as a model system. *J Invertebr Pathol*. 133:87-94. DOI: 10.1016/j.jip.2015.11.010.
- [68] Khan, S., Guo L., Maimaiti Y., Mijit M., Qiu D., 2012. Entomopathogenic fungi as microbial biocontrol agent. *Molecular Plant Breeding*, 3(7):63-79 (DOI: 10.5376/mpb.2012.03.0007).
- [69] Khudhair M. W., Alrubeai, H. F., Khalaf, M. Z., Shbar, A. K., Hamad, B. S., Khalaf, H. S., 2014. Occurrence and distribution of entomopathogenic fungi in Iraqi agro-ecosystems. *Int. J. Entomol. Res*. 2: 117-124.
- [70] Koduru UD, Galidevara S, Reineke A, Pathan AAK., 2014. Bioinformatic tools in the analysis of determinants of pathogenicity and ecology of entomopathogenic fungi used as microbial insecticides in crop protection. In: Kavi Kishor PB, Rajib Bandopadhyay, Prashanth Suravajhala (eds). *Agricultural Bioinformatics*. 215-234.
- [71] Lacey, L. A., Fransen, J. J., Carruthers, R. I., 1996. Global distribution of naturally occurring fungi of *Bemisia*, their biologies and use as biological control agents. In *Bemisia 1995: taxonomy, biology, damage, control and management* (D. Gerling & R. Mayer, eds): 401-433. Intercept, Andover.
- [72] Leão MP, Tiago PV, Andreote FD, de Araújo WL, de Oliveira NT., 2015. Differential expression of the pr1A gene in *Metarhizium anisopliae* and *Metarhizium acridum* across different culture conditions and during pathogenesis. *Genet Mol Biol*. 38(1):86-92. DOI: 10.1590/S1415-475738138120140236.

- [73] Leger, R. J., Lokesh Joshi, Michael J. Bidochka, and Donald W. Roberts., 1996. Construction of an improved mycoinsecticide overexpressing a toxic protease. *Proc. Natl. Acad. Sci.* 93:6349-6354.
- [74] Li, J., Li, H., Bi, X., Zhang, K. Q., 2013. Multiple gene genealogical analyses of a nematophagous fungus *Paecilomyces lilacinus* from China. *Journal of Microbiology.* 51(4), 423–429.
- [75] Liu, M., Chaverri, P., Hodge, K., T., 2006. A taxonomic revision of the insect biocontrol fungus *Aschersonia aleyrodidis*, its allies with white stromata and their *Hypocrella* sexual states. *Mycological Research*, 110(5): 537–554.
- [76] Lobo LS, Luz C, Fernandes ÉK, Juárez MP, Pedrini N., 2015. Assessing gene expression during pathogenesis: Use of qRT-PCR to follow toxin production in the entomopathogenic fungus *Beauveria bassiana* during infection and immune response of the insect host *Triatoma infestans*. *J Invertebr Pathol.* 128:14-21. DOI: 10.1016/j.jip.2015.04.004.
- [77] López-Llorca, L. V., and Jansson, H. B., 2001. Biodiversidad del suelo: control biológico de nemátodos fitopatógenos por hongos nematófagos. *Cuadernos de biodiversidad: publicación cuatrimestral del Centro Iberoamericano de la Biodiversidad*, (6), 12-15. [Soil biodiversity: biological control of phytopathogenic nematodes by nematophagous fungi. *Biodiversity Notebooks: quarterly publication of the Ibero-American Center for Biodiversity*, (6), 12-15].
- [78] Luan, F., Zhang, S., Wang, B., Huang, B., Li, Z., 2012. Genetic diversity of the fungal pathogen *Metarhizium* spp., causing epizootics in Chinese burrower bugs in the Jingting Mountains, eastern China. *Mol Biol Rep.* 40(1), 515-23.
- [79] Luo X, Keyhani NO, Yu X, He Z, Luo Z, Pei Y, Zhang Y., 2012. The MAP kinase Bbslt2 controls growth, conidiation, cell wall integrity, and virulence in the insect pathogenic fungus *Beauveria bassiana*. *Fungal Genet Biol* 49(7):544-55. DOI: 10.1016/j.fgb.2012.05.002.
- [80] MacLeod, D., M., Muller-Kogler, E., Wilding N., 1976. Entomophthora species with E. muscae-like conidia. *Mycologia*, 68: 1–29.
- [81] Mains, E., B., 1959. North American Species of *Ashersonia* Parasitic on Aleyrodidae. *Journal of Insect Pathology*, 1:43-47.
- [82] Marmaras, V. J., N. D. Charalambidis, C. G. Zervas., 1996. Immune response in insects: the role of phenoloxidase in defense reactions in relation to melanization and sclerotization. *Archives of Insect Biochemistry and Physiology* 31:119-133.
- [83] Mayerhofer, J., Lutz, A., Widmer, F., Rehner, S. A., Leuchtman, A., Enkerli, J., 2015. Multiplexed microsatellite markers for seven *Metarhizium* species. *Journal of Invertebrate Pathology.* 132, 132–134.
- [84] McCoy, C. W., 1981. Pest Control by the fungus *Hirsutella thompsonii* pp. 449-512. In: Burges H. D. (ed) Academic Press, Londres. *Microbial Control of Pest and Plant Diseases* 1970-1980.
- [85] Meyling NV, Thorup-Kristensen K; Eilenberg J., 2011. Below- and aboveground abundance and distribution of fungal entomopathogens in experimental conventional and organic cropping systems. *Biological Control*, 59: 180-186.
- [86] Meyling, N. V., Eilenberg, J., 2007. Ecology of the entomopathogenic fungi *Beauveria bassiana* and *Metarhizium anisopliae* in temperate agroecosystems: potential for conservation biological control. *Biol. Control* 43, 145–155.

- [87] Meyling, N. V., Lübeck, M., Buckley E. P., Eilenbera, J., Rehner, S. A., 2009. Community composition, host range and genetic structure of the fungal entomopathogen *Beauveria* in adjoining agricultural and seminatural habitats. *Molecular Ecology*. 18, 1282-1293.
- [88] Meyling, N. V., Pilz, C., Keller, S., Widmer, S. Enkerli, J., 2012. Diversity of *Beauveria* spp. isolates from pollen beetles *Meligethes aeneus* in Switzerland. *Journal of Invertebrate Pathology*. 109, 76-82.
- [89] Motta-Delgado, P. A., and Murcia-Ordoñez, B., 2011. Hongos entomopatógenos como alternativa para el control biológico de plagas/Entomopatogenic fungi as an alternative for biological pest control/fungos entomopatogênicos como alternativa para o controle biológico de pragas. [Entomopathogenic fungi as an alternative for the biological control of pests / Entomopatogenic fungi as an alternative for biological pest control / entomopathogenic fungi as an alternative for or biological control of pragas]. *Revista Ambiente & Água*, 6(2), 77.
- [90] Nielsen, C., Jensen, A. B., Eilenberg, J., 2008. Survival of entomophthoralean fungi infecting aphids and higher flies during unfavorable conditions and implications for conservation biological control. In: Ekesi, S., Maniania, N. K. (Eds.), Use of Entomopathogenic Fungi in Biological Pest Management. *Research SignPost*, Kerala, India, 13–38.
- [91] Oliveira, I., Pereira, J. A., Quesada-Moraga, E., Lino-Neto, T., Bento, A., & Baptista, P., 2013. Effect of soil tillage on natural occurrence of fungal entomopathogens associated to *Prays oleae* Bern. *Scientia Horticulturae*, 159, 190-196.
- [92] Ortiz-Urquiza A, Luo Z, Keyhani NO., 2015. Improving mycoinsecticides for insect biological control. *Appl Microbiol Biotechnol.*, 99(3):1057-68. DOI: 10.1007/s00253-014-6270-x.
- [93] Ortiz-Urquiza, A., Keyhani, N.O., 2013. Action on the surface: Entomopathogenic fungi versus the insect cuticle. *Insects*, 4:357-374; DOI:10. 3390/insects4030357.
- [94] Oulevey, C., Widmer, F., Kölliker, R., Enkerli, J., 2009. An optimized microsatellite marker set for detection of *Metarhizium anisopliae* genotype diversity on field and regional scales. *Mycol. Res.* 113, 1016–1024.
- [95] Pantou, M. P., Mavridou, A., Typas, M. A., 2003. IGS sequence variation, group-I introns and the complete nuclear ribosomal DNA of the entomopathogenic fungus *Metarhizium*: excellent tools for isolate detection and phylogenetic analysis. *Fungal Genet. Biol.* 38, 159.
- [96] Pérez, C. N., 2004. *Manejo ecológico de plagas*. [Ecological management of pests]. Primera edición. CEDAR.
- [97] Peterson, S. W., 2008. Phylogenetic analysis of *Aspergillus* species using DNA sequences from four *loci*. *Mycologia* 100, 205-226.
- [98] Punya, J., Swangmaneecharern, P., Pinsupa, S., Nitistaporn, P., Phonghanpot, S., Kunathigan, V., Cheevadhanarak, S., Tanticharoen, M., Amnuaykanjanasin, A., 2015. Phylogeny of type I polyketide synthases (PKSs) in fungal entomopathogens and expression analysis of PKS genes in *Beauveria bassiana* BCC 2660. *Fung. Biol.* 119, 538-550.
- [99] Qin Y, Ortiz-Urquiza A, Keyhani NO., 2014. A putative methyltransferase, *mtrA*, contributes to development, spore viability, protein secretion and virulence in the

- entomopathogenic fungus *Beauveria bassiana*. *Microbiology*. 160(Pt 11):2526-37. DOI: 10.1099/ mic.0.078469-0.
- [100] Quiroz-Velasquez, P. F., Abiff, S. K., Fins, K. C., Conway, Q. B., Salazar, N. C., Delgado, A. P., Dawes, J. K., Douma, L. G., Tartar, A., 2014. Transcriptome analysis of the entomopathogenic oomycete *Lagenidium giganteum* reveals putative virulence factors. *Appl. Environ. Microbiol.* 80(20): 6427-6436.
- [101] Rehner, S. A., Buckley, E. P., 2003. Isolation and characterization of microsatellite loci from the entomopathogenic fungus *Beauveria bassiana* (Ascomycota: Hypocreales). *Mol. Ecol. Notes* 3, 409–411.
- [102] Rehner, S. A., Buckley, E. P., 2005. A *Beauveria* phylogeny inferred from nuclear ITS and EF1 α sequences: evidence for cryptic diversification and links to *Cordyceps telemorphs*. *Mycol.* 97, 84–98.
- [103] Rehner, S. A., Posada, F., Buckley, E. P., Infante, F., Castillo, A., Vega, F. E., 2006. Phylogenetic origins of African and Neotropical *Beauveria bassiana* s.l. pathogens of the coffee berry borer, *Hypothenemus hampei*. *J. Invert. Pathol.* 93, 11–21.
- [104] Reineke, A., Brischoff-Schaefer, M., Rondot, Y., Galidevara, S., Hirsch, J., Devi, K. U., 2014. Microsatellite markers to monitor a commercialized isolate of the entomopathogenic fungus *Beauveria bassiana* in different environments: Technical validation and first applications. *Biological Control*. 70, 1–8.
- [105] Rosas-García NM, Ávalos-de-León O, Villegas-Mendoza JM, Mireles-Martínez M, Barboza-Corona JE, Castañeda-Ramírez JC., 2014. Correlation between pr1 and pr2 gene content and virulence in *Metarhizium anisopliae* strains. *J. Microbiol. Biotechnol.* 24(11): 1495-1502 <http://dx.Doi.org/10.4014/jmb.1404.04044>.
- [106] Rosewich, U. L., McDonald, B. A., 1994. DNA fingerprinting in fungi. *Methods Mol. Cel. Biol.* 5, 41-48.
- [107] Rostami, F., Khosravi, M. F., Kazem, S. S., Saeidi, S., 2015. Comparison of PCR-RFLP based on ribosomal regions and ssr markers in genetic diversity of pistachio die-back caused by *Paecilomyces variotii*. *Gene Cell Tissue.* 2(1), 2-4.
- [108] Samson, R. A., 1974. *Paecilomyces* and some allied Hyphomycetes. *Studies in Mycology*, 6: 80-83.
- [109] Samson, R. A., McCoy, C. W., O'Donnell, K. L., 1980. Taxonomy of the acarine parasite *Hirsutella thompsonii*. *Mycologia*, 72: 359-377.
- [110] Samson, R. A., Visagie, C. M., Houbraken, J., Hong, S. B., Hubka, V., Klaassen, C. H. W., Perrone, G., Seifert, K. A., Susca, A., Tnney, J. B., Varga, J., Kocsubé, S., Szigeti, G., Yaguchi, T., Frisvad, J. C., 2014. Phylogeny, identification and nomenclature of the genus *Aspergillus*. *Studies in Mycology* 78, 141-173.
- [111] Sánchez-Pérez, L. C, Barranco-Florido, J. E., Rodríguez-Navarro, S., Cervantes-Mayagoitia, J. F., Ramos-López, M. Á., 2014. Enzymes of entomopathogenic fungi, advances and insights. *Advances in Enzyme Research*, 2:65-76. <http://dx.Doi.org/10.4236/aer.2014.22007>.
- [112] Sansores-Lara LI, Palomo-Kumul J, Mijangos C., 2014. *Manual de producción y formulación de hongos entomopatógenos: Manejo de hongos entomopatógenos*, EAE. [Manual of production and formulation of entomopathogenic fungi: Management of entomopathogenic fungi, EAE].
- [113] Sandhu, S. S., Sharma A. K., Beniwal V., Goel G., Batra P., Kumar A., Jaglan S., Sharma AK., Malhotra S., 2012. Myco-biocontrol of insect pests: Factors involved, mechanism

- and regulation. *J. Pathogens*, Vol. 2012. 10.1155/2012/126819. DOI:10.1155/2012/126819.
- [114] Schoch, C. L., Seifert, K. A., Huhndorf, S., et al., 2012. Nuclear ribosomal internal transcribed spacer (ITS) region as a universal DNA barcode marker for Fungi. *Proceedings of the National Academy of Sciences of the USA* 109, 6241–6246.
- [115] Sevim A, Donzelli BGG, Wu D, Demirbag Z, Gibson DM, Turgeon BG., 2012. Hydrophobin genes of the entomopathogenic fungus, *Metarhizium brunneum*, are differentially expressed and corresponding mutants are decreased in virulence. *Curr Genet* 58:79-92. DOI 10.1007/s00294-012-0366-6.
- [116] Shahid, A. A., Rao, A. Q., Bakhsh, A., Husnain, T., 2012. Entomopathogenic fungi as biological controllers: New insights into their virulence and pathogenicity. *Arch. Biol. Sci.*, Belgrade, 64 (1): 21-42, DOI:10. 2298/ABS1201021S21.
- [117] Steinwender, B. M., Enkerli, E., Widmer, F., Eilenber, J., Thorup-Kristensen, K., Meyling, N. V., 2014. Molecular diversity of the entomopathogenic fungal *Metarhizium* community within an agroecosystem. *Journal of Invertebrate Pathology*. 123, 6–12.
- [118] Sugimoto, M., Koike M., Hiyama N., Nagao H., 2002. Genetic, morphological, and virulence characterization of the entomopathogenic fungus *Verticillium lecanii*. *Journal of Invertebrate Pathology* 82:176–187.
- [119] Tanada, Y. and H. K. Kaya., 1993. *Insect pathology*. Primera edición. Academic Press, San Diego, California.
- [120] Tellez, J. A; Cruz, R. M. G; Mercado, F. Y; Assaf, T.; Arana, C. A., 2009. Mecanismos de acción y respuesta en la relación de hongos entomopatógenos e insectos. [Mechanisms of action and response in the relation of entomopathogenic fungi and insects]. *Revista Mexicana de Micología*. 30: 73-80.
- [121] Tigano-Milani, M. S., Honeycutt, J., Lacey, L. A., Assis, R., McClelland, M., Sobral, W. S., 1995. Genetic variability of *Paecilomyces fumosoroseus* isolates revealed by molecular markers. *J. Inv. Pathol*. 65, 274-282.
- [122] Tulloch, M., 1976. Genus *Metarhizium*. *T. Brit. Mycol. Soc.* 66, 407–411.
- [123] Valero-Jiménez CA, Wiegers H, Zwaan BJ, Koenraadt CJM, van Kan JAL., 2016. Genes involved in virulence of the entomopathogenic fungus *Beauveria bassiana*. *Journal of Invertebrate Pathology*. 133: 41-49.
- [124] Vega, F. E., Goettel, M. S., Blackwell, M., Chandler, D., Jackson, M. A., Keller, S., and Pell, J. K., 2009. Fungal entomopathogens: new insights on their ecology. *Fungal Ecology*, 2(4), 149-159.
- [125] Velasquez, V. B., Carcamo, M. P., Merino, C. R., Iglesias, A. F., Duran, J. F., 2007. Intraspecific differentiation of Chilean isolates of the entomopathogenic fungi *Metarhizium anisopliae* var. *anisopliae* as revealed by RAPD, SSR and ITS markers. *Genetics Mol. Biol.* 30, 89–99.
- [126] Ventura-Sobrevilla J., Boone D., Rodriguez-Herrera R., Martinez-Hernandez J. L., Aguilar C., 2015. Microbial biosynthesis of enzymes for food applications. In edited by Rickey Yada. *Improving and Tailoring Enzymes for Food Quality and Functionality*. Elsevier. 85-99.
- [127] Vilcinskas, A., 2010. Coevolution between pathogen-derived proteinases and proteinase inhibitors of host insects, *Virulence*, 1:3, 206-214, DOI: 10.4161/viru.1.3.12072.

- [128] Visalakshy, P. G., Krishnamoorthy, A., & Kumar, A. M., 2005. Note: Effects of plant oils and adhesive stickers on the mycelia growth and conidiation of *Verticillium lecanii*, a potential entomopathogen. *Phytoparasitica*, 33(4), 367-369.
- [129] Wang C and St Leger RJ., 2007. The MAD1 Adhesin of *Metarhizium anisopliae* links adhesion with blastospore production and virulence to insects, and the MAD2 adhesin enables attachment to plants. *Eukaryot Cell* 6(5): 808-816. DOI: 10.1128/EC.00409-06.
- [130] Wang Z, Meng H, Zhuang Z, Chen M, Xie L, Huang Bo., 2013. Molecular cloning of a novel subtilisin-like protease (Pr1A) gene from the biocontrol fungus *Isaria farinose*. *Appl Entomol Zool* 48:477–487 DOI: 10.1007/s13355-013-0208-0.
- [131] Wang ZL, Li F, Li C, Feng MG., 2014. Bbssk1, a response regulator required for conidiation, multi-stress tolerance, and virulence of *Beauveria bassiana*. *Appl Microbiol Biotechnol*. 98(12):5607-18. DOI: 10.1007/s00253-014-5644-4.
- [132] White, T. J., Bruns, T., Lee, S., Taylor, J., 1990. Amplification and direct sequencing of fungal ribosomal RNA genes for phylogenetics. In *PCR Protocols: a guide to methods and applications* (M. A. Innis, D. H. Gelfand, J. J. Sninsky & T. J. White, eds): 315-322. Academic Press, San Diego.
- [133] Wichadakul, D., Kobmoo, N., Ingsriswang, S., Tangphatsornruang, S., Chantasingh, D., Luangsa, J. J., Eurwilaichitr, L., 2015. Insights from the genome of *Ophiocordyceps polyrhachis-furcata* to pathogenicity and host specificity in insect fungi. *BMC Genomics*, 16:881.
- [134] Xiao, G., Ying, S. H., Zheng, P., Wang, ZL., Zhang, S., Xie, X. Q., Shang, Y., Leger, R. J., Zhao, G. P., Wang, Ch., Feng, M.G., 2012. Genomic perspectives on the evolution of fungal entomopathogenicity in *Beauveria bassiana*. *Scientific Reports*, 2: 483. DOI: 10.1038/srep00483.
- [135] Xie X-Q, Wang J, Huang B-F, Ying S-H, Feng M-G., 2010. A new manganese superoxide dismutase identified from *Beauveria bassiana* enhances virulence and stress tolerance when overexpressed in the fungal pathogen. *Appl Microbiol Biotechnol* 86:1543-1553. DOI: 10.1007/s00253-010-2437-2.
- [136] Xie XQ, Guan Y, Ying SH, Feng MG., 2013. Differentiated functions of Ras1 and Ras2 proteins in regulating the germination, growth, conidiation, multi-stress tolerance and virulence of *Beauveria bassiana*. *Environ Microbiol*. 15(2):447-62. DOI: 10.1111/j.1462-2920.2012. 02871.x.
- [137] Zare, R., Gams, W., 2001. A revision of *Verticillium* sect. Prostata IV. The genera *Lecanicillium* and *Simplicillium* gen. *Nova Hedwigia*, 73:1-50.
- [138] Zhang S, Xia YX, Kim B, Keyhani NO., 2011. Two hydrophobins are involved in fungal spore coat rodlet layer assembly and each play distinct roles in surface interactions, development and pathogenesis in the entomopathogenic fungus, *Beauveria bassiana*. *Mol Microbiol*. 80(3):811-26. DOI: 10.1111/j.1365-2958.2011.07613.x.
- [139] Zheng, P., Xia, Y., Xiao, G., Xiong, Ch., Hu, X., Zhang, S., Zheng, H., Huang, Y., Zhou, Y., Wang, S., Zhao, G. P., Liu, X., Leger, R. J., Wang, Ch., 2011. Genome sequence of the insect pathogenic fungus *Cordyceps militaris*, a valued traditional Chinese medicine. *Genome Biology*, 12:R116.
- [140] Zimmermann, G., 2007. Review on safety of the entomopathogenic fungus *Metarhizium anisopliae*. *Biocontrol Sci. Technol*. 17, 879–920.
- [141] Zimmermann, G., 1986. The Galleria baits method for detection of entomopathogenic fungi in soil. *J. Appl. Entomol*. 102:213-215.

BIOGRAPHICAL SKETCH

Name: Raúl Rodríguez-Herrera

Affiliation: School of Chemistry- Universidad Autónoma de Coahuila

Education: Agriculture minor Horticulture. 1981. Universidad Autónoma Agraria Antonio Narro. México

M. S. Plant Breeding. 1986. Universidad Autónoma Agraria Antonio Narro. México

Ph. D. Plant Breeding 1999. Texas A&M University. USA.

Post-doctorate and sabbatical training 1999, 2014. United States Department of Agriculture. USA

Address: School of Chemistry

Universidad Autónoma de Coahuila

Blvd. V. Carranza e Ing. José Cárdenas S/n

Col. Republica Saltillo Coahuila 25280 México

Research and Professional Experience:

1999-to date Full professor- Universidad Autónoma de Coahuila- México

1999 Post-doctorate training Texas A&M University-USDA USA

1997-1998 Graduate Assistant Texas A&M University-USA

1986-1998 Researcher-National Research Institute for Forestal, Agricultural and Livestock-Mexico

1982-1984 Administrator- Ministry of Agrarian Reform-Mexico

Professional Appointments:

Mexican Society of Biotechnology and Bioengineering. México

Directive Committee Symposium “Genomes and Proteomes in XXI Century.” México

Institute of Food Technologists. USA

Mexican Association of Food Science (AMECA). Mexico

Honours

1981. Graduated with honours from the UAAAN.

1988. Award to outstanding research. PCCMCA Congress. San José, Costa Rican.

1990. Award to outstanding research. PCCMCA Congress. San Salvador, El Salvador.

2001. Distinguished as National Researcher- Level 2 Mexico.

2003. National Prize on Food Science and Technology. CONACYT-Coca Cola México, D. F.

2005. Coahuila State Prize for outstanding Food Research Project. Saltillo Coahuila, Mexico.

2005. National Agro-Bio Prize on Agricultural Biotechnology-Mexico

Publications Last 3 Years:

Selected Publications (Total 216 refereed publications, 60 divulgation papers, 51 book chapters, 5 books, 6 patents, 8 bulletins, recent refereed publications are listed).

1. Meléndez Rentería, NP., Aguilar, CN., Rodríguez Herrera, R., Nevárez Moorillón, GV., 2013. Microbiological effect of fermented Mexican oregano (*Lippia berlandieri* Schauer) waste. *Waste and Biomass Valorization*. DOI 10.1007/s12649-013-9222-2
2. M. Cruz-Requena, H. De La Garza-Toledo, C. N. Aguilar-González, A. Aguilera-Carbó, H. Reyes-Valdés, M. Rutiaga and R. Rodríguez-Herrera. 2013. Chemical and molecular properties of sotol plants (*Dasyilirion cedrosanum*) of different sex and its fermentation products. *International Journal of Basic and Applied Chemical Sciences* 3 (1):. 41-49.
3. Diana B. Muñiz-Márquez, Guillermo C. Martínez-Ávila, Jorge E. Wong-Paz, Ruth Belmares-Cerda, Raúl Rodríguez-Herrera, Cristóbal N. Aguilar. 2013. Ultrasound-assisted extraction of phenolic compounds from *Laurus nobilis* L. and their antioxidant activity.” *Ultrasonics Sonochemistry* 20: 1149–1154. doi: 10.1016/j.ultsonch.2013.02.008.
4. Juan A. Ascacio-Valdés, José J. Buenrostro, Reynaldo De la Cruz, Leonardo Sepúlveda, Antonio F. Aguilera, Arely Prado, Juan C. Contreras, Raúl Rodríguez and Cristóbal N. Aguilar. 2014. Fungal biodegradation of pomegranate ellagitannins. *Basic Journal of Microbiology*. 54:28-34. DOI 10.1002/jobm.201200278.
5. Ascacio-Valdés, Juan, Burboa, Edgardo, Aguilera-Carbo Antonio F., Aparicio Mario, Pérez-Schmidt Ramón, Rodríguez Raúl, Aguilar Cristóbal N. (2013). An antifungal ellagitannin isolated from *Euphorbia antisiphilitica* Zucc. *Asian Pac J Trop Biomed*. 3(1): 41–46. doi: 10.1016/S2221-1691(13)60021-0. IF = 0.371.
6. T. López, A. Prado-Barragán, G.V. Nevárez-Moorillón, J. C. Contreras, R. Rodríguez and C.N. Aguilar. 2013. Incremento de la capacidad antioxidante de extractos de pulpa de café por fermentación láctica en medio sólido CyTA – *Journal of Food*, [Increase in antioxidant capacity of coffee pulp extracts by lactic acid fermentation in solid medium CyTA - *Journal of Food*], 11(4):359-365. dx.doi.org/10. 1080/19476337.2013.773563.
7. Luis C. Mata, Catalina Chavez, Raúl Rodríguez-Herrera, Daniel Hernández-Castillo and Cristobal N. Aguilar. 2013. Growth inhibition of some phytopathogenic bacteria by cell-free extracts from *Enterococcus* sp. *British Biotechnology Journal*, 3(3): 359-366.
8. Emilio Ochoa-Reyes, Gabriela Martínez-Vazquez, Saul Saucedo-Pompa, Julio Montañez, Romeo Rojas-Molina, Miguel A. de Leon-Zapata, Raúl Rodríguez-Herrera and Cristóbal N. Aguilar. 2013. Improvement of shelf life quality of green bell peppers using edible coating formulations. *Journal of Microbiology, Biotechnology and Food Sciences*, 2 (6) 2448-2451
9. Tanya Cecilia Espinoza-Hernández, Raúl Rodríguez-Herrera, Cristóbal Noé Aguilar-González, Faustino Lara-Victoriano, Manuel Humberto Reyes-Valdés and Francisco Castillo- Reyes. 2013. Characterization of three novel pigment-producing *Penicillium* strains isolated from the Mexican semi-desert. *African Journal of Biotechnology* 12(22): 3405-3413.
10. Lorena Lara-Fernández, Heliodoro de la Garza-Toledo, Jorge E. Wong-Paz, Ruth Belmares, Raúl Rodríguez-Herrera and Cristóbal N. Aguilar. 2013. Separation conditions and evaluation of antioxidant properties of boldo (*Peumus boldus*) extracts. *Journal of Medicinal Plant Research*, 7(15): 911-917.
11. Delgado-García, M. Cruz Hernández, M.A., De la Garza-Rodríguez, I., Balagurusamy N, Aguilar, C., Rodríguez-Herrera, R. 2013. Characterization of halophilic microorganisms isolated from Mexican saline soils, *ARPN Journal of Agricultural and Biological Science*, 8(6):457-464.

12. Ascacio-Valdés, J., Aguilera-Carbo Antonio F., Rodríguez Raúl, Aguilar Cristóbal N. (2013). Análisis de ácido elágico en algunas plantas del semidesierto mexicano. [Analysis of ellagic acid in some plants of the Mexican semi-desert]. *Revista Mexicana de Ciencias Farmacéuticas*. 44(2): 36-40.
13. Y.N. Mora, J.C. Contreras, C. N. Aguilar, P. Meléndez, I. De la Garza, R. Rodríguez, 2013. Chemical composition and functional properties from different sources of dietary fiber. *American Journal of Food and Nutrition*. 1(3):27-33. DOI:10.12691/ajfn-1-3-2.
14. Mondragon-Preciado, G., Escalante-Minakata, P., Osuna-Castro, J.A., Ibarra-Junquera, V., Morlett-Chavez, J.A., Aguilar-Gonzalez, C.N., Rodríguez-Herrera, R., 2013. Bacteriocinas: características y aplicación en alimentos. [Bacteriocins: characteristics and application in food]. *Investigacion y Ciencia* 59: 64-70.
15. M. Cruz-Requena, C.N. Aguilar-González, J. Espinoza-Velázquez, M.H. Reyes-Valdés and R. Rodríguez-Herrera. 2013. AFLPs loci associated with polyembryonic maize using selective genotyping analysis. *Israel Journal of Plant Sciences*. 61(1-4):46-50. DOI: 10.1080/07929978.2014.950045
16. Jorge E. Wong-Paz, Diana B. Muñiz-Márquez, Pedro Aguilar-Zárate, Raúl Rodríguez-Herrera and Cristóbal N. Aguilar. 2013. Microplate quantification of total phenolic content from plant extracts obtained by conventional and ultrasound methods. *Phytochem. Anal.* DOI 10.1002/pca.2512.
17. D.B. Muñiz-Márquez, R. Rodríguez, N. Balagurusamy, M.L. Carrillo, R. Belmares, J.C. Contreras, G.V. Nevárez & C.N. Aguilar (2014) Phenolic content and antioxidant capacity of extracts of *Laurus nobilis* L., *Coriandrum sativum* L. and *Amaranthus hybridus* L., *CyTA - Journal of Food*, 12:3, 271-276, DOI: 10.1080/19476337.2013.847500
18. V. Padilla-García, F. Castillo-Reyes, C. N. Aguilar-González, A. Cuenca-Arana, A. Téllez-Jurado, M. H. Reyes-Valdez and R. Rodríguez-Herrera 2014. Molecular characterization of six tannase-producers *Aspergillus* strains using PCR-RFLP of ITS and IGS regions and RAPD's. *International Journal of Molecular Biology* 8(12): 451-459.
19. GCG Martinez-Avila, A.F. Aguilera, S. Saucedo, R. Rojas, R. Rodriguez and CN Aguilar. 2014. Fruit wastes fermentation for phenolic antioxidants production and their application in manufacture of edible coatings and films, *Critical Reviews in Food Science and Nutrition*. 54(3):303-11 DOI:10.1080/10408398.2011.584135 ISSN 1040-8398 (Print), 1549-7852 (Online). JCR FI= 4.510.
20. Salvador Eduardo Vásquez Rivera, Mayra Alejandra Escobar-Saucedo, Diana Morales, Cristóbal Noé Aguilar and Raúl Rodríguez-Herrera. 2014. Synergistic effects of ethanolic plant extract mixtures against food-borne pathogen bacteria. *African Journal of Biotechnology* 13(5): 699-704.
21. Norma P. Meléndez, Virginia Nevárez-Moorillón, Raúl Rodríguez-Herrera, José C. Espinoza and Cristóbal N. Aguilar. 2014. A microassay for quantification of 2,2-diphenyl-1-picrylhydrazyl (DPPH) free radical scavenging. *African Journal of Biochemistry Research* 8(1):14-18.
22. Dulce A. Flores-Maltos, Solange I. Mussatto, Juan C. Contreras Esquivel, Juan J. Buenrostro, Raúl Rodríguez, José A. Teixeira and Cristóbal N. Aguilar. 2014. Typical Mexican agroindustrial residues as supports for solid-state fermentation. *American Journal of Agricultural and Biological Sciences* 9 (3): 289-293.

23. Reynaldo De la Cruz Quiroz, Sevastianos Roussos, Daniel Hernández, Raúl Rodríguez, Francisco Castillo, and Cristóbal N. Aguilar. 2014. Challenges and opportunities of the bio-pesticides production by solid-state fermentation: filamentous fungi as a model. *Critical Reviews in Biotechnology*. (doi:10.3109/07388551.2013.857292).
24. Veana F., Fuentes-Garibay J.A., Aguilar C.N., Rodríguez-Herrera R., Guerrero-Olazarán M., Viader-Salvadó J. M. (2014). Gene encoding a novel invertase from a xerophilic *Aspergillus niger* strain and production of the enzyme in *Pichia pastoris*. *Enzyme and Microbial Technology* 63: 28-33.
25. M. Delgado, J. Mendez, R. Rodríguez-Herrera, C.N. Aguilar, M. Cruz-Hernández & N. Balagurusamy (2014): Characterization of phosphate-solubilizing bacteria isolated from the arid soils of a semi-desert region of north-east Mexico, *Biological Agriculture & Horticulture: An International Journal for Sustainable Production Systems*, 30(3):211–217. DOI: 10.1080/01448765.2014.909742
26. Jorge E. Wong-Paz, Juan C. Contreras-Esquivel, Diana Muñoz-Marquez, Ruth Belmares, Raul Rodriguez, Patricia Flores and Cristobal N. Aguilar. 2014. Microwave-assisted extraction of phenolic antioxidants from semiarid plants. *American Journal of Agricultural and Biological Sciences* 9 (3): 299-310. ISSN: 1557-4989doi:10.3844/ajabssp.2014. 299.310.
27. K. Cruz-Aldaco, C. N. Aguilar, A. Ilyina, R. Rodríguez-Herrera and J. L. Martínez-Hernández. 2014. Surface adhesion fermentation for lipase production by *Mucor griseocyanus* *Micol. Apl. Int.*, 26(1): 9-16.
28. J.A. Borrego-Terrazas, F. Lara-Victoriano, A.C. Flores-Gallegos, F. Veana, C.N. Aguilar and R. Rodríguez-Herrera. 2014. Nucleotide and amino acid variation of tannase gene from different xerophylic *Aspergillus* strains. *Canadian Journal of Microbiology*. 60: 509–516. 10.1139/cjm-2014-0163.
29. Chávez-González, Mónica L.; Guyot, Sylvain; Rodríguez-Herrera, Raúl, Prado-Barragán, Arely; Aguilar, Cristóbal N. 2014. Production profiles of phenolics from fungal tannic acid biodegradation in submerged and solid-state fermentation. *Process Biochemistry*. 49(4):541-546.
30. Ayerim Hernández-Almanza, Julio Montanez-Sáenz, Cristian Martínez-Ávila, Raúl Rodríguez-Herrera, Cristóbal N. Aguilar. 2014. Carotenoid production by *Rhodotorula glutinis* YB-252 in solid-state fermentation. *Food Bioscience* 7:31–36.
31. Hernández-Almanza, Ayerim; Cesar Montanez, Julio; Aguilar-González, Miguel A.; Martínez-Ávila, Cristian; Rodríguez-Herrera, Raúl; Aguilar, Cristóbal N. 2014. *Rhodotorula glutinis* source of pigments and metabolites for food industry. *Food Bioscience*. 5(3):64-72.
32. Burboa E.A., Ascacio V., J.A., Zugasti C., A., Rodriguez H., R., Aguilar G., C.N. 2014. Capacidad antioxidante y antibacteriana de extractos de residuos de candelilla. [Antioxidant and antibacterial capacity of candelilla waste extracts]. *Revista Mexicana de Ciencias Farmaceuticas*. 45(1):51-56.
33. Sepúlveda, L.; Buenrostro -Figuroa, J. J.; Ascacio-Valdés, J A.; Aguilera-Carbó, A. F.; Rodríguez -Herrera, R.; Contreras-Esquivel, J. C.; Aguilar, C. N. 2014. Submerged culture for production of ellagic acid from pomegranate husk by *Aspergillus niger* GH1. *Micología Aplicada Internacional*, 26(2): 27-35.

34. Veana F., Martínez-Hernández J. L., Aguilar C. N., Rodríguez-Herrera R. and Michelena G. 2014. Utilization of molasses and sugar cane bagasse for production of fungal invertase in solid state fermentation using *Aspergillus niger* GH1. *BMJ*, 45(2): 373-377.
35. JuanBuenrostro-Figueroa, Alberto Ascacio-Valdés, Leonardo Sepúlveda, Reynaldo De la Cruz, Arely Prado-Barragán, Miguel A. Aguilar-González, Raúl Rodríguez, Cristóbal N. Aguilar (2014). Potential use of different agroindustrial by-products assupports for fungal ellagitannase production undersolid-state fermentation. *Food and Bioproducts Processing* 9(2): 376–382.
36. Buenrostro, J., Huerta, S., Prado, A., Ascacio, J., Sepulveda, L., Rodriguez-Herrera R., Aguilera, A., Aguilar, C.N., 2014. Continuous production of ellagic acid in a packed-bed reactor. *Process Biochemistry* 49: 1595–1600.
37. Reynaldo de la Cruz, Juan A. Ascacio, José J. Buenrostro, Leonardo Sepúlveda, Raúl Rodríguez, Arely Prado-Barragán, Juan C. Contreras, Antonio Aguilera & Cristóbal N. Aguilar. 2014. Optimization of ellagitannase production by *Aspergillus niger* GH1 by solid state fermentation. *Preparative Biochemistry & Biotechnology*. 45:7, 617-631, DOI:10.1080/10826068.2014.940965.
38. Flores-Gallegos AC., Morlett-Chávez JA., Aguilar CN., Riutort M., Rodríguez-Herrera R., 2014. Gene encoding inulinase isolated from *Penicillium citrinum* ESS and its molecular phylogeny. *Applied Biochemistry and Biotechnology*. 175:1358-1370 DOI 10.1007/s12010-014-1280-9
39. Flores-Gallegos, A.C., Morlett-Chavez J.A., Rodríguez-Herrera R. 2014. Inulin potential for enzymatic obtaining of prebiotic oligosaccharides. *Critical Reviews in Food Science and Nutrition*. DOI:10.1080/10408398.2013.807220.
40. Delgado-García, M., Aguilar, C.N., Contreras-Esquivel, J.C. and Rodríguez-Herrera, R. 2014. Screening for extracellular hydrolytic enzymes production by different halophilic bacteria. *Mycopath* 12(1): 17-23.
41. F. Veana, D. G. Rodríguez-Reyna, E. T. Aréchiga-Carvajal, C. N. Aguilar and R. Rodríguez-Herrera. 2014. Isolation of a putative Invertase gene from the xerophilic *Aspergillus niger* GH1 strain. *Mycopath* (2014) 12(2): 77-82.
42. M. Govea Salas, A.M. Rivas Estilla, Jesus Morlett, Sonia Amelia Lozano Sepúlveda, R. Rodríguez Herrera, C.N. Aguilar González. 2014. P420 gallic acid has antiviral effect against hepatitis c virus (hcv), which is mediated by its antioxidant activity. *Journal of Hepatology* (Impact Factor: 11.34). 60(1):S208. DOI: 10.1016/S0168-8278(14)60582-.
43. Naivy Y. Nava-Cruz, Miguel A. Medina-Morales, Jose L. Martinez, R. Rodriguez, and Cristobal N. Aguilar. 2015. Agave biotechnology: an overview. *Critical Reviews in Biotechnology*. 35(4):546-559. (doi:10.3109/07388551.2014.923813).
44. Silvia Marina González-Herrera, Raul Rodriguez Herrera, Mercedes Guadalupe López, Olga Miriam Rutiaga, Cristobal Noe Aguilar, Juan Carlos Contreras Esquivel, Luz Araceli Ochoa Martínez, (2015), "Inulin in food products: prebiotic and functional ingredient," *British Food Journal*, 117 (1): 371 – 387.
45. Jorge E. Wong-Paz, Juan C. Contreras-Esquivel, Raúl Rodríguez-Herrera, María L. Carrillo-Inungaray, Lluvia I. López, Guadalupe V. Nevárez-Moorillón, Cristóbal N. Aguilar. (2015). Total phenolic content, in vitro antioxidant activity and chemical composition of plant extracts from semiarid Mexican region. *Asian Pacific Journal of Tropical Medicine* 8(2): 104-111.

46. Virgilio Cruz, Romeo Rojas, Saúl Saucedo-Pompa, Dolores G. Martínez, Antonio F. Aguilera-Carbó, Olga B. Alvarez, Raúl Rodríguez, Judith Ruiz, and Cristóbal N. Aguilar. 2015. Improvement of shelf life and sensory quality of pears using a specialized edible coating. *Journal of Chemistry* 1-7. Article ID 138707.
47. Miguel A. De León-Zapata, Aide Sáenz-Galindo, Romeo Rojas-Molina, Raúl Rodríguez-Herrera, Diana Jasso-Cantú, Cristóbal N. Aguilar. 2015. Edible candelilla wax coating with fermented extract of tarbush improves the shelf life and quality of apples. *Food Packaging and Shelf Life*. 3:70-75 doi:10.1016/j.fpsl.2015.01.001.
48. Jose Antonio Fuentes-Garibay, Cristobal Noe Aguilar, Raul Rodriguez-Herrera, Martha Guerrero-Olazarán, Jose Maria Viader-Salvado. 2015. Tannase sequence from a xerophilic *Aspergillus niger* strain and production of the enzyme in *Pichia pastoris*. *Molecular Biotechnology* 57:439–447. DOI 10.1007/s12033-014-9836-z.
49. Flores-Gallegos, A.C., Morlett-Chávez, J.M., Contreras-Esquivel, J.C., Aguilar, C.N., Rodríguez-Herrera, R. 2015. Comparative study of Mexican fungal strains for thermostable inulinase production. *Journal of Bioscience and Bioengineering*. 119 (4):421-426, <http://dx.doi.org/10.1016/j.jbiosc.2014.09.020>.
50. Castillo-Reyes, F., Hernández-Castillo, FD., Gallegos-Morales, G., Flores-Olivas, A., Rodríguez-Herrera, R., Aguilar, CN., 2015. Efectividad in vitro de Bacillus y polifenoles de plantas nativas de México sobre Rhizoctonia-Solani. [In vitro effectiveness of Bacillus and polyphenols of plants native to Mexico on Rhizoctonia-Solani]. *Revista Mexicana de Ciencias Agrícolas*. 6(3): 549-562.
51. Marselino Avendaño-Sánchez, José Espinoza-Velázquez, Alondra Gutiérrez-López, Adriana Carolina Flores-Gallegos, Raúl Rodríguez-Herrera. 2015. Secuencias nucleotídicas de la región ITS en familias S₁ y PL de maíces poliembriónicos. [Nucleotide sequences of the ITS region in S₁ and PL families of polyembryonic maizes]. *Revista Mexicana de Ciencias Agrícolas* 6(3): 509-521.
52. Garduño-Velázquez, S., Quero-Carrillo, AR., Bonnett, D., Rodríguez-Herrera, R., Pérez-Hernández, A., Hernández-Garay, A. 2015. Nivel de ploidía en poblaciones de *Leptochloa dubia* (Kunth) Nees nativas de México. [Level of ploidy in populations of *Leptochloa dubia* (Kunth) Nees native to Mexico]. *Revista Mexicana de Ciencias Agrícolas*, 6(3):539-548.
53. González-Morales S., Cruz-Requena, M., Rodríguez-Vidal A, Aguilar-González C.N., Rebolloso-Padilla O, and Rodríguez-Herrera R 2015. Persistence of transgenic genes and proteins during soybean food processing. *Food Bioscience*, 1 1: 4 3 – 4 7.
54. Marco Mata-Gómez, Solange I. Mussatto, Raul Rodríguez, Jose A. Teixeira, Jose L. Martinez, Ayerim Hernandez, Cristóbal N. Aguilar. 2015. Gallic Acid Production with Mouldy Polyurethane Particles Obtained from Solid State Culture of *Aspergillus niger* GH1. *Applied Biochemistry and Biotechnology* (online). DOI 10.1007/s12010-015-1634-y
55. Santiago Garduño Velázquez, Raúl Rodríguez Herrera, Adrián Raymundo Quero Carrillo, Javier Francisco Enríquez Quiroz, Alfonso Hernández Garay y Alejandra Pérez Hernández. 2015. Diversidad morfológica, citológica y valor nutritivo de siete nuevos genotipos de pasto buffel (*Cenchrus ciliaris* L.) y un cultivar tolerante al frío. [Morphological, physiological and nutritive value of seven new genotypes of buffel grass (*Cenchrus ciliaris* L.) and a cold tolerant cultivar]. *Revista Mexicana de Ciencias Agrícolas* 6(7): 1679-1687.

56. Juan de Dios Hernández-Quintero, M. Humberto Reyes-Valdés, Dulce V. Mendoza-Rodríguez, Martha Gómez-Martínez y Raúl Rodríguez-Herrera. (2015). Estudio de los cromosomas mitóticos y meióticos del sotol (*Dasyilirion cedrosanum* Trel.). [Study of the mitotic and meiotic chromosomes of sotol (*Dasyilirion cedrosanum* Trel.)]. *YTON International Journal of Experimental Botany*. 84(1):107-113
57. Mayra Alejandra Escobar-Saucedo, Raúl Rodríguez-Herrera, Alfonso Reyes-López, Marisol Cruz-Requena, Héctor Flores-Chávez, Cristóbal N. Aguilar. 2014. Análisis genético y bromatológico de siete mutantes de manzano (*Malus domestica* Borkh) del cultivar Golden Delicious. [Genetic and bromatological analysis of seven apple mutants (*Malus domestica* Borkh) from the cultivar Golden Delicious]. *Ecosistemas y Recursos Agropecuarios* 1(3):269-279.
58. Aarón Casas Acevedo, Cristóbal Noé Aguilar González, Heliodoro De La Garza Toledo, Jesús Antonio Morlett Chávez, Didier Montet, Raúl Rodríguez Herrera. 2015. Importancia de las levaduras no-*Saccharomyces* durante la fermentación de bebidas alcohólicas. *Investigacion y Ciencia* 65(5):73-79. Muñiz-Marquez, D. M., Contreras-esquivel J.C., Rodriguez-Herrera R., Wong-Paz, J.E., Texeira A., Musato SI., Aguilar CN., 2015. Influence of thermal effect on sugars composition of Mexican Agave syrup. *CyTA - Journal of Food*, 13:4, 607-612 DOI. .org/10.1080/1947 6337.2015.1028452.
60. Gabriela Macias de la Cerda, Fabiola Veana, Juan Carlos Contreras Esquivel, Cristóbal Noé Aguilar y Raúl Rodríguez Herrera. 2015. Cinética de crecimiento de *Fusarium oxysporum* cultivado en diferentes niveles de glucosa y pectina. [Growth kinetics of *Fusarium oxysporum* grown in different levels of glucose and pectin]. *Revista Investigación y Ciencia* (ACEPTADO).
61. Mayela Govea-Salas, Ana Maria Rivas-Estilla, Raul Rodríguez-Herrera, Sonia A. Lozano-Sepúlveda, Cristobal N. Aguilar-Gonzalez, Alejandro Zugasti-Cruz, Tanya B. Salas-Villalobos and Jesus Antonio Morlett-Chávez. 2015. Gallic acid decreases hepatitis C virus expression through its antioxidant capacity. *Experimental and Therapeutic Medicine*. DOI: 10.3892/etm.2015.2923.
62. Eduardo Osorio, Francisco Daniel Hernández Castillo, Raúl Rodríguez Herrera, Sostenes Edmundo Varela Fuentes, Benigno Estrada Drouaillet y José Alberto López Santillan. 2015. Actividad antagonica de *trichoderma* spp. sobre *Rhizoctonia solani* in vitro. [Antagonistic activity of trichoderma spp. about *Rhizoctonia solani* in vitro]. *Revista Investigación y Ciencia* (ACEPTADO).
63. Francisco Castillo-Reyes, Francisco Daniel Hernández-Castillo, Julio Alberto Clemente-Constantino, Gabriel Gallegos-Morales, Raúl Rodríguez-Herrera and Cristóbal Noé Aguilar. 2015. In vitro antifungal activity of polyphenols-rich plant extracts against *Phytophthora cinnamomi* Rands. *Africa Journal of Agricultural Research*. 10(50):4554-4560.
64. Mónica L. Chávez-González, Lluvia I. López-López, Raúl Rodríguez-Herrera, Juan C. Contreras-Esquivel, Cristóbal N. Aguilar. 2015. Enzyme-assisted extraction of citrus essential oil. *Chemical Papers*. DOI: 10.1515/chempap-2015-0234.
65. Pedro Aguilar-Zárate, Mario A. Cruz, Julio Montañez, Raúl Rodríguez-Herrera, Jorge E. Wong-Paz, Ruth E. Belmares, Cristóbal N. Aguilar, 2015. Gallic acid production under anaerobic submerged fermentation by two bacilli strains. *Microbial Cell Factories*. 14:209.

66. De la Cruz R, Ascacio JA, Buenrostro J, Sepúlveda L, Rodríguez R, Prado-Barragán A, Contreras JC, Aguilera A, Aguilera CN. 2015. Optimization of ellagitannase production by *Aspergillus niger* GH1 by solid-state fermentation. *Preparative Biochem Biotechnol.* 45(7). Pp 617-631. doi: 10.1080/10826068.2014.940965.
67. Ascacio-Valdes, J.A., Aguilera-Carbo, A.F., Buenrostro, J.J., Prado-Barragan, A., Rodriguez-Herrera, R. Aguilar, C.N. 2015. The complete biodegradation pathway of ellagitannins by *Aspergillus niger* in solid-state fermentation. *J. Basic Microbiol.* 55, 1–8.
68. Silvia Marina Gonzalez-Herrera, Olga Miriam Rutiaga-Quñones, Cristobal Noe Aguilar, Luz Araceli Ochoa-Martínez, Juan Carlos Contreras-Esquivel, Mercedes G. Lopez, Raúl Rodríguez-Herrera. 2016. Dehydrated apple matrix supplemented with Agave fructans, inulin, and oligofructose. *LWT - Food Science and Technology.* 65:1059-1065.
69. Miguel A. De León-Zapata, Lorenzo Pastrana-Castro, María Luisa Rua-Rodríguez, Olga Berenice Alvarez-Pérez, Raul Rodríguez-Herrera & Cristóbal N. Aguilar. 2016. Experimental protocol for the recovery and evaluation of bioactive compounds of tarbush against postharvest fruit fungi. *Food Chemistry* 198: 62–67.
70. Dulce A. Flores-Maltos, Solange I. Mussatto, Juan C. Contreras-Esquivel, Raúl Rodríguez-Herrera, J.A. Teixeira, Cristóbal N. Aguilar 2016. Biotechnological production and applications of fructooligosaccharides *Critical Reviews in Biotechnology.* 36(2):18:1-9. doi:10.3109/07388551.2014.953443.

Books

- Aguilar, C.N., R. Rodríguez H., S. Saucedo P., D. Jasso C. 2008. Phyto-chemicals from Mexican semi-desert: from plant to natural chemicals and biotechnology. Ed. Path Design. Saltillo Coahuila México 579 p (in Spanish).
- Soto-Cruz, O., P.M. Ángel, A. Gallegos-Infante and R. Rodríguez-Herrera. 2008. Advances in Food Science and Food Biotechnology in Developing Countries. AMECA. Saltillo Coahuila México 330 p.
- Rodríguez H., R., O. Soto C., J.L. Martínez, C.N. Aguilar. 2009. Proteomes and Genomes in the XXI century: Environmental Biotechnology. Ed. Valle del Candamo, Monclova Coah. México 356 p. (in Spanish).
- Rodríguez Herrera, R., Aguilar González, C.N., Simpson-Williamson, J.K., Gutiérrez Sánchez, G.. 2011 Phytopathology in the omics era. Signpost Research. Kerala India. 366p. ISBN 978-81-308-0438-5.
- Flores AC., González VM., Aguilar CN., Rodríguez R., 2014. Microbial Biofertilizers. Ed. Plaza y Valdes. 434p. ISBN: 978-607-402-177-1. Mexico (in Spanish).

Chapter 68

**MITOCHONDRIAL POPULATION GENETICS
INFERENCES ABOUT THE PHYLOGEOGRAPHY
AND SYSTEMATICS OF THE TAYRA (*EIRA BARBARA*,
MUSTELIDAE, CARNIVORA)**

***Manuel Ruiz-García^{1,*}, Nicolás Lichilín-Ortíz¹,
Yolanda Mejía¹, Juan Manuel Ortega¹
and Joseph Mark Shostell²***

¹Laboratorio de Genética de Poblaciones-Biología Evolutiva,
Unidad de Genética. Departamento de Biología,
Facultad de Ciencias. Pontificia Universidad Javeriana,
Bogotá DC, Colombia

²Department of Math Science and Technology,
University of Minnesota Crookston, Crookston, MN, US

ABSTRACT

We sequenced the mitochondrial (mt) *ND5* gene of 100 specimens of *Eira barbara* (Mustelidae, Carnivora). The samples represented six out of the seven putative morphological subspecies recognized for this Mustelidae species (*E. b. inserta*, *E. b. sinuensis*, *E. b. poliocephala*, *E. b. peruana*, *E. b. madeirensis*, and *E. b. barbara*) throughout Panama, Colombia, Venezuela, French Guiana, Brazil, Ecuador, Peru, Bolivia, Paraguay, and Argentina. The main results show that the genetic diversity levels for the overall samples and within each one of the aforementioned putative taxa were very high. The phylogenetic analyses showed that the ancestor of the Central and South-American *E. barbara* originated during the Miocene or Pliocene (6.3-4 millions of years ago, MYA). Furthermore, the ancestors of some geographical groups, (we detected at least four) originated during the Pliocene (3.7-2.5 MYA). These four groups (or lineages) were placed in the Cesar-Antioquia Departments (northern Colombia), Bolivia and northwestern

* Corresponding Author's Email: mruizgar@yahoo.es, mruiz@javeriana.edu.co.

Argentina, northern-central Peru, and in the trans-Andean area of Ecuador. However, during the Pleistocene, this species experienced a strong population expansion and many haplotypes expanded their geographical distributions. They became superimposed on the geographical areas of older geographical groups that originally differentiated during the Pliocene. Until new molecular studies are completed, including those with nuclear markers, we proposed the existence of only two subspecies of *E. barbara* (*E. b. inserta* in southern Central America, and *E. b. barbara* for all South America). All of the demographic analyses showed a very strong population expansion for this species in the last 400,000 YA during the Pleistocene.

Keywords: Tayra, *Eira barbara*, mitochondrial *ND5* gene, putative geographical subspecies, genetic diversity, phylogeography, population expansion during the Pleistocene

INTRODUCTION

Tayra (*Eira barbara*) is a Mustelidae (Carnivora, Mammalia) with a long, and slender body. Its length varies from 56 to 71 cm, not including a 37 to 46 cm long bushy tail. Its weight ranges from 2.7 to 7 kg with males larger than females. This species has short, dark brown to black fur that is relatively uniform across the body, limbs, and tail, except for a yellow or orange spot on the chest. The fur on the head and neck is much paler, typically tan or greyish in color. The head has small, rounded, ears, long whiskers and black eyes with a blue-green shine. The feet have toes of unequal length with tips that form a strongly curved line when held together. The claws are short and curved, but strong, being adapted for climbing and running rather than digging.

This species occurs from southern Veracruz (Mexico) throughout Central America and across South America to northern Argentina save for the high Andes and the Caatinga and Cerrado (eastern Brazil; Emmons and Feer 1990). It is one of the most common medium-size predators throughout its range (Emmons and Feer 1990).

E. barbara is a diurnal, sometimes crepuscular species (González-Maya et al., 2015), with a solitary behavior and large home range (Sunquist et al., 1989). Emmons and Feer (1990) showed that the tayra inhabits tropical and subtropical forests, secondary rain forests, gallery forests, gardens, cloud forests, and dry scrub forests. Hall and Dalquest (1963) affirmed that it can live near human disturbed habitats. It frequently occurs in agricultural areas and along the edge of human settlements. Tayra usually inhabits areas below 1,200 m, but there are reports of it being in areas as high as 2,400 m (Eisenberg 1989, Emmons and Feer 1990) and it is common at 2,000 m (Cuarón et al., 2016). Its diet is omnivorous, including fruits, carrion, small vertebrates, insects, honey and small vertebrates such as marsupials, rodents, and iguanids among others (Cabrera and Yepes 1960, Hall and Dalquest 1963, Emmons and Feer 1990). This species is listed as Least Concern (Cuarón et al., 2016).

Cabrera (1957) and Hall (1981) recognized seven morphological subspecies, two in Central America and five in South-America: 1- *E. b. senex* (Thomas in 1900). The type locality is Hacienda Tortugas, Jalapa, Veracruz, Mexico; 2- *E. b. inserta* (Allen in 1908), with the type locality in Ulse, Matagalpa, Nicaragua; 3- *E. b. sinuensis* (Humboldt in 1812), with the type locality for the Sinu River in the Bolivar Department in northern Colombia; 4- *E. b. poliocephala* (Traill in 1821), with type locality Demerara in Guyana; 5- *E. b. madeirensis* (Lonnberg in 1913), with type locality in Humaitá, Madeira River, Brazilian Amazon; 6- *E. b.*

peruana (Nehring in 1886), with type locality in YuracYaku in the San Martin Department in Peru and 7- *E. b. barbara* (Linnaeus in 1758) with the type locality assigned by Lonnberg (1913) to Pernanbuco, Brazil. See Figure 1.

Although it is a relatively common species, only one preliminary study on its molecular population genetics and infra-specific systematics has been published (Ruiz-García et al., 2013).

Therefore, we expanded upon our initial molecular population genetics study with mitochondrial genes of the tayra with the following main aims: 1- To estimate the mitochondrial levels of genetic diversity in the overall tayra population and in some putative morphological subspecies; 2- To determine if there is a correlation between the molecular clades obtained in the phylogenetic analyses with the traditional putative morphological and geographical subspecies of tayras; 3- To estimate the possible temporal splits in the mitochondrial diversification within the evolution of the tayra; and 4- To determine if demographic evolutionary changes have characterized the natural history of the tayra.

MATERIALS AND METHODS

Samples

We sequenced 100 tayras at the mt *ND5* gene. The samples came from 11 countries and represent seven of the eight putative morphological subspecies (Table 1 & Figure 1). They are: 1- Argentina, eight individuals (putative *E. b. barbara*); 2- Bolivia, 16 specimens (putative *E. b. barbara*); 3- Brazil, nine exemplars (four putative *E. b. barbara*; five putative *E. b. madeirensis*); 4- Colombia, 12 individuals (three putative *E. b. madeirensis*; nine putative *E. b. sinuensis*); 5- Ecuador, 27 specimens (four putative *E. b. sinuensis*; 23 putative *E. b. madeirensis*); 6- French Guiana, five exemplars (putative *E. b. poliocephala*); 7- Panama, one individual (putative *E. b. inserta*); 8- Paraguay, four specimens (putative *E. b. barbara*); 9- Peru, 17 exemplars (nine putative *E. b. peruana*; eight putative *E. b. madeirensis*); 10- Trinidad and Tobago, one individual (putative *E. b. poliocephala*). Thus, these samples represent six out of the seven putative morphological subspecies recognized for this species.

The DNA of some of the tayra individuals we analyzed was extracted from hairs obtained from animals found alive in diverse Indian communities throughout Central and South America. Another fraction of the DNA was obtained from skins, bones, and teeth of hunted individuals of *E. barbara*. We requested permission to collect biological materials from these skins, bones, and teeth that were already present in the Indian communities. In the case of the skins, we sampled 1-2 cm². Communities were visited only once. All sample donations were voluntary, and no financial or other incentive was offered for supplying specimens for analysis. For more information about sample permissions, see the Acknowledgment section.

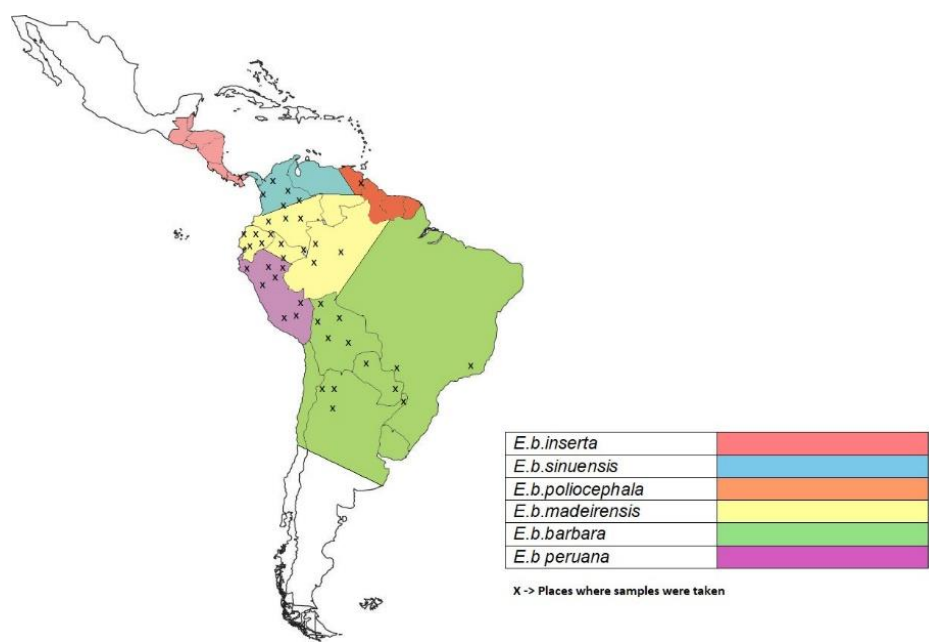


Figure 1. Map with the approximate geographical distributions of the six putative geographical tayra's subspecies (*Eira barbara*) sequenced at the mitochondrial *ND5* gene. X represent localities where samples were obtained.

Molecular Analyses

The DNA from skins and bones was extracted using the phenol-chloroform procedure (Sambrook et al., 1989), whereas DNA samples from hairs and teeth were extracted with 10% Chelex resin (Walsh et al., 1991). Primers and PCR conditions for the *ND5* gene (265 bp) were brought to a volume of 25 µl with 13.5 µl of Mili-Q H₂O, 3 µl of MgCl₂ 1 mM, 1 µl of dNTPs 0.2 mM, 1 µl of each primer (0.1 µM), 2.5 µl of buffer 10X, and one unity of Taq Polymerase with 50-100 ng of DNA. We used the primers L12673 and H12977 (5'-GGTGCAACTCCAAATAAAAGTA -3' and 5'-AGAATTCTATGATGGATCATGT 3'; Waits et al., 1999). The PCR temperatures were 95° for 5 minutes followed by 10 cycles of 1 minute at 95°C, 1 minute at 64°C and 1.5 minute at 72°C, 25 cycles of 1 minute at 95°C, 1 minute at 60°C and 1.5 minute at 72°C and one final extension of 15 minutes at 72°C. All amplifications, including positive and negative controls, were checked in 2% agarose gels. Those samples that amplified were purified using membrane-binding spin columns (Qiagen). The PCR products were sequenced in both directions using the Big Dye™ kit in an ABI 377A automated DNA sequencer. A consensus of the forward and reverse sequences was determined using the Sequencher program.

It is possible that some of the sequences represent numts (mitochondrial DNA fragments inserted into the nuclear genome) rather than true mtDNA (Chung and Steiper, 2008). However, we note that all amino acid translations of the obtained sequences showed the presence of initial start and terminal stop codons and the absence of premature stop codons. Protein translation was also checked to evaluate the possible presence of numts. Nevertheless, the mutations we observed were synonymous changes, thus suggesting that there were no numts in the sequences.

Table 1. Samples of taya (*Eira barbara*), by countries, localities and putative geographic subspecies, sequenced at the mitochondrial ND5 gene for this work

Country	Number of samples studied	Localities	Putative geographic subspecies
Panama	1	1 Chiriqui	<i>E. b. inserta</i>
Colombia	12	2 Agustín Codazzi-Cesar 1 PNN Los Katios-Chocó 1 Zaragoza-Antioquia 1 Yarumal-Antioquia 1 PNN Tama, Norte de Santander 2 El Tuparro-Vichada 1 San Martín-Meta 1 Playa Blanca-Guainia 1 Puerto Arara-Amazonas 1 Leticia-Amazonas	9 <i>E. b. sinuensis</i> 3 <i>E. b. madeirensis</i>
Trinidad & Tobago	1	1 Rio Claro	<i>E. b. poliocephala</i>
French Guiana	5	5 Camopi River	<i>E. b. poliocephala</i>
Ecuador	27	3 Yarinacocha-Pastaza 2 Sucua-Morona Santiago 2 La Perla-Santo Domingo de Tsáchilas 2 Miashi-Zamora 2 Pillaro-Tungurahua 2 Canelos-Pastaza 2 Sarayaku-Pastaza 1 Coca-Napo 1 Misahuallí-Napo 1 La Bonita-Napo 1 Macuma-Morona Santiago 1 Loreto-Napo 1 Hushafindi-Napo 1 Pichincha 1 Miazal-Morona Santiago 1 Yangana-Loja 1 Cononaco-Pastaza 1 Tinigua-Napo 1 Nuevo Rocafuerte-Napo	4 <i>E. b. sinuensis</i> 22 <i>E. b. madeirensis</i>
Peru	17	2 Iquitos-Loreto 1 Caballococha-Loreto 1 Nauta-Loreto 1 Luceropata-Loreto 1 Puerto Venus-Loreto 1 Lamas-San Martin 1 Moyobamba-San Martin 1 Rioja-San Martin 1 Nuevo Cajamarca-San Martin 1 Bagua Grande-Amazonas 1 Oxamarca-Cajamarca 1 Puerto Bermudez-Pasco	8 <i>E. b. madeirensis</i> 9 <i>E. b. peruana</i>

Table 1. (Continued)

Country	Number of samples studied	Localities	Putative geographic subspecies
		1 Bolognesi-Ucayali 1 Seshea-Ucayali 1 Manu-Madre de Dios 1 Marcapata-Cusco	
Bolivia	16	3 Ballivian-Beni 1 Piso Firme-Beni 1 Nicolas Suarez-Pando 1 Sena-Pando 1 Franz Tamayo-La Paz 1 Coripata-La Paz 1 Cajuata-La Paz 1 Totorá-Cochabamba 1 Pojo-Cochabamba 1 Vila vila-Cochabamba 1 Julpe River-Cochabamba 1 El Cerro-Santa Cruz 1 Puerto Pailas-Santa Cruz 1 San Jose de Chuiquitos-Santa Cruz	<i>E. b. barbara</i>
Brasil	9	3 Foz de Iguazu-Parana 3 Novo Airao-Negro River-Amazonas 1 Moora-Negro River-Amazonas 1 Paumari-Yavari River-Amazonas 1 Tres Rios-Rio de Janeiro	4 <i>E. b. barbara</i> 5 <i>E. b. madeirensis</i>
Paraguay	4	2 Los Cedrales 2 Loma Grande	<i>E. b. barbara</i>
Argentina	8	2 Salta 1 Abra Pampa-Jujuy 1 Humahuaca-Jujuy 1 La Cocha-Tucuman 1 Burruyacu-Tucuman 1 El Dorado-Misiones 1 San Javier-Misiones	<i>E. b. barbara</i>

Data Analyses

Genetic Diversity

The statistics used to determine the genetic diversity in the overall tayra sample and within the five South American putative tayra subspecies were as follows: the haplotypic diversity (H_d), the nucleotide diversity (π), the average number of nucleotide differences (K), and the θ statistic by sequence. These statistics were obtained using the DNAsp 5.1 software (Librado and Rozas, 2009).

Phylogenetics Analyses

The sequence alignments were carried out manually as well as with the DNA Alignment program (Fluxus Technology Ltd.). MrModeltest v2.3 software (Nylander, 2004) and Mega 6.05 software (Tamura et al., 2013) were applied to determine the best evolutionary mutation model. The Akaike and Bayesian information criteria (AIC and BIC; Akaike, 1974; Schwarz, 1978) were used to determine the best evolutionary nucleotide model in the overall sequence set of *E. barbara*.

Phylogenetic trees were constructed by using two procedures: Maximum Likelihood (MLT) and Bayesian analysis (BI). The ML trees were obtained using the RAxML v.7.2.6 software (Stamatakis, 2006). To select the best fitting model, 50 independent iterations were run using three data partitions (codon 1, codon 2, and codon 3). Additionally, 50 iterations were run using two data partitions (codons 1+2 combined, and codon 3). For each sequence data set, the GTR + G model (General Time Reversible + gamma distributed rate variation among sites; Tavaré, 1986) was used to search for the ML tree and topologic support was estimated with 500 bootstrap replicates using GTR.

A BI tree was completed with the BEAST v. 1.8.1 program (Drummond et al., 2012). Four independent iterations were run using three data partitions (codon 1, codon 2, and codon 3) with six MCMC chains sampled every 10,000 generations for 30 million generations after a burn-in period of 3 million generations. We checked for convergence using Tracer v1.6 (Rambaut et al., 2013). We plotted the likelihood versus generation and estimated the effective sample size (ESS > 200) of all parameters across the four independent analyses to determine convergence and optimal results. The results from different runs were combined using LogCombiner v1.8.0 and TreeAnnotator v1.8.0 software (Rambaut and Drummond, 2013). A Yule speciation model and a relaxed molecular clock with an uncorrelated log-normal rate of distribution (Drummond et al., 2006) were used. Posterior probability values provide an assessment of the degree of support of each node on the tree. The tree was visualized in FigTree v. 1.4 software (Rambaut, 2012). This BI tree was used to estimate the time to most recent common ancestor (TMRCA) for the different nodes. We used a prior of 24.0 ± 1 MYA (95% confidence interval: 26.24–22.36 MYA) for the split between the ancestors of *Eira* and one Procyonidae, as *Potos flavus*. This prior followed the results of Koepfli et al., (2008).

Following Pennington and Dick (2010), the previous BI temporal estimates belong to one of two different approaches for inferring divergence times. The first approach is based on fossil-calibrated DNA phylogenies. The second approach is named “borrowed molecular clocks” and uses direct nucleotide substitution rates inferred from other taxa. For this second approach, we used a median joining network (MJN) with the help of Network 4.6.10 software from Fluxus Technology Ltd (Bandelt et al., 1999). The ρ statistic (Morral et al., 1994) was estimated and transformed into years of divergence among the haplotypes studied. To determine the temporal splits, it is necessary to estimate a mutation rate at the mt *ND5* gene. We used a nucleotide divergence of 1.22% per each million years (Culver et al., 2000), which yielded one mutation each 309,310 years. This estimate was obtained for Felidae. In this work, we assumed that this mutation rate could be similar in Mustelidae. The networks are more appropriate for intraspecific phylogenies than tree algorithms because they explicitly allow for the co-existence of ancestral and descendant haplotypes, whereas trees treat all sequences as terminal taxa (Posada and Crandall, 2001).

Heterogeneity Analyses

Several procedures were carried out to estimate the genetic heterogeneity among the diverse putative *tayra* subspecies analyzed. To determine the overall genetic heterogeneity in *E. barbara*, we used the statistics G_{ST} , γ_{ST} , N_{ST} and F_{ST} (Nei, 1973; Hudson et al., 1992). Additionally, we relied on the H_{ST} , K_{ST} , K_{ST}^* , Z , Z^* , and S_{nn} tests (Hudson, 2000), and the chi-square test on the haplotypic frequencies with permutation tests using 10,000 replicates to measure genetic heterogeneity. Also, we estimated the genetic heterogeneity by subspecies pairs within *E. barbara*. For this task, we used three procedures: 1- Exact tests with Markov chains, 10,000 dememorizations parameters, 20 batches, and 5,000 iterations per batch; 2- Indirect gene flow estimates (N_m) from the F_{ST} statistic with a n -dimensional island model (Slatkin, 1985; Ruiz-García, 1993, 1994, 1997, 1999; Ruiz-García and Álvarez, 2000); and 3- Kimura 2P genetic distances (Kimura, 1980). These genetic heterogeneity statistics were completed with DNAsp 5.1 (Librado and Rozas, 2009) and Arlequin 3.5.1.2 (Excoffier and Lischer, 2010).

Demographic Changes

We relied on three procedures to detect possible historical population changes in the *tayra*: 1- We used the Strobeck's S statistic (Strobeck, 1987), F_u and $L_i D^*$ and F^* tests (Fu and Li, 1993), the $F_u F_s$ statistic (Fu, 1997), the Tajima D test (Tajima, 1989) and the R_2 statistic (Ramos-Onsins and Rozas, 2002). A 95% confidence interval and probabilities were obtained with 10,000 coalescence permutations. 2- The mismatch distribution (pairwise sequence differences) was obtained following the method of Rogers and Harpending (1992) and Rogers et al., (1996). We used the raggedness rg statistic to determine the similarity between the observed and the theoretical curves. 3- A Bayesian skyline plot (BSP) was obtained by means of the BEAST v. 1.8.1 and Tracer v1.6 software. The Coalescent-Bayesian skyline option in the tree priors was selected with four steps and a piecewise-constant skyline model with 30,000,000 generations (the first 3 million discarded as burn-in), kappa with log Normal [1, 1.25] and Skyline population size with uniform [0, infinite; initial value 80]. In the Tracer v1.6, the marginal densities of temporal splits were analyzed and the Bayesian Skyline reconstruction option was selected for the trees log file. A stepwise (constant) Bayesian skyline variant was selected with the maximum time as the upper 95% high posterior density (HPD) and the trace of the root height as the treeModel.rootHeight. To determine the time range for possible demographic changes for *E. barbara*, we consider that the evolution of this taxon occurred during the last 4 MY.

RESULTS

Genetic Diversity and Phylogenetic Inferences

The BIC showed that the best nucleotide substitution model was T92 + G (7,649.51). In contrast, the AIC detected GTR + G + I (5,881.29) as the best model.

The genetic diversity levels in the overall studied sample of *tayra* were very high. For the 100 individuals analyzed, we found 70 different haplotypes with $H_d = 0.983 \pm 0.006$, $\pi = 0.0422 \pm 0.0048$ and $k = 11.175 \pm 5.117$. The genetic diversity for four out of five South American putative morphological subspecies were very similar, all of them with very high genetic

diversity levels ($H_d = 0.991-0.960$ and $\pi = 0.0562-0.0308$). The genetic diversity of *E. b. poliocephala* was somewhat lower ($H_d = 0.600$ and $\pi = 0.0176$), although the sample size for this putative subspecies was the lowest (Table 2).

The MLT and BI can be seen in Figures 2 and 3. Both phylogenetic trees showed that the first diverging branch represented the animal sampled in northcentral Panama (putatively, *E. b. inserta*) (MLT: low bootstrap 28%; BI: $p = 1$). All of the South American specimens we analyzed were placed in the remaining cluster. However, although putatively animals from five different subspecies were included, very few significant clades were observed, and only partially related to the morphological subspecies. The first diverging cluster in the South American animals was one composed by three animals from northern Colombia (Cesar and Antioquia Departments; 50% and $p = 0.97$, respectively), which corresponded with the putative *E. b. sinuensis*. Nevertheless, a Bolivian exemplar and many other specimens “a priori” classified as *E. b. sinuensis* by their geographical origins that did not belong to this cluster, were present in the BI, within this clade. Henceforth, there was only a partial correspondence between this clade and *E. b. sinuensis*. There were other interesting clades in both phylogenetic trees. 1- One was composed of three individuals from the Pacific area of trans-Andean Ecuador (61% and $p = 1$, respectively), which also partially corresponded with *E. b. sinuensis*; 2- Another cluster was composed of individuals from different areas of Bolivia and mainly by individuals from northwestern Argentina (Salta, Jujuy and Tucuman provinces) in MLT (41%). In the BI, this group was only composed of five individuals from northwestern Argentina ($p = 0.96$). This cluster was partially correlated with *E. b. barbara*. In the BI there was another cluster with several Bolivian and Argentinian specimens ($p = 0.84$). **It was separated from the first Argentinian cluster we aforementioned.** However, as we commented for *E. b. sinuensis*, this relationship was incomplete because other individuals “a priori” classified as *E. b. barbara* were dispersed by other clusters; 3- Another cluster of certain relevance was detected in the MLT and BI. It was composed of individuals of central Peru and one individual from the northern Peruvian Amazon (80% and $p = 0.72$, respectively). This cluster also partially supported the existence of the morphological subspecies *E. b. peruana*; 4- Small clusters of animals from the Ecuadorian and Colombian Amazon were present. One of them contained two animals from the Ecuadorian and Colombian Amazon (62% and $p = 1$, respectively) and other three from the Ecuadorian Amazon (73% and $p = 1$; 89% and $p = 1$; 28% and $p = 0.8$, respectively). These very locally restricted clusters were inside the putative morphological subspecies, *E. b. madeirensis*. Many other individuals of *E. b. madeirensis* were distributed in clusters with other individuals “a priori” considered different morphological subspecies.

Table 2. Genetic diversity in the overall sample of *Eira barbara* and in the five putative South American morphological subspecies at the mt ND5 gene represented by the number of haplotypes (NH), the haplotype diversity (H_d), the nucleotide diversity (π), and the average number of nucleotide differences (K)

<i>Eira barbara</i> taxa	NH	H_d	π	K
Overall Sample	70	0.983 ± 0.006	0.0422 ± 0.0048	11.175 ± 5.117
<i>E. b. sinuensis</i>	12	0.987 ± 0.065	0.0562 ± 0.0311	14.884 ± 8.344
<i>E. b. poliocephala</i>	14	0.600 ± 0.147	0.0176 ± 0.0093	4.667 ± 2.556
<i>E. b. peruana</i>	13	0.989 ± 0.063	0.0517 ± 0.0299	13.703 ± 7.324
<i>E. b. madeirensis</i>	26	0.960 ± 0.009	0.0308 ± 0.0071	8.162 ± 3.233
<i>E. b. barbara</i>	25	0.991 ± 0.012	0.0412 ± 0.0092	10.928 ± 4.111

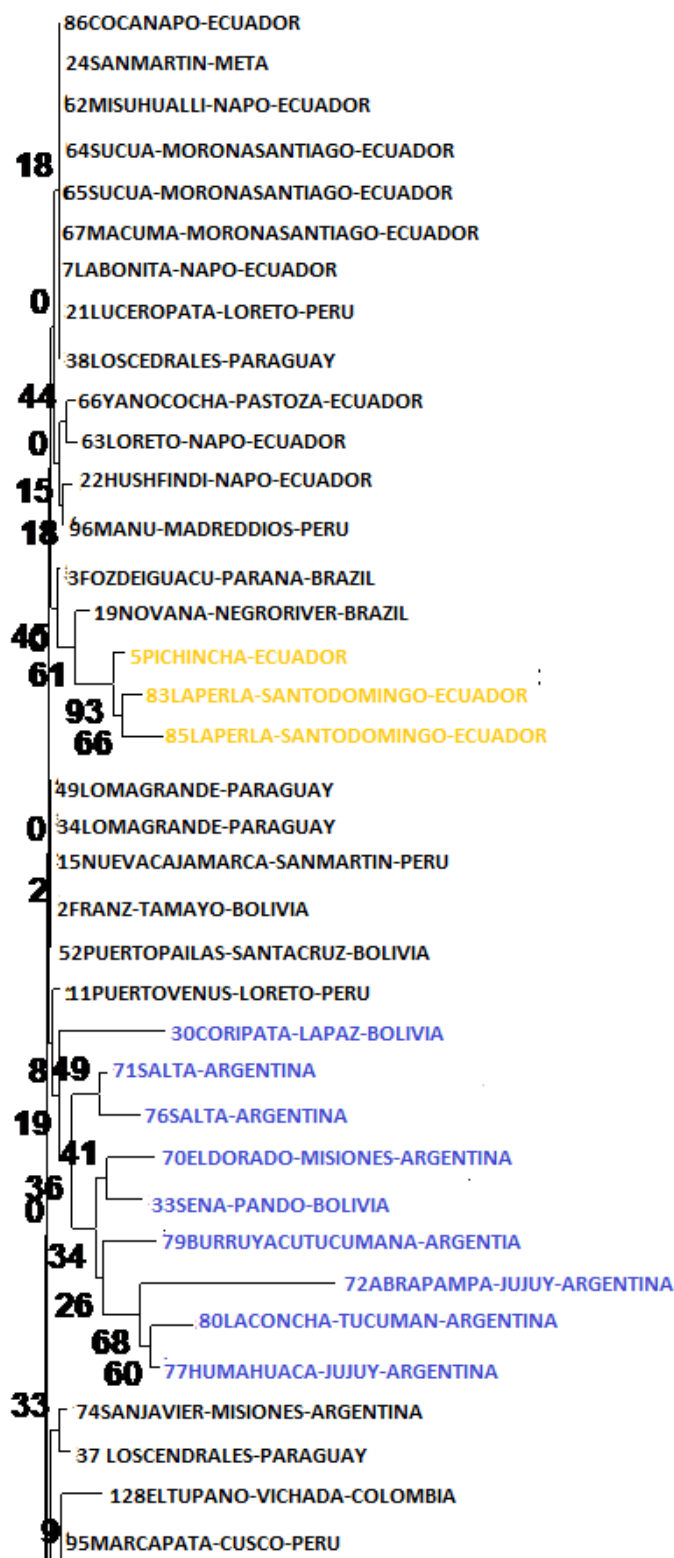


Figure 2. (Continued).

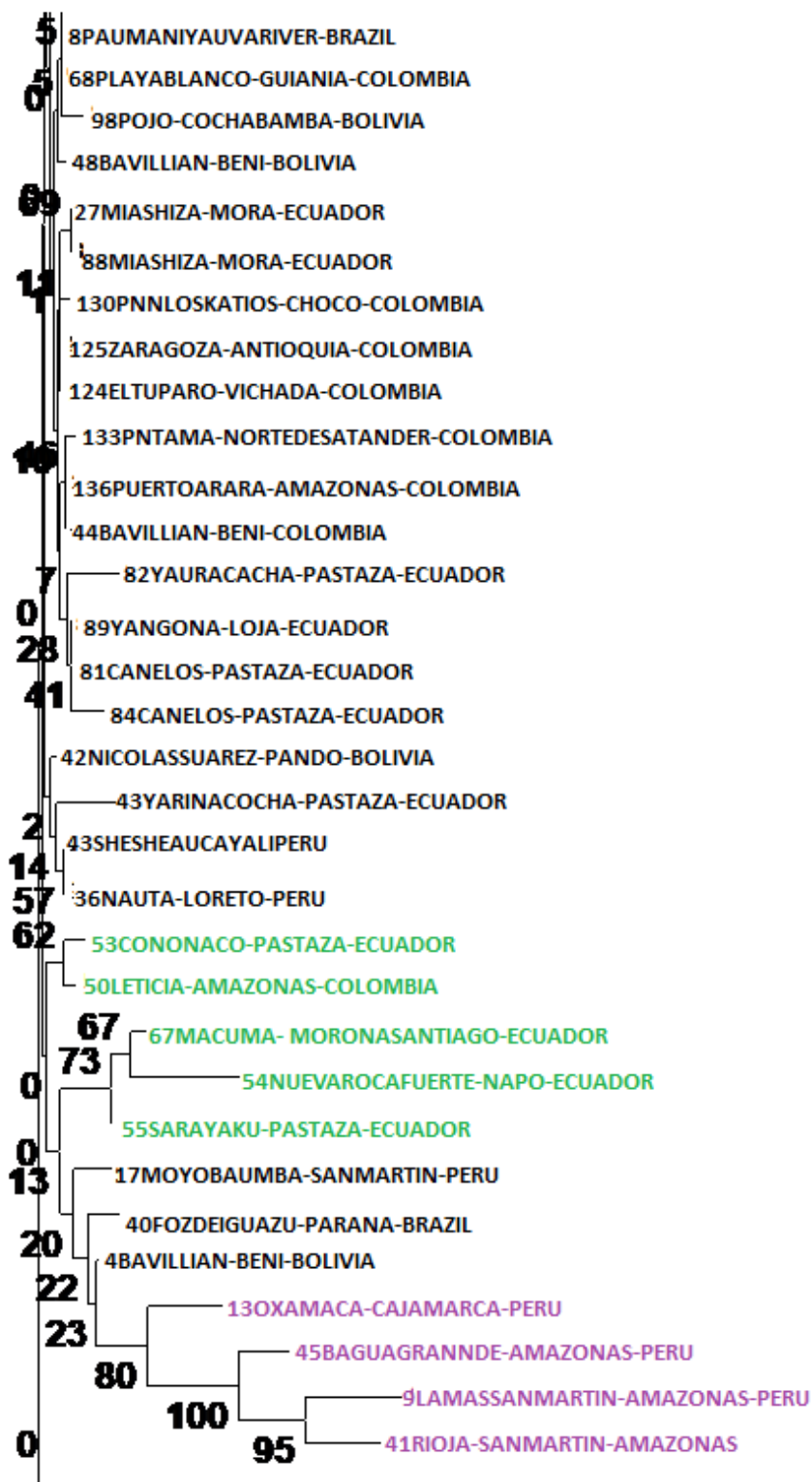


Figure 2. (Continued).

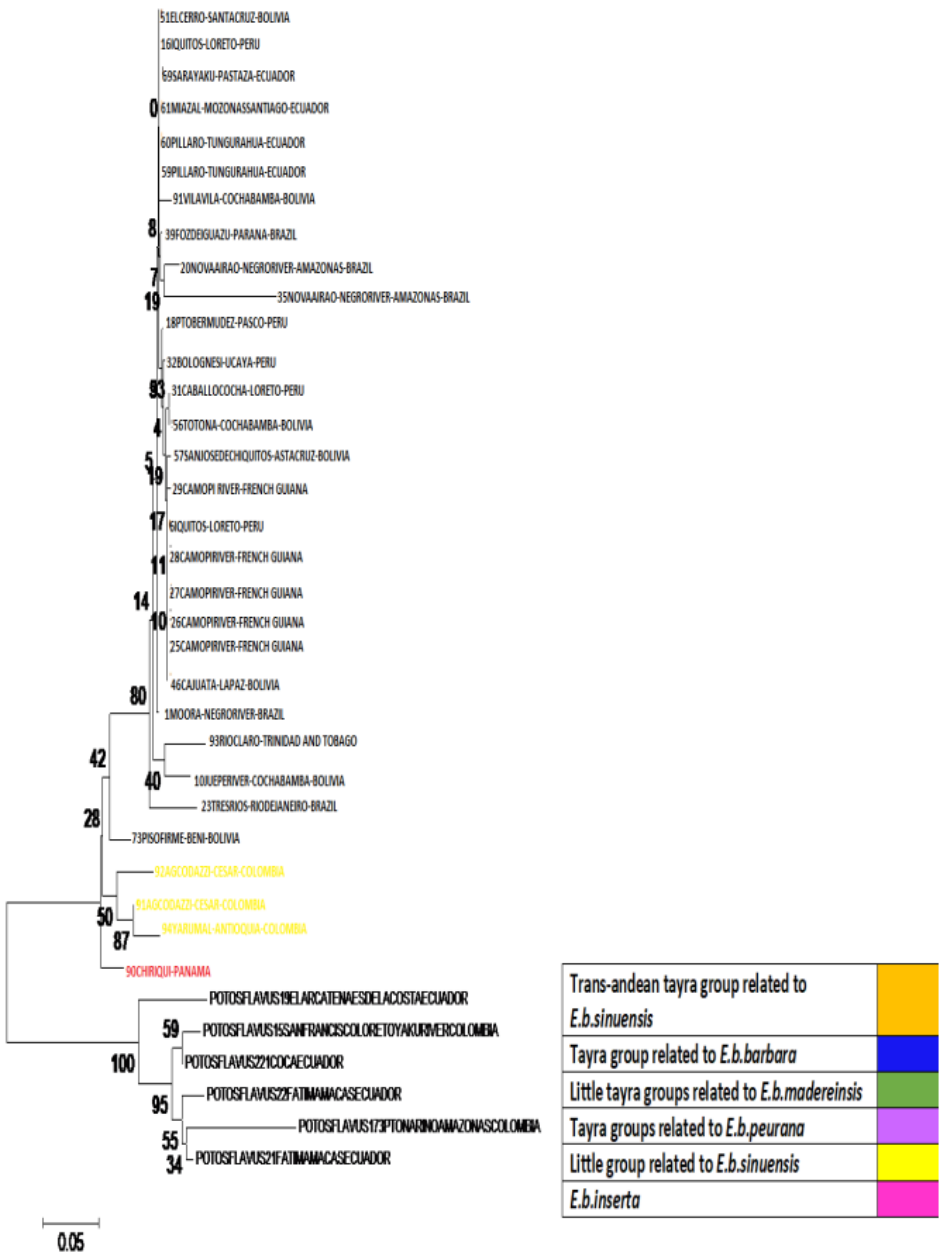


Figure 2. Maximum likelihood tree with the 100 specimens of tayra (*Eira barbara*) sequenced at the mitochondrial *ND5* gene. The number in the nodes are the bootstrap percentages. The procyonidae, *Potos flavus*, was employed as outgroup. In different colors, some relevant clusters which showed a limited correspondence with some putative morphological geographic subspecies of *E. barbara*.

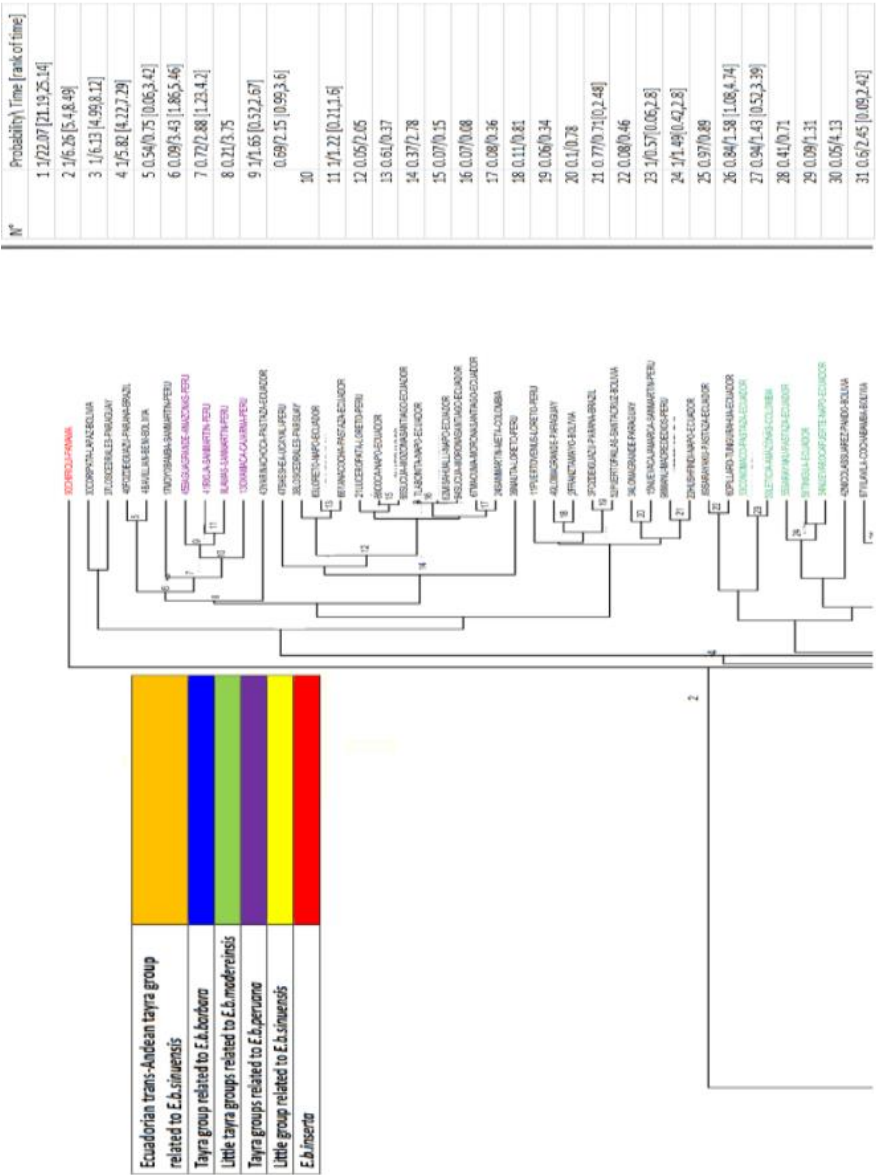


Figure 3. (Continued).

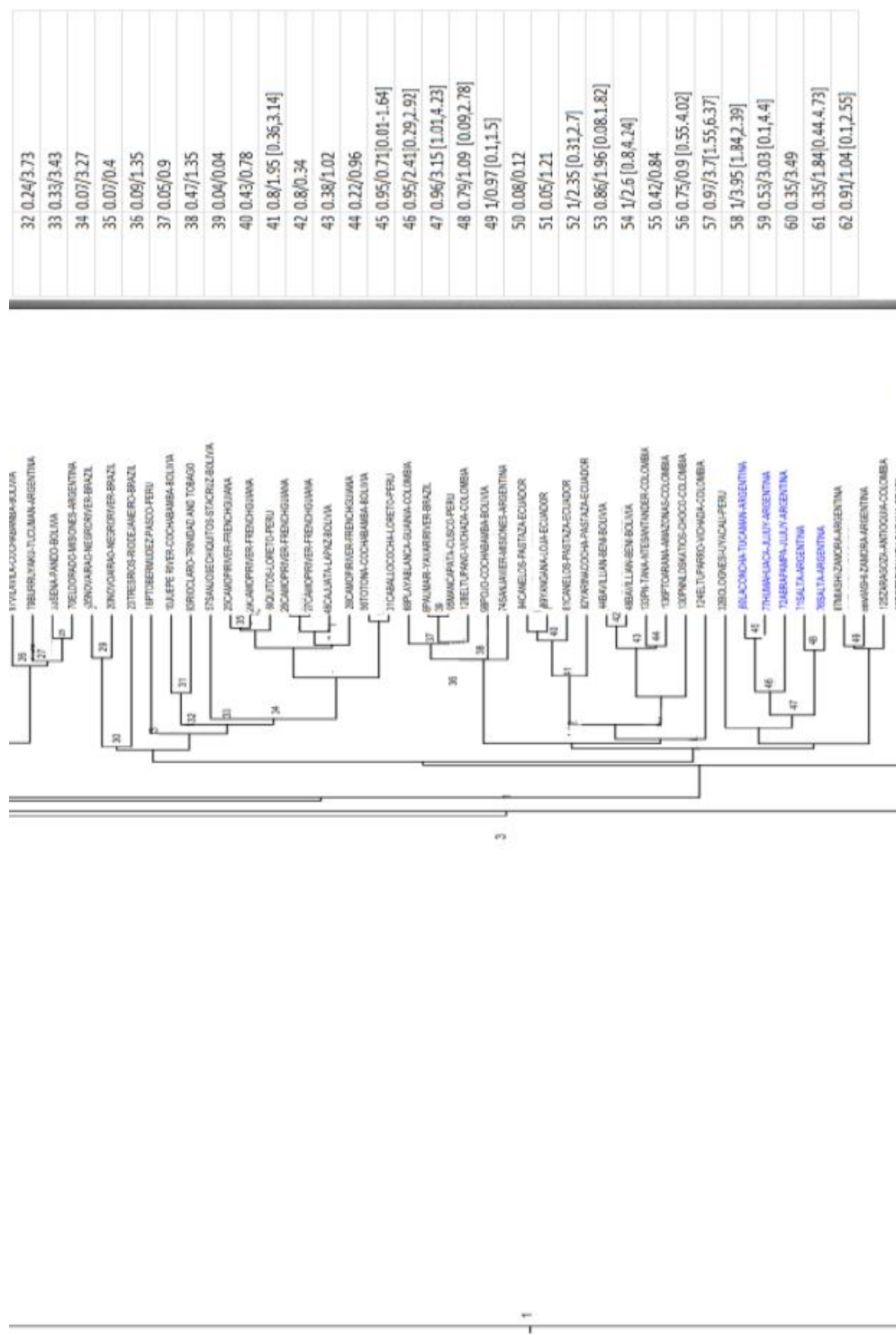


Figure 3. (Continued).

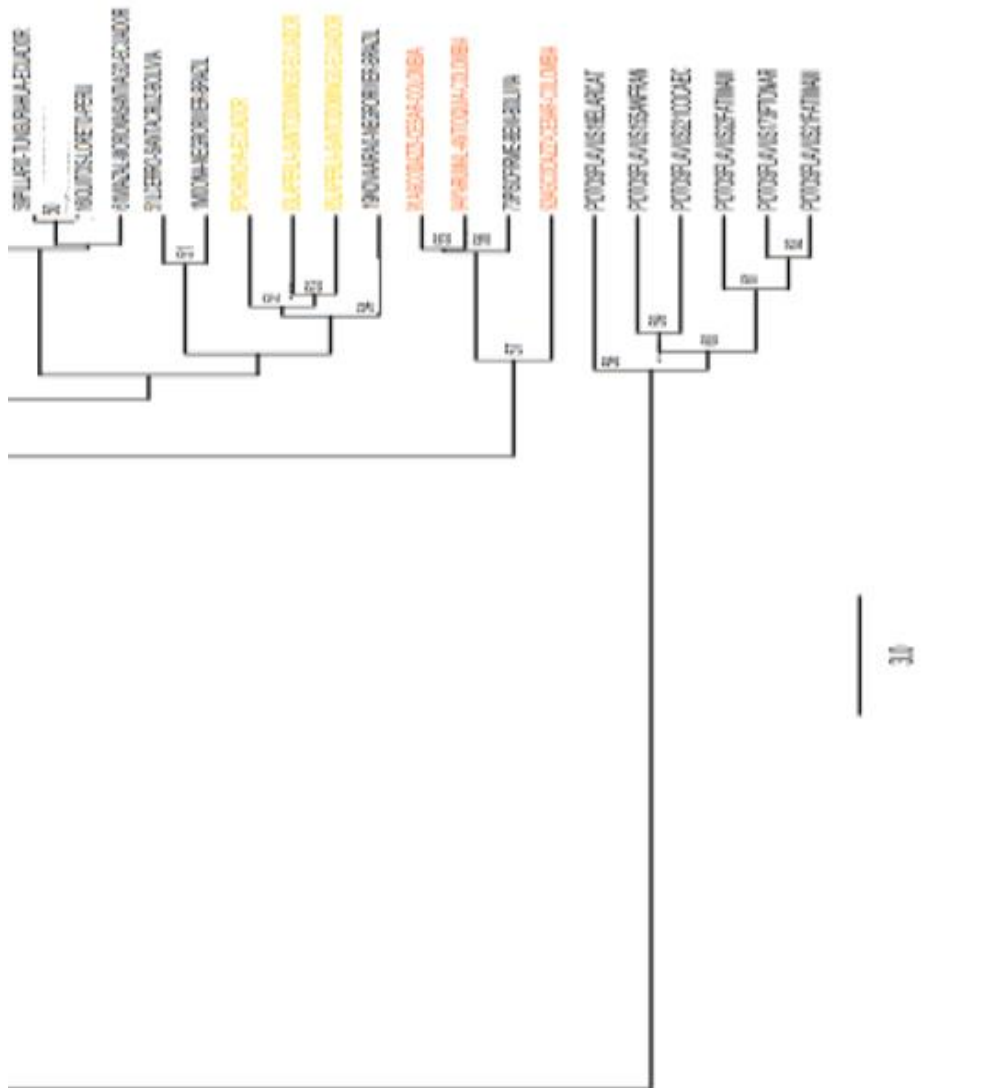


Figure 3. Bayesian tree with the 100 specimens of *tayra* (*Eira barbara*) sequenced at the mitochondrial *ND5* gene. The three numbers in the nodes are the posteriori probabilities, estimated temporal splits in the nodes in millions of years, and the 95% high posterior density of these temporal splits. The *Procyonidae*, *Potos flavus*, was employed as outgroup. In different colors, some relevant clusters which showed a limited correspondence with some putative morphological geographic subspecies of *E. barbara*.

The BI temporal split estimate showed an initial divergence of the ancestors of the Panamanian individual (*E. b. inserta*) and South-American specimens around 6.26 MYA (95%: 5.4-8.49 MYA; Miocene divergence). The ancestor of the clade from northern Colombia split around 3.7 MYA (1.55-6.37 MYA). The ancestor of the animals from northwestern Argentina diverged around 3.15 MYA (1.01-4.23 MYA). In contrast, the ancestors of animals of north-central Peru, trans-Andean Ecuador and one of the Ecuadorian Amazonas clusters diverged 2.88 MYA (1.23-4.2 MYA), 2.35 MYA (0.31-2.7 MYA), and 1.49 MYA (0.42-3.74 MYA), respectively.

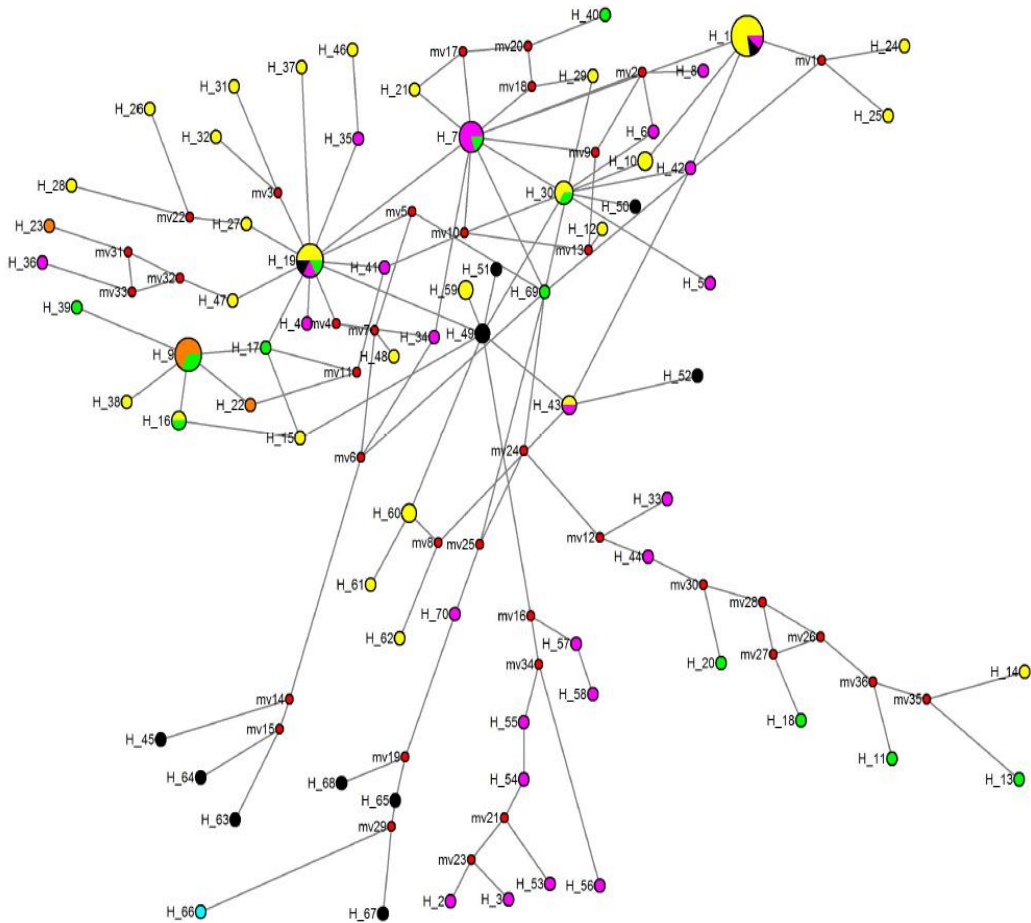


Figure 4. Median Joining Network (MJN) for the haplotypes detected in 100 tayra (*Eira barbara*) sequenced at the mitochondrial *ND5* gene. In light blue, one haplotype of *E. b. inserta*; in pink, haplotypes of *E. b. barbara*; in green, haplotypes of *E. b. peruana*; in yellow, haplotypes of *E. b. madeirensis*; in black, haplotypes of *E. b. sinuensis*; and in brown, haplotypes of *E. b. poliocephala*. Therefore, the five putative geographical subspecies of South American tayras and one Central America subspecies were represented in this analysis. Little red circles are extinct or not found haplotypes.

The MJN analysis revealed a view very similar to the phylogenetic trees (Figure 4). The major fraction of haplotypes were distributed irrespective of the geographical distribution of the morphological subspecies. For instance, the most frequent haplotypes (H1, H30, H7, H49, H19 and H9) included individuals of different putative subspecies: H1 contained exemplars classified “a priori” as *madeirensis*, *sinuensis* and *barbara*; H30 and H7 were composed of *madeirensis* and *peruana* individuals; H49 enclosed individuals of *sinuensis*; H19 consisted of specimens of *sinuensis*, *peruana*, *madeirensis*, and *barbara*, whilst H9 included *poliocephala* and *peruana*. Therefore, some haplotypes were widely distributed, which agrees quite well with extensive gene flow of this species across all of South America. Many of these main haplotypes presented other small haplotypes in star-like form, which is highly related to possible population expansions across the entire South American range of the tayra. Nevertheless, the MJN, as the phylogenetic trees, detected some haplotype clusters to be well delimited geographically. For example, there were the cases of the Central American individual (H66),

the Cesar-Antioquia cluster (H65, H67, and H68), the trans-Andean Ecuadorian cluster (H45, H63, and H64), and the central Peruvian cluster (H11, H13, H18, and H20). The MJN temporal splits were slightly less than that those obtained with BI, but relatively similar. The temporal splits among these haplotypes and the main groups can be seen in Table 3. Some of these time splits are interesting. The divergence between the Panamanian individual and H7 was estimated to occur around 4.02 ± 0.31 MYA. The temporal divergence between clusters from the areas of northwestern Argentina, Cesar-Antioquia area (northern Colombia), trans-Andean Ecuador, and north-central Peru in reference to H7 were 3.73 ± 0.65 MYA, 3.29 ± 0.57 , 1.04 ± 0.31 MYA, and 2.89 ± 0.47 MYA, respectively.

Therefore, the phylogenetics tree and the MJN analyses showed that the ancestor of Central and South American tayras originated during the Miocene or Pliocene. Also, the ancestors of some geographical groups, at least four of certain relevance, originated during the Pliocene and first part of the Pleistocene. However, during the Pleistocene (as we will show later), tayra experienced a strong population expansion and many haplotypes expanded their geographical distributions. They superimposed onto the geographical areas of these older and geographical groups that originally differentiated during the Pliocene.

Genetic Distances and Genetic Heterogeneity among Putative Morphological Subspecies of *Eira barbara*

The Kimura 2P genetic distances among all of the comparison pairs of the six putative morphological subspecies of tayra are shown in Table 4. The differentiation between the Central American subspecies (*E. b. inserta*) and the five South American subspecies was elevated (5.5% - 7.8%), which confirmed that the Central America taxon is, at least, a different subspecies. It is interesting to note that the less differentiated South American taxon with regard to the Central American one was *E. b. sinuensis* (5.5%). It was the South American taxon closest geographically. The genetic distances with the other four South American taxa ranged from 7.1%-7.8%.

In contrast, the genetic distances among the five South American subspecies were very small. They ranged from 0.1% to 1.2%. The pairs of subspecies with the greatest genetic distances were *E. b. poliocephala*-*E. b. sinuensis* (1.2%) and *E. b. poliocephala*-*E. b. barbara* (0.9%).

The overall genetic heterogeneity for all five South American tayra subspecies taken together was significant (Table 5), but the genetic heterogeneity was relatively small. For example, the F_{ST} and the γ_{ST} statistics showed values of 0.095 and 0.109, respectively. Their respective gene flow estimates of 4.75 and 4.10, were relatively high among the putative South-American subspecies.

The analysis of subspecies pair comparisons with exact probability tests (Table 6) only showed two significant pairs: *E. b. madereinsis*-*E. b. barbara* ($p = 0.0066 \pm 0.0017$) and *E. b. madereinsis*-*E. b. poliocephala* ($p = 0.0135 \pm 0.0034$). In this analysis, the Central American taxon was deleted because only one sequence was analyzed. The estimates of N_m by subspecies pair comparisons (Table 7) clearly yielded that the values lower than 1 (which is considered the limit for low gene flow; Wright, 1943) always implied the Central American taxon: *E. b. inserta*-*E. b. peruana* ($N_m = 0.584$), *E. b. inserta*-*E. b. barbara* ($N_m = 0.459$), *E. b. inserta*-*E.*

b. madereinsis ($N_m = 0.286$), *E. b. inserta*-*E. b. poliocephala* ($N_m = 0.137$). Only a South American taxon, *E. b. sinuensis*, ($N_m = 1.202$) had a more substantial gene flow with *E. b. inserta*. This agrees quite well with that determined in the phylogenetic trees and in the genetic distance analysis. The gene flow estimates among the South American taxa were all above 1, ranging from 2.469 to 11.847. These values strongly correlate to elevated historical gene flows among the populations of tayra throughout South America.

Table 3. Temporal splits among different *Eira barbara*'s lineages estimated by means of a Median Joining Network (MJN). Values of temporal splits are in millions of years. SD = Standard deviation

Lineages compared	$\rho \pm SD$	Temporal divergence
Between the Panamanian haplotype (<i>inserta</i>) and H49 (<i>sinuensis</i>)	13.000 ± 1.000	4.021 ± 0.309
Between the Bolivian and northern-western Argentinian haplotypes and H7 (<i>madereinsis</i> , <i>peruana</i>)	12.077 ± 2.097	3.735 ± 0.648
Between the Cesar-Antioquia (northern Colombia) haplotypes and H7 (<i>madereinsis</i> , <i>peruana</i>)	10.625 ± 1.829	3.286 ± 0.565
Between the trans-Andean Ecuadorian haplotypes and H7 (<i>madereinsis</i> , <i>peruana</i>)	3.375 ± 1.008	1.043 ± 0.311
Between the northcentral Peru and H7 (<i>madereinsis</i> , <i>peruana</i>)	9.333 ± 1.523	2.886 ± 0.471
Between H9 (<i>poliocephala</i> , <i>peruana</i>) and H19 (<i>sinuensis</i> , <i>peruana</i> , <i>madeirensis</i> , <i>barbara</i>)	1.000 ± 0.500	0.309 ± 0.154
Between H9 (<i>poliocephala</i> , <i>peruana</i>) and H7 (<i>madereinsis</i> , <i>peruana</i>)	1.363 ± 0.454	0.422 ± 0.140
Between H19 (<i>sinuensis</i> , <i>peruana</i> , <i>madeirensis</i> , <i>barbara</i>) and H7 (<i>madereinsis</i> , <i>peruana</i>)	0.454 ± 0.454	0.141 ± 0.141
Between H49 (<i>sinuensis</i>) and H7 (<i>madereinsis</i> , <i>peruana</i>)	1.429 ± 0.714	0.441 ± 0.220
Between H30 (<i>sinuensis</i>) and H7 (<i>madereinsis</i> , <i>peruana</i>)	0.625 ± 0.625	0.193 ± 0.193
Between H9 (<i>poliocephala</i> , <i>peruana</i>) and H1 (<i>madeirensis</i> , <i>sinuensis</i>)	2.400 ± 0.600	0.742 ± 0.185
Between H19 (<i>sinuensis</i> , <i>peruana</i> , <i>madeirensis</i> , <i>barbara</i>) and H1 (<i>madeirensis</i> , <i>sinuensis</i>)	1.200 ± 0.600	0.371 ± 0.185
Between H49 (<i>sinuensis</i>) and H1 (<i>madeirensis</i> , <i>sinuensis</i>)	2.454 ± 0.818	0.759 ± 0.253
Between H7 (<i>madereinsis</i> , <i>peruana</i>) and H1 (<i>madeirensis</i> , <i>sinuensis</i>)	0.643 ± 0.643	0.198 ± 0.198
Between H30 (<i>sinuensis</i>) and H1 (<i>madeirensis</i> , <i>sinuensis</i>)	1.500 ± 0.790	0.463 ± 0.231

Table 4. Kimura 2P genetic distance (Kimura, 1980) in percentages (%) among six different morphological subspecies of *Eira barbara* (Mustelidae) (below main diagonal) and standard deviations in percentages (%) (above main diagonal) at the mt ND5 gene.

**1 = *E. b. barbara*; 2 = *E. b. peruana*; 3 = *E. b. madeirensis*;
4 = *E. b. poliocephala*; 5 = *E. b. sinuensis*; 6 = *E. b. inserta*;
7 = *Potos flavus* (Procyonidae)**

	1	2	3	4	5	6	7
1		0.1	0.1	0.4	0.1	1.4	3.5
2	0.2		0.1	0.2	0.1	1.5	3.5
3	0.2	0.1		0.4	0.1	1.5	3.7
4	0.9	0.5	0.8		0.5	1.5	3.2
5	0.3	0.5	0.4	1.2		1.2	3.5
6	7.1	7.7	7.4	7.8	5.5		3.5
7	27.5	27.5	28.5	24.6	26.5	24.9	

Table 5. Overall genetic heterogeneity and gene flow (Nm) statistics for the five putative South American subspecies of *Eira barbara* at the mt ND5 gene. * P < 0.05; ** P < 0.01

Estimated Genetic Differentiation	P	Gene flow	
$\chi^2 = 313.351$ df = 272	0.0429*	G _{ST} = 0.0336	Nm = 14.40
H _{ST} = 0.0236	0.0001**	γ_{ST} = 0.0951	Nm = 4.75
K _{ST} = 0.0559	0.0001**	N _{ST} = 0.1108	Nm = 4.01
K _{ST} * = 0.0362	0.0001**	F _{ST} = 0.1086	Nm = 4.10
Z _S = 2279.890	0.0001**		
Z _S * = 7.360	0.0001**		
S _{nn} = 0.511	0.0001**		

Table 6. Exact probability tests (P) (below main diagonal) and standard deviations (above main diagonal) among six different putative morphological subspecies of *Eira barbara* by means of the mt ND5 gene. 1 = *E. b. barbara*; 2 = *E. b. peruana*;

3 = *E. b. madeirensis*; 4 = *E. b. poliocephala*; 5 = *E. b. sinuensis*;

6 = *E. b. inserta*. * = Significant probability

	1	2	3	4	5	6
1		0.0000	0.0017	0.0138	0.0197	-----
2	1.0000		0.0071	0.0132	0.0088	-----
3	0.0066*	0.0679		0.0034	0.0201	-----
4	0.2307	0.3472	0.0135*		0.0036	-----
5	0.4044	0.4796	0.1032	0.0893		-----
6	-----	-----	-----	-----	-----	

Demographic Evolutionary Changes in the Tayra

All of the demographic change statistics indicated population expansion in the tayra (Strobeck's S statistic, $P = 0.0001$; Tajima $D = -2.258$, $p = 0.0040$; Fu & Li $D^* = -3.175$, $p = 0.0115$; Fu & Li $F^* = -3.335$, $P = 0.0052$; Fu's $F_s = -52.632$, $P = 0.00001$; and $R_2 = 0.037$, $P = 0.0041$).

Table 7. Gene flow (Nm) estimates (below main diagonal) among six different putative morphological subspecies of *Eira barbara* by means of the mt *ND5* gene.
1 = *E. b. barbara*; 2 = *E. b. peruana*; 3 = *E. b. madeirensis*;
4 = *E. b. poliocephala*; 5 = *E. b. sinuensis*; 6 = *E. b. inserta*

	1	2	3	4	5	6
1						
2	9.8245					
3	9.2893	11.8471				
4	2.6879	5.3236	2.2686			
5	7.1919	5.8729	4.8435	2.4691		
6	0.4597	0.5843	0.2862	0.1373	1.2019	

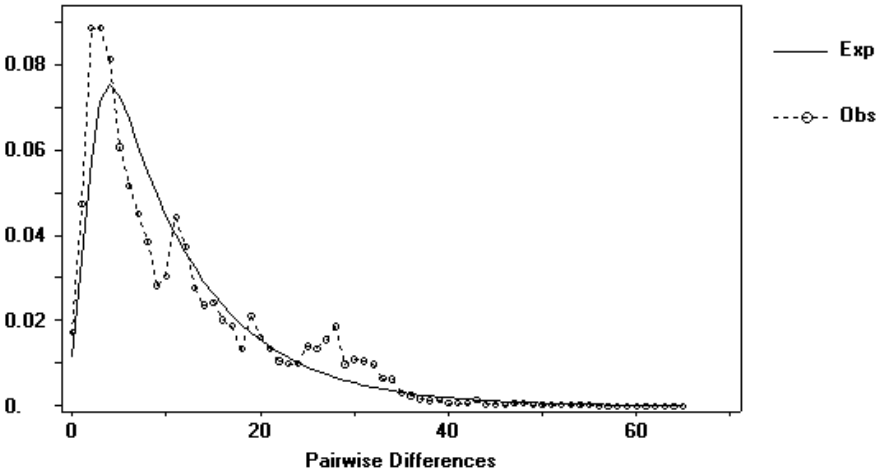


Figure 5. Historical demographic analysis by means of the mismatch distribution procedure (pairwise sequence differences) for the mitochondrial *ND5* gene studied in the overall sample of *Eira barbara*. The analysis showed a clear population expansion of this species during the Pleistocene.

The mismatch distributions also indicated population expansion ($rg = 0.0040$, $P = 0.00280$) (Figure 5). Assuming one year as one generation in the tayra, the population expansion began 343,586 YA, during the Pleistocene.

The BSP analyses also determined a strong female population expansion during the Pleistocene for the tayra (Figure 6). The analyses showed the beginning of the expansion around 400,000 YA, very similar to the temporal estimate previously showed. Therefore, there is incontrovertible evidence that the tayra experienced a strong population expansion during the Pleistocene, as was previously suggested by the MJN analysis.

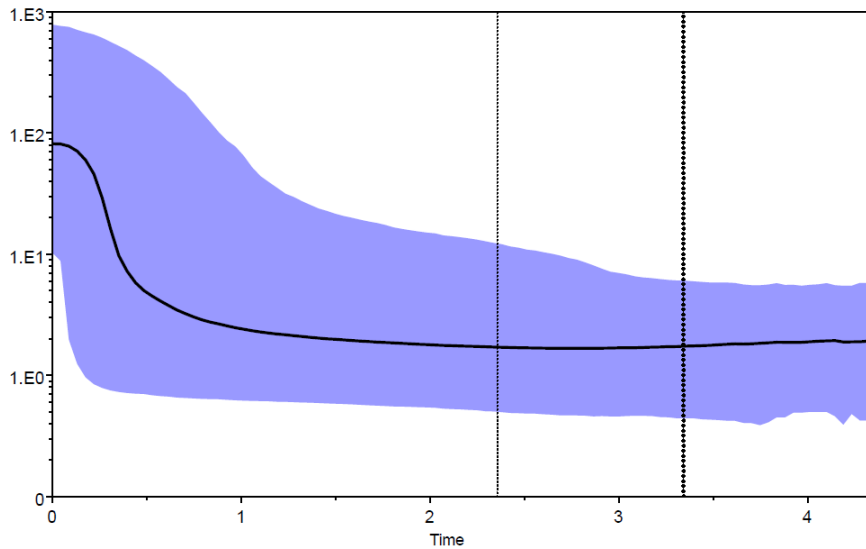


Figure 6. Bayesian skyline plot analysis (BSP) to determine possible demographic changes across the natural history of the overall sample of tayra (*E. barbara*) sequenced at the mitochondrial *ND5* gene. The analysis showed a clear female population expansion during the Pleistocene. On the x-axis, time in millions of years; on the y-axis, log effective population size of females.

DISCUSSION

Genetic Diversity

The levels of nucleotide diversity found in *E. barbara* ($\pi = 0.0422$) were quite high. They were higher than those found in many other neotropical carnivores. For example, they were higher than three fox species (*Lycalopex culpaeus*: $\pi = 0.008$, Ruiz-García et al., 2013a; *Lycalopex sechurae*: $\pi = 0.015$, Ruiz-García et al., 2013a; *Cerdocyon thous*: $\pi = 0.019$, Tchaika et al., 2006), three otter species (*Lontra felina*: $\pi = 0.005$, Vianna et al., 2010; *Lontra longicaudis*: $\pi = 0.011$, Ruiz-García et al., 2017a; *Pteronura brasiliensis*: $\pi = 0.0086$, Ruiz-García et al., 2017a), and two vulnerable Neotropical cats (*Leopardus jacobita*: $\pi = 0.0047$, Ruiz-García et al., 2013b; *Leopardus guigna*: $\pi = 0.00461$, Napolitano et al., 2013). The values of *E. barbara* were similar to those found in certain Neotropical cats, which are characterized by very elevated genetic diversity levels (*Puma yaguaroundi*: $\pi = 0.0661$; Ruiz-García and Pinedo-Castro, 2013 and Ruiz-García et al., 2017b; *Leopardus pajeros*: $\pi = 0.0513$, Ruiz-García et al., 2013b; *Leopardus pardalis*: $\pi = 0.068$, Eizirik et al., 1998; *Leopardus wiedii*: $\pi = 0.035$ - 0.074 , Ruiz-García et al., 2017c).

These high levels of genetic diversity in “a priori” neutral markers, as that we studied, could be related with the fact that many other genes in the genome of a given species contains enough variability for the action of natural selection (Kimura, 1986). This could be the origin of the great morphological variation and behavior plasticity found in the tayra. Emmons and Freer (1990) determined that tayras could live in a wide variety of habitats such as tropical and subtropical forests, primary rain forest (as throughout the Amazonian forest in Brazil, Peru,

Colombia and Ecuador), secondary rain and gallery forests (as in the Llanos of Venezuela and Colombia), gardens, plantations, and cloud forests. They also inhabit dry scrub and deciduous forests (as in the Pantanal in Brazil, Paraguay and Bolivia) and tall grass savannas (as in Argentina, Bolivia, and Paraguay). Sunquist et al., (1989) showed that the extreme plasticity of this species for habitat preferences, activity periods, and diet preferences may reduce interspecific competition between *E. barbara* and other carnivores. This could also be the explanation why Konecny (1989) found no significant habitat preference for this mustelid in Belize. The abundance of the tayra throughout much of Central and South America may be a consequence of its ecological flexibility compared to sympatric carnivores. Associated with this, the tayra is a generalist predator, consuming a variety of fruits, carrion, small and medium vertebrates, insects, and honey (Cabrera and Yepes, 1960; Galef et al., 1976; Hall and Dalquest, 1963; Konecny, 1989; Sunquist et al., 1989).

Genetic Heterogeneity, Gene Flow and the Systematics of the Tayra

Our results clearly showed that the specimen sampled in Central America was highly divergent from all of the individuals sampled in South America. However, the genetic heterogeneity among the putative morphological subspecies of South American tayras, although significant, is very small as we found with the F_{ST} statistic, exact probability tests, and genetic distances. The indirect gene flow estimates were clearly higher than 1. Wright (1943) stated that in an island model, if $Nm > 1$, then gene flow is important enough to erase the genetic heterogeneity among populations. In a stepping-stone model, this amount must be larger than 4 (Trexler, 1988). In both models, $Nm < 0.5$ means that the populations are highly disconnected from a reproductive point of view. For instance, the gene flow estimates between *E. b. inserta* and *E. b. poliocephala* ($Nm = 0.137$), *E. b. inserta* and *E. b. madereinsis* ($Nm = 0.286$) and *E. b. inserta* and *E. b. peruana* ($Nm = 0.459$) showed that the Central American taxon is completely isolated from these three South American taxa. However, recall that certain genetic relatedness was detected between the Central American taxon and the most northern South American taxon (*E. b. sinuensis*). Additionally, we found several gene flow comparison pairs between South American taxa, such as *E. b. madereinsis*-*E. b. barbara* ($Nm = 9.289$) and *E. b. madereinsis*-*E. b. poliocephala* ($Nm = 2.168$). They were elevated, although these comparison pairs showed significant heterogeneity. This might be explained according to Alledorf and Phelps (1981), who argued that the most correct interpretation of $Nm > 1$ is that the populations share the same alleles, although not necessarily with the same allele frequencies. By means of simulations, these authors showed that significant allele divergence occurred in 50% of the generations with a gene flow of $Nm = 50$. Significant allele divergence happened on most occasions when $Nm = 10$.

The tayra seems to have a strong dispersion capacity, which could be related with these high gene flow estimates detected for all the putative South American subspecies. For instance, in the Venezuelan and Colombian Llanos, tayras are usually found along gallery forests. However, tayras cross these extensive grasslands at night, presumably moving from one forest to another covering long distances (Defler, 1980).

Taking into consideration all these facts, we suggest that the six putative morphological subspecies analyzed could be reduced to two different subspecies: *E. b. inserta* for southern Central America and *E. b. barbara* for all South America. The name should be *E. b. barbara*

because it was given by Linneus in 1758 versus *E. b. sinuensis* given by Humboldt in 1812, *E. b. peruana* given by Nehring in 1886, *E. b. madeirensis* given by Lonnberg in 1913 and *E. b. poliocephala* given by Traill in 1821. In reference to the putative northern Central America subspecies (*E. b. senex*), we cannot add any comment on its systematics because no individual of this putative subspecies was analyzed. Therefore, it is essential to sample tayras from this putative subspecies to determine its relationships with other tayra taxa.

Here we suggest another alternative point of view. As we showed, the first tayra's splits originated during the Miocene-Pliocene and beginning of the Pleistocene. However, during the Pleistocene, tayra experienced a strong population expansion and many haplotypes expanded their geographical distributions and they became superimposed on the geographical areas of older geographical groups that originally differentiated during the Pliocene. We suggest that future studies analyze nuclear genes to determine if there was hybridization between the older geographical groups (northern Colombia, part of Bolivia and northwestern Argentina, trans-Andean area of Ecuador, and north-central Peru) and the tayra's population that expanded throughout South America during the Pleistocene. If data support this, then our view of a unique tayra's subspecies in South America should be valid. On the contrary, if there was little or no hybridization between the original groups and Pleistocene colonizers in sympatry, then the number of subspecies in South America could be higher. Therefore, the northern Colombian population (Cesar, Antioquia and possibly nearby areas) should be named *E. b. sinuensis* and the northern central Peruvian population should be named *E. b. peruana*. Also, the Bolivian and especially the northwestern Argentinian population should be defined as a new subspecies (tentatively *E. b. saltensis*). The trans-Andean Ecuadorian population should be defined as a new subspecies (tentatively *E. b. aequatorialis*). The remaining populations of tayra in South America should be named as *E. b. barbara*. Additionally, the range distribution of *E. b. sinuensis* and *E. b. peruana* should be more restricted than traditionally accepted (see Presley, 2000). Even, if some reproductive isolation mechanism had emerged between the Central and the South America tayras due to the old split estimated during the Miocene-Pliocene, both tayra populations should be consider two different species (*E. barbara* and *E. inserta*; this last should be *E. senex* if both Central American forms of tayra were genetically undifferentiated because *senex* was firstly named by Thomas in 1900 and *inserta* was named by Allen in 1908).

Only future nuclear genetic studies can clarify which of the two points of view is more acceptable.

Temporal Splits in the Tayras

Our results showed that the divergence between the Central and the South American tayras occurred around 6.3 to 4 MYA (Miocene or Pliocene periods depending of the temporal estimation). Johnson and O'Brien (1997) and Johnson et al., (2006) showed that seven of the eight primary lineages of felids radiated in the early part of the Late Miocene (10.8-6.2 MYA). There was a noteworthy cooling of the global climate near the end of the Middle Miocene. This period of cooling coincides with formation of a permanent Antarctic ice sheet in the Middle and in the Late Miocene and an Arctic ice sheet in the Pliocene. A large peak of diversification in many vertebrate taxa occurred during the Pliocene epoch. The cold and dry climate during the Pliocene, coincides with the onset of high latitude glacial cycles, causing an explosive expansion of low-biomass vegetation, including grasslands and steppe at mid-latitudes and

development of taiga at high latitudes of Eurasia and North America. These changes were correlated with the diversification of prey species such as muroid rodents and passerine birds that exploited these new habitats, which in turn provided new niches for little or medium carnivores, such as the tayra. Additionally, this Pliocene period agrees quite well with the last phase of rising of the Andes as shown by Dollfus (1974) and Clapperton (1993) (see, for instance, the rising of the “tablazos” of Piura, Peru) and very high volcanic activity in the Andes, with replacement of rainforests with steppe and grassland environments.

Our initial divergence estimates in the tayra agrees relatively well with other molecular studies in the reconstruction of the phylogenetic relationships in the Mustelidae (Bryant et al., 1993; Dragoo and Honeycutt, 1997; Koepfli and Wayne, 1998, 2003; Sato et al., 2003, 2004; Flynn et al., 2005; Fulton and Strobeck, 2006; Koepfli et al., 2008). Two molecular studies are fundamental to understanding the phylogenetics of the Mustelidae (Koepfli and Wayne, 2003; Koepfli et al., 2008). In the first study, the authors used five nuclear gene segments and the mt *Cyt-b* gene. The genes *APOB*, *FES*, *GHR*, *RH01* and *mtCyt-b* clustered *E. barbara* together with *Martes americana*, *Martes pennanti* and *Gulo gulo*. On the other hand, *CHRNA1* clustered *E. barbara* with *Meles meles* and *Arctonyx collaris*. The major part of the trees generated by these authors showed that Mustelinae and Melinae were polyphyletic within the Mustelidae, whereas Lutrinae was monophyletic. The authors of the second study analyzed 22 nuclear and mitochondrial gene segments and determined Mustelidae to consist of seven primary groups. These groups include four major clades and three monotypic lineages. It also included *Eira barbara* clustered into the subfamily Martinae, together with *Martes* and *Gulo*, the most divergent taxa within this subfamily (100% of bootstrap and posterior probability of 1). In that study, the branch of *E. barbara* diverged from the other Mustelidae around 6.7-7.7 MYA (calculated using the mean values), which agrees with our estimate.

These molecular results are not in conflict with the fossil record we know and understand for Mustelidae in America. Many species of Mustelidae appeared in North America during the Late Miocene-Early Pliocene. For instance, *Cernictis hesperus*, from the Pinole Tuff Local Fauna of California, has been dated radiometrically to have lived 5.3-5.5 MYA (Tedford et al., 2004; Baskin, 2011). Other cases of extinct genera appearances are *Trogonictis* and *Sminthosinis* as well as extant genera such as *Lutra* and *Mustela* during the Hemphillian period (4.7-5.9 MYA, Tedford et al., 2004). There is also *Legionarictis fortidens* from the Barstovian (Middle Miocene) marine Temblor Formation in California (Tseng et al., 2009). This form has shown a very close resemblance to other Miocene Mustelidae genera such as *Dehmictis*, *Eirictis*, *Iberictis*, and *Trochictis*, all from the Old World. The form also closely resembles *Sminthosinis*, *Trigonictis* (all from the New World), and, especially, with two extant genera, *Galictis* and *Eira* (Ray et al., 1981; Ginsburg and Morales, 1992; Baskin, 1998; Ginsburg, 1999). In fact, the cladistic analysis of Tseng et al., (2009) determined that this genus could be an evolutionary basal stage closely related to *Eira*. At the Longdan Fauna of the Gansu Province in China, a similar fossil to *Eira* was found by Qiu et al., (2004) from the Late Pliocene (*Eirictis robusta*; 2.58-2.15 MYA). However, the cladistic analysis of Tseng et al., (2009) determined that this mustelid was not very close to *Eira* and it is probably younger than *Eira*. Two possible fossil species of *Eira* have been described from post-Pliocene deposits of Maryland and Virginia under the names *Galera macrodon* and *G. perdicida*. However, the former has been assigned to *Trigonictis* based on additional material collected from deposits of the Blancan land mammal age from Washington, Idaho, Nebraska, Kansas, Texas, North Carolina, and Florida (Ray et al., 1981). *Trigonictis* is considered an intermediate form between *Galictis* and *Eira*

and could be ancestral to both. The second species may be *Mephitis* (Alston, 1882). In addition, extinct species of *Eira* were noted from the Pliocene of the Eastern Hemisphere, but specific names were not given (Scott, 1937). Hershkovitz, (1972) claimed that *Eira* and other endemic monotypic mustelid genera, such as *Lyncodon* and *Pteronura*, may have evolved in South America and moved north as part of the north and south American interchange across the Panamanian land bridge. In contrast, fossil records suggest that *Eira* may have a North American origin (Ray et al., 1981).

Therefore, the process of diversification within *Eira* could be at the end of the first mustelid diversification peak or at the beginning of the second mustelid diversification peak detected by Koepfli et al., (2008). In either case, this mitochondrial diversification process occurred before the Panamanian land bridge (3-2.5 MYA). Therefore, *Eira*'s could have radiated in North America before South America in concordance with the fossil record (Ray et al., 1981) and against the view of Hershkovitz (1972). If so, tayras arrived in South America before the complete formation of the Panamanian land bridge, coinciding with the Choco-Panama island bridge (Galvis, 1980), which could have been used by the ancestors of the current *E. barbara* to colonize northern South America from Central America. During the upper Pliocene orogeny, the present Tuira, Atrato and Sinu river basins and the nearby lowlands were raised above sea level. Thus, the mountains of southern Central America and of the northern Andes were uplifted to about their present elevation (Van der Hammen, 1961). Although the Nicaraguan, Panamanian and Colombian portals remained open (upper Miocene-middle Pliocene), numerous volcanic islands existed from the lower Atrato Valley and the Tuira River Basin of eastern Panama to the Nicaraguan portal. They could have been used by *Eira*'s ancestor to migrate southward. The Cuchillo Bridge of the Uraba region, connecting the Tertiary Western Colombian Andes with the Panamanian islands was probably above sea level during this period. Henceforth, tayras could be another "island hopper" species (Simpson 1950, 1965, 1980).

Our results could also be considered as indirect evidence of a Miocene origin of the Isthmus of Panama (Montes et al., 2012, 2015). Indeed, the Isthmus of Panama formation began earlier and seems to be associated with the Northern Andean uplift, around 24 MYA (Farris et al., 2011).

Therefore, the tayra could have arrived in South America before other Mustelidae. Koepfli et al., (2008) claimed that genera and species of mustelids found in South America today are largely descended from North American immigrants that arrived as part of the GABI following the rise of the Panamanian isthmus, 3.0-2.5 MYA. Several informational points support the statement of Koepfli et al., (2008). 1-For example, there is the clade of New World otters where *Lutra canadensis* is sister to *Lontra felina*, and *Lontra longicaudis*. The latter two species are found in South America and are estimated to have diverged 2.8–3.4 MYA (95% HPD: 1.6–5.2 MYA) overlapping with the formation of the Panamanian land bridge. 2-The long-tailed weasel, *Mustela frenata*, ranges from North America to northern South America. In addition, *Mustela africana* and *Mustela felipei* are endemic to South America. Fossil evidence clearly indicates that *Mustela* colonized South America from the north, apparently well after the Panamanian isthmus was in place. 3- Fossils of the current *Lyncodon patagonicus*, (and a fossil form very related as *Lycodon bosei*) and *Stipanicia* sp (closely related to the extant *Galictis cuja*) have been registered in the Ensenadean and Bonaerense periods of the Argentinian Pleistocene (Forasiepi et al., 2007). However our results could be ratified by other results provided by the same authors (Koepfli et al., 2008). They found that *Pteronura*, *Galictis* and *Eira* could have a Eurasian origin for each genus with posterior diversification in North

America. For example, *Pteronura* may be related to the extinct genus *Satherium* from the Pliocene of North America. Additionally, *Eira* may be related to *Trigonictis* and *Legionarictis*, also North American fossils (Tseng et al., 2009). Fossil evidence suggests mustelids colonized the New World across Beringia during different intervals when the land bridge between Eurasia and North America was open. Multiple genera of mustelids migrated into North America during the Late Miocene (around 11.2–5.3 MYA), prior to the first opening of the Bering Strait 5.4–5.5 MYA, which severed the route across Beringia. Many genera that colonized North America during the Late Miocene or earliest Pliocene became extinct. However, *Eira* could be one of the surviving genera that began to diversify in North America and also in South America if we accept that they arrived in South America before the complete formation of the Panamanian land bridge.

The mitochondrial diversification of the oldest groups of South American tayra occurred 3.7–2.3 MYA. This coincided with the climatic changes that originated from the completion of the Panamanian land bridge (3.1–2.8 MYA; Marshall et al., 1979, 1982; Marshall, 1985, 1988; Webb, 1985, 1997; Coates and Obando, 1996) in the Last Pliocene. Diversification in the tayra occurred close to the Gelasian period (2.5–1.8 MYA), a period characterized by the last stages of a global cooling trend that led to the quaternary ice ages (International Commission on Stratigraphy 2007). Around 2.5 MYA, the Andean forests were transformed into open cold dry savannah ('paramo'), which could have potentially isolated populations of different species. They could have crossed the Northern Andes coming from Central America. Van der Hammen (1992) demonstrated that the mean temperature in the Colombian Andes was 4 °C lower than today. He also stated that the rain level descended below the level reported for today (500–1,000 mm). At 2,500 meters above sea level (masl), the temperature was 10 °C lower than it is today. Tayra's diversification could have been affected by the rapid uplift that resulted in a significant elevation of the Northern Andes. The mountain range's height climaxed around 2.7 MYA when the northern Colombian Andes reached its present day elevation (Gregory-Wodzicki, 2000). This also coincides with the last formation of the Central Andes. All of the Andean chain between Cajamarca and Huancavelica in Peru appeared by volcanism in this period.

Much of the mitochondrial diversification process of the typical Pleistocene colonizer haplotypes occurred around 1.5–0.8 MYA. This divergence could have been initiated by the pre-Pastonian glacial period (1.3–0.8 MYA), which had the highest glacial peak of the first Quaternary glacial period (Günz). This glacial period was extremely dry, and there was a great degree of forest fragmentation. This period was a time for haplotype diversification. It was also a time of separation for many carnivores as it was previously determined for the Pampas cat (Cossíos et al. 2009), and for the foxes of the *Lycalopex* genus (Ruiz-García et al. 2013a). Around 1.3 MYA, the Buenos Aires's fauna transformed into a typical semi-arid Patagonian fauna, represented by the guanaco, *Lestodelphys* and *Lyncodon*. Therefore, the climate was considerably colder and drier than today and could have influenced the mitochondrial fragmentation within the tayra.

The strong population expansion detected around 0.4 MYA for the tayra agrees well with an interglacial period (0.39–0.20 MYA, West 1967) characterized by higher temperatures and humidity and forest expansions (Hoxniense in the British Islands, Yarmouth in North America, Holstein in northern Europe and Mindel-Riss interglacial period in central Europe).

Future analyses with nuclear markers are needed as well as samples from Central America (especially, southern Mexico, Guatemala, Belize and Honduras), the Pacific areas of Colombia

and Ecuador, other areas of Central Peru, and the Guyana shield. These markers and additional samples will help us to determine the exact number of subspecies or ESUs (Moritz, 1994). This information is crucial for the development of effective conservation plans for this species.

ACKNOWLEDGMENTS

Special thanks to Dr. Diana Álvarez, Dr. Francois Catzeflis, Pablo Escobar-Armel, and Lina Arguello for their help in obtaining samples of tayras throughout Panama, Colombia, Trinidad, French Guiana, Peru, Ecuador, Brazil, Bolivia, Paraguay and Argentina. Thanks also go to the Instituto von Humboldt (Colombia), the Ministry of Environment, Consejo Nacional del Ambiente and the Instituto Nacional de Recursos Naturales in Peru, to the Colección Boliviana de Fauna (Dr. Julieta Vargas), to CITES in Bolivia and to the Ministerio del Ambiente in Coca and Santo Domingo de Tsáchilas (Ecuador) for their role in facilitating the obtainment of the collection permits in Colombia, Peru, Bolivia and Ecuador. Also thanks go to the Panamanian and Argentinian Ministries of Environment for their roles in obtaining permissions in these countries. Additional thanks to the many people of the diverse Indian tribes in Peru (Bora, Ocaina, Shipigo-Comibo, Capanahua, Angoteros, Orejón, Cocama, Kishuarana, and Alamas), Colombia (Jaguas, Ticunas, Huitoto, Cocama, Tucano, Nonuya, Yuri, Yucuna, Curripacos, and Desano), and Ecuador (Kichwa, Huaorani, Shuar, and Achuar) for their help in obtaining tayra samples.

REFERENCES

- Akaike, H. (1974). A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, AC, 19, 716–723.
- Allendorf, F. W., Phelps, S. R. (1981). Use of allelic frequencies to describe population structure. *Canadian Journal of Fish and Aquatic Sciences*, 38: 1507-1514.
- Alston, E. R. (1882). *Biologia Centrali-Americana: Mammalia*. [Central American Biology: Mammalia]. R. H. Porter, London, United Kingdom.
- Bandelt, H. J, Forster, P., and Rohl, A. (1999) Median-joining networks for inferring intraspecific phylogenies. *Molecular Biology and Evolution*, 16, 37-48.
- Baskin, J. A. (1998). Mustelidae. In C. M. Janis, K. M. Scott, and L. L. Jacobs (Eds.), *Evolution of Tertiary Mammals of North America Volume 1* (pp. 152-173). Cambridge, England: Cambridge University Press.
- Baskin, J. A. (2011). A new species of *Cernictis* (Mammalia, Carnivora, Mustelidae) from the Late Miocene Bidahochi Formation of Arizona, USA. *Palaeontologia Electronica*, 14, 1-7.
- Bryant, H. N, Russell, F. L. S, A. P., and Fitch, W. D. (1993). Phylogenetic relationships within the extant Mustelidae (Carnivora): appraisal of the cladistic status of the Simpsonian subfamilies. *Zoological Journal of Linnean Society*, 108, 301-334.
- Cabrera, A. (1957). Catálogo de los mamíferos de América del Sur. [Catalog of mammals of South America]. I (Metatheria, Unguiculata, Carnívora). *Revista del Museo Argentino de Ciencias Naturales “Bernardo Rivadavia,” Zoología* 4:1-307.

- Cabrera, A., and Yepes, J. (1960). *Mamíferos sudamericanos. Second Ed.* [South American mammals. Second Ed.]. Buenos Aires, Argentina: Ediar.
- Chung, W. K., Steiper, M. E. (2008). Mitochondrial *COII* introgression into the nuclear genome of Gorilla gorilla. *International Journal of Primatology*, 29, 1341–1353.
- Clapperton, C. (1993). *Quaternary Geology and Geomorphology of South America*. Elsevier, Amsterdam.
- Coates, A. G., and Obando, J. A. (1996). The geological evolution of the Central American isthmus. In J. B. C. Jackson, A. F. Budd, and A. G. Coates (Eds.), *Evolution and Environment in Tropical America* (pp. 21-56). Chicago, Illinois: University of Chicago Press.
- Cossíos, E. D., Lucherini, M., Ruiz-García, M., and Angers B. (2009). Influence of ancient glacial periods on the Andean fauna: the case of the Pampas cat (*Leopardus colocolo*). *BMC Evolutionary Biology*, 9, 68-79.
- Cuarón, A. D., Reid, F., Helgen, K., and González-Maya, J. F. (2016). *Eira barbara*. The IUCN Red List of Threatened Species 2016: e.T41644A45212151. <http://dx.doi.org/10.2305/IUCN.UK.2016-1.RLTS.T41644A45212151.en>. Downloaded on 05 July 2017.
- Culver, M., Johnson, W. E., Pecon-Slattery, J. and O'Brien, S. J. (2000). Genomic ancestry of the American puma (*Puma concolor*). *Journal of Heredity*, 91, 186-197.
- Defler, T. R. (1980). Notes on interactions between tayra (*Eira barbara*) and the white-fronted capuchin (*Cebus albifrons*). *Journal of Mammalogy*, 61, 156.
- Dollfus, O. (1974). La cordillera de los Andes: Présentation de los problemas geomorfológicos. [Presentation of geomorphological problems]. *Bulletin de l'Institut Francois d'Etudes Andines*, 3, 1-36.
- Dragoo, J. W., and Honeycutt, R. L. (1997). Systematics of mustelid-like carnivores. *Journal of Mammalogy*, 78, 426-443.
- Drummond, A. J., Ho, S. Y. W., Phillips, M. J., and Rambaut, A. (2006). Relaxed phylogenetics and dating with confidence. *PLOS Biology*, 4(5), e88.
- Drummond, A. J., Suchard, M. A., Xie, D., and Rambaut, A. (2012). Bayesian phylogenetics with BEAUti and the BEAST 1.7. *Molecular Biology and Evolution*, 29, 1969–1973.
- Eisenberg, J. F. (1989). *Mammals of the Neotropics*. Volume 1. The Northern Neotropics. University of Chicago Press, USA. Pp. 1-449.
- Eizirik, E., Bonatto, S. L., Johnson, W. E., Crawshaw Jr. P. G., Vié, J. C., Brousset, D. M., O'Brien, S. J. and Salzano, F. M. (1998). Phylogeographic patterns and evolution of the mitochondrial DNA control region in two Neotropical cats (Mammalia, Felidae). *Journal of Molecular Evolution*, 47, 613-624.
- Emmons, L. H., and Freer, F. (1990). *Neotropical rainforest mammals: a field guide*. University of Chicago Press, Illinois.
- Excoffier, L., Lischer, H. E. L. (2010). Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows. *Molecular Ecology Resour.*, 10, 564–567.
- Farris, D. W., Jaramillo, C., Bayona, G., et al., (2011). Fracturing of the Panamanian Isthmus during initial collision with South America. *Geology*, 39, 1007–1010.
- Flynn, J. J., Finarelli, J. A., Zehr, S., Hsu, J., and Nedbal, M. A. (2005). Molecular phylogeny of the Carnivora (Mammalia): assessing the impact of increased sampling on resolving enigmatic relationships. *Systematic Biology*, 54, 317-337.

- Forasiepi, A., Martinelli, A., and Blanco, J. (2007). *Bestiario Fósil. Mamíferos del Pleistoceno de la Argentina* [Fossil bestiary. Pleistocene mammals of Argentina] (pp. 1-190) Buenos Aires, Argentina: Editorial Albatros.
- Fu, Y. X. (1997). Statistical tests of neutrality against population growth, hitchhiking and background selection. *Genetics*, 147, 915-925.
- Fu, Y. X., and Li, W. H., 1993. Statistical tests of neutrality of mutations. *Genetics*, 133, 693-709.
- Fulton, T. L., and Strobeck, C. (2006). Molecular phylogeny of the Arctoidea (Carnivora): effect of missing data on supertree and supermatrix analyses of multiple gene data sets. *Molecular Phylogenetics and Evolution*, 41, 165-181.
- Galef, B. G., JR., Mittermeier, R. A., and Bailey, R. C. (1976). Predation by the tayra (*Eira barbara*). *Journal of Mammalogy*, 57, 760-761.
- Galvis, J. (1980). Un arco de islas Terciario en el Occidente colombiano. [An arch of Tertiary islands in the Colombian West]. *Geología Colombiana*, 11, 7-43.
- Ginsburg, L. (1999). Order Carnivora. In *The Miocene Land Mammals of Europe*. Rössner, G. E., and Heissig, K. (Eds.). Verlag Dr. Friedrich Pfeil, München. Pp. 109-148.
- Ginsburg, L., and Morales, J. (2000). Origin and evolution of the Melinae (Mustelidae, Carnivora, Mammalia). *Earth Planet Science*, 330, 221-225.
- González-Maya, J. F., Zárrate-Charry, D., Vela-Vargas, I. M., Jiménez-Alvarado, J. S., and Gómez-Hoyos, D. (2015). Activity patterns of Tayra *Eira barbara* populations from Costa Rica and Colombia: evidence of seasonal effects. *Rev. Biodivers. Neotrop.*, 5, 96-104.
- Gregory-Wodzicki, K. M. (2000). Uplift history of the central and northern Andes: a review. *Geol. Soc. of Am. Bull.*, 112, 1091-1105.
- Hall, E. R. (1981). *The mammals of North America*. Second Ed. New York, New York: John Wiley and Sons.
- Hall, E. R., and Dalquest, W. W. (1963). The mammals of Veracruz. University of Kansas Publication Museum of Natural History, 14, 165-362.
- Hershkovitz, P. (1972). The recent mammals of the Neotropical region: a zoogeographical and ecological review. In A. Keast, F. Erk, and B. Glass (Eds.). *Evolution, mammals, and southern continents* (pp. 311-431). Albany, New York: State University of New York Press.
- Hudson, R. R. (2000). A new statistic for detecting genetic differentiation. *Genetics*, 155, 2011-2014.
- Hudson, R. R., Boss, D. D., and Kaplan, N. L. (1992). A statistical test for detecting population subdivision. *Molecular Biology and Evolution*, 9, 138-151.
- International Commission on Stratigraphy. (2007). *International Stratigraphic Chart*. <http://www.stratigraphy.org/chus.pdf>.
- Johnson, W. E., and O'Brien, S. J. (1997). Phylogenetic reconstruction of the Felidae using 16S rRNA and NADH-5 mitochondrial genes. *Journal of Molecular Evolution*, 44 (Suppl 1), S98-S116.
- Johnson, W. E., Eizirik, E., Pecon-Slatery, J., Murphy, W. J., Antunes, A., Teeling, E., and O'Brien, S. J. (2006). The Late Miocene radiation of the modern Felidae: a genetic assessment. *Science*, 311, 73-77.
- Kimura, M. (1980). A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences. *Journal of Molecular Evolution*, 16, 111-120.

- Kimura, M. (1986). DNA and the neutral theory. *Philosophical Transactions of the Royal Society of London, series B*, 312: 343-354.
- Koepfli, K-P., and Wayne, R. K. (1988). Phylogenetic relationships of otters (Carnivora: Mustelidae) based on mitochondrial cytochrome *b* sequences. *Journal of Zoology* (London), 246, 401-416.
- Koepfli, K-P., and Wayne, R. K. (2003). Type-1 STS markers are more informative than cytochrome *b* in phylogenetic estimation of the Mustelidae (Mammalia: Carnivora). *Systematic Biology*, 52, 571-593.
- Koepfli, K. P., Deere, K. A., Slater, G. J., Begg, C., Begg, K., Grassman, L., Lucherini, M., Veron, G., and Wayne, R. K. (2008). Multigene phylogeny of the Mustelidae: Resolving relationships, tempo and biogeographic history of a mammalian adaptive radiation. *BMC Evolutionary Biology*, 6, 10.
- Konecny, M. J. (1989). Movement patterns and food habits of four sympatric carnivore species in Belize, Central America. In K. Redford and J. Eisenberg (Eds.), *Advances in Neotropical Mammalogy* (pp. 243-264). Gainesville, Florida: Sandhill Crane Press.
- Librado, P., Rozas J. (2009). DnaSP v5: A software for comprehensive analysis of DNA polymorphism data. *Bioinformatics*, 25, 1451-1452 | doi: 10.1093/bioinformatics/btp187.
- Lönnberg, E. (1913). Mammals from Ecuador and related forms. *Arkiv Zoologi*, 16, 1–36.
- Marshall, L. G. (1985). Geochronology and land-mammal biochronology of the transamerican faunal interchange. In F. G. Stehli and S. D. Webb (Eds.), *The Great American Interchange* (pp. 49-85). New York, New York: Plenum Press.
- Marshall, L. G. (1988). Land mammals and the great American interchange. *American Science*, 76, 380–388.
- Marshall, L., Butler, R., Drake, R., Curtis, G., and Tedford, R. (1979). Calibration of the Great American Interchange. *Science*, 204, 272-279.
- Marshall, L. G., Webb, S. D., Sepkoski Jr., J. J., and Raup, D. M. (1982). Mammalian evolution and the great American interchange. *Science*, 215, 1351–1357.
- Montes, C., Bayona, G., Cardona, A., Buchs, D. M., Silva, C. A., Morón, S., Hoyos, N., Ramírez, D. A., Jaramillo, C. A., and Valencia, V. (2012). Arc-continent collision and orocline formation: Closing of the Central American seaway. *J. Geophysical Research*, 117, B04105, doi:10.1029/2011JB008959.
- Montes, C., Cardona, A., Jaramillo, C., Pardo, A., Silva, J. C., Valencia, V., Ayala, V. C., Pérez-Angel, L. C., Rodríguez-Parra, L. A., Ramirez, V., et al., (2015). Middle Miocene closure of the Central American sea way. *Science*, 348, 226–229.
- Moritz, C. (1994). Defining evolutionary significant units for conservation. *Trends in Ecology and Evolution*, 9, 373-375.
- Morral, N., Bertrantpetit, J., and Estivill, X (1994). The origin of the major cystic fibrosis mutation (delta F508) in European populations. *Nature Genetics*, 7, 169-175.
- Napolitano, C., Sanderson, J., Johnson, W. E., O'Brien, S. J., Rus Hoelzel, A., Freer, R., Dunstone, N., Ritland, K., and Poulin E. (2013). Population Genetics of the felid *Leopardus guigna* in Southern South America: Identifying intraspecific units for conservation. In Ruiz-García, M and Shostell, J. M. (Eds.), *Molecular Population Genetics, Phylogenetics, Evolutionary Biology and Conservation of the Neotropical Carnivores*. New York, New York: Nova Science Publishers.
- Nei, M. (1973). Analysis of gene diversity in subdivided populations. *Proc Natl Acad Sci USA*, 70, 3321–3323.

- Nylander, J. A. (2004). *MrModeltest* v2. Evolutionary Biology Center, Uppsala University.
- Pennington, R. T., Dick, C. W. (2010). Diversification of the Amazonian flora and its relation to key geological and environmental events: a molecular perspective. In: Hoorn C, Wesselingh F (eds). *Amazonia, Landscape and Species Evolution: A Look into the Past*. Wiley-Blackwell, Oxford. Pp 373-385.
- Posada, D., Crandall, K. A. (2001). Intraspecific gene genealogies: trees grafting into networks. *Trends in Ecology and Evolution*, 16, 37–45.
- Presley, S. J. (2000). *Eira barbara*. *Mammalian Species*, 636, 1-6.
- Qiu, Z. X., Deng, T. and Wang, B. Y. (2004). Early Pleistocene mammalian fauna from Longdan, Dongxiang, Gansu, China. *Palaeontologia Sinica Series C*, 191, 1–198.
- Rambaut, A. (2012). *FigTree* v1.4. <http://tree.bio.ed.ac.uk/software/figtree/>.
- Rambaut, A., Drummond, A. J. (2013). *TreeAnnotator* v1.8.0. <http://beast.bio.ed.ac.uk/>.
- Rambaut, A., Suchard, M. A., Xie, W., Drummond, A.J. (2013). *Tracer* v1.6. <http://tree.bio.ed.ac.uk/software/tracer/>.
- Ramos-Onsins, S. E., and Rozas, J. (2002). Statistical properties of new neutrality tests against population growth. *Molecular Biology and Evolution*, 19, 2092-2100.
- Ray, C. E., Anderson, E., and Webb, S. D. (1981). The Blancan carnivore *Trigonictis* (Mammalia: Mustelidae) in the eastern United States. *Brimleyana*, 5, 1-36.
- Rogers, A. R., Fraley, A. E., Bamshad, M. J., Watkins, W. S., and Jorde, L. B. (1996). Mitochondrial mismatch analysis is insensitive to the mutational process. *Molecular Biology and Evolution*, 13, 895-902.
- Rogers, A. R., and Harpending, H. C. (1992). Population growth makes waves in the distribution of pairwise genetic differences. *Molecular Biology and Evolution*, 9, 552-569.
- Ruiz-García, M. (1993). Analysis of the evolution and genetic diversity within and between Balearic and Iberian cat populations. *Journal of Heredity*, 84, 173-180.
- Ruiz-García, M. (1994). Genetic profiles from coat genes of natural Balearic cat populations: An Eastern Mediterranean and North-African origin. *Genetics Selection and Evolution*, 26, 39-64.
- Ruiz-García, M. (1997). Genetic relationships among some new cat populations sampled in Europe: a spatial autocorrelation analysis. *Journal of Genetics*, 76, 1-24.
- Ruiz-García, M. (1999). Genetic structure of different cat populations in Europe and South America at a microgeographic level: Importance of the choice of an adequate sampling level in the accuracy of population genetics interpretations. *Genetics and Molecular Biology*, 22, 493-505.
- Ruiz-García, M., Alvarez, D. (2000). Genetic microstructure in two Spanish cat populations I: Genic diversity, gene flow and selection. *Genes and Genetics Systems*, 75, 269-280.
- Ruiz-García, M., and Pinedo-Castro, M. (2013). Population genetics and phylogeographic analyses of the jaguarundi (*Puma yagouaroundi*) by means of three mitochondrial markers: the first molecular population study of this species. In Ruiz-García, M., Shostell, J. (Eds.), *Molecular Population Genetics, Phylogenetics, Evolutionary Biology and Conservation of the Neotropical Carnivores*. New York, New York: Nova Science Publishers. Pp. 245-288.
- Ruiz-García, M., Lichilín, N., Jaramillo, M. F. (2013a). Molecular phylogenetics of two Neotropical Carnivores, *Potos flavus* (Procyonidae) and *Eira barbara* (Mustelidae): No clear existence of putative morphological subspecies. In Ruiz-García, M., Shostell, J. (Eds.), *Molecular Population Genetics, Phylogenetics, Evolutionary Biology and*

- Conservation of the Neotropical Carnivores*. New York, New York: Nova Science Publishers. Pp. 37-84.
- Ruiz-García, M., Rivas-Sánchez, D., and Lichilín-Ortiz, N. (2013b). Phylogenetics relationships among four putative taxa of foxes of the *Pseudoalopex* genus (Canidae, Carnivora) and molecular population genetics of *Ps. culpaeus* and *Ps. sechurae*. In Ruiz-García, M., Shostell, J. (Eds.), *Molecular Population Genetics, Phylogenetics, Evolutionary Biology and Conservation of the Neotropical Carnivores*. New York, New York: Nova Science Publishers. Pp. 97-128.
- Ruiz-García, M., Cossíos, D., Lucherini, M., Yáñez, J., Pinedo, M., and Angers, B. (2013c). Population genetics and spatial structure in two Andean cats (the Pampas cat, *Leopardus pajeros* and the Andean mountain cat, *L. jacobita*) by means of nuclear and mitochondrial markers and some notes on skull biometrics. In Ruiz-García, M., Shostell, J. (Eds.), *Molecular Population Genetics, Phylogenetics, Evolutionary Biology and Conservation of the Neotropical Carnivores*. New York, New York: Nova Science Publishers. Pp. 187-246.
- Ruiz-García, M., Escobar-Armel, P., De Thoisy, B., Martínez-Agüero, M., Pinedo-Castro, M., and Shostell, J. M. (2017a). Biodiversity in the Amazon: origin hypotheses, intrinsic capacity of species colonization, and comparative phylogeography of river otters (*Lontra longicaudis* and *Pteronura brasiliensis*, Mustelidae, Carnivora) and pink river dolphin (*Inia sp.*, Iniidae, Cetacea). *Journal of Mammalian Evolution*, 24, 1-28. DOI 10.1007/s10914-016-9375-4.
- Ruiz-García, M., Pinedo-Castro, M., Shostell, J. M. (2017b). Mitogenomics of the jaguarundi (*Puma yagouaroundi*, Felidae, Carnivora): unclear correlation between morphological subspecies and molecular data. *PLOS ONE*, (in press).
- Ruiz-García, M., Pinedo-Castro, M., Shostell, J. M. (2017c). Mitogenomics of the margay (*Leopardus wiedii*, Felidae, Carnivora): some geographic regional clusters but low correspondence with putative morphological subspecies. *Journal of Heredity*, (in press).
- Sambrock, J., Fritsch, E., Maniatis, T. (1989). *Molecular Cloning: A Laboratory manual*. 2nd edition. V1. Cold Spring Harbor Laboratory Press. New York.
- Sato, J. J., Hosoda, T., Wolsan, M., Tsuchiya, K., Yamamoto, Y., and Suzuki, H. (2003). Phylogenetic relationships and divergence times among mustelids (Mammalia: Carnivora) based on nucleotide sequences of the nuclear interphotoreceptor retinoid binding protein and mitochondrial cytochrome *b* genes. *Zoological Scripta*, 20, 243-264.
- Sato, J. J., Hosoda, T., Wolsan, M., and Suzuki, H. (2004). Molecular phylogeny of arctoids (Mammalia: Carnivora) with emphasis on phylogenetic and taxonomic positions of ferret badgers and skunks. *Zoological Scripta*, 21, 111-118.
- Schwarz, G. E. (1978). Estimating the dimension of a model. *Annals of Statistics*, 6, 461-464.
- Scott, W. B. (1937). *A History of Land Mammals in the Western Hemisphere*. New York, New York: Macmillan.
- Simpson, G. G. (1950). History of the fauna of Latin America. *American Scientist*, 38, 361-389.
- Simpson, G. G. (1965). *The geography of evolution. Collected essays*. Philadelphia, Pennsylvania and New York, New York: Chilton Books.
- Simpson, G. G. (1980). *Splendid Isolation: The Curious History of South American Mammals*. New Haven, Connecticut: Yale University Press.
- Slatkin, M. (1985). Rare alleles as indicators of gene flow. *Evolution*, 39, 53-65.
- Stamatakis, A. (2006). RAxML-VI-HP: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. *Bioinformatics*, 22, 2688- 1243.

- Strobeck, C. (1987). Average number of nucleotide differences in a sample from a single subpopulation: a test for population subdivision. *Genetics*, 117, 149-153.
- Sunquist, M. E., Sunquist, F., and Daneke, D. E. (1989). Ecological separation in a Venezuelan llanos carnivore community. In Redford, K., and Eisenberg, J. (Eds.), *Advances in Neotropical mammalogy* (pp. 197-232). Gainesville, Florida: Sandhill Crane Press.
- Tajima, F. (1989). Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics*, 123, 585-595.
- Tamura K., Stecher G., Peterson D., Filipski A., and Kumar S. (2013). MEGA6: Molecular Evolutionary Genetics Analysis version 6.0. *Molecular Biology and Evolution*, 30, 2725-2729.
- Tavaré, S. (1986). Some Probabilistic and Statistical Problems in the Analysis of DNA Sequences. *Lectures on Mathematics in the Life Sciences*, 17, 57-86.
- Tchaicka, L., Eizirik, E., De Oliveira, T., Cándido, J. F., and Freitas, T. (2006). Phylogeography and population history of the crab-eating fox (*Cerdocyon thous*). *Molecular Ecology*, 16, 819-838.
- Tedford, R. H., Albright, L. B. III., Barnosky, A. D., Ferrusquia-Villafranca, I., Hunt, R. M. Jr, Storer, J. E., Swisher, C. C. III, Voorhies, M. R., Webb, S. D., and Whistler, D. P. (2004). Mammalian biochronology of the Arikarean through Hemphillian interval (Late Oligocene through Early Pliocene epochs. In Woodburne, M. O. (Eds.), *Late Cretaceous and Cenozoic mammals of North America* (pp. 169-231). New York, New York: Columbia University Press.
- Trexler, J. C. (1988). Hierarchical organization of genetic variation in the sailfin molly, *Poecilia latipinna* (Pisces: Poeciliidae). *Evolution*, 42, 1006-1017.
- Tseng, Z. J., Wang, X., and Stewart, J. D. (2009). A new immigrant mustelid (Carnivora, Mammalia) from the middle Miocene Temblor Formation of central California. *PaleoBios*, 29, 13-23.
- Van der Hammen, T. (1961). The Quaternary climatic changes of northern South America. *Annals of the New York Academy of Sciences*, 95, 676-683.
- Van der Hammen, T. (1992). La paleoecología de Suramérica Tropical. Cuarenta años de investigación de la historia del medio ambiente y de la vegetación. [The paleoecology of Tropical South America. Forty years of research into the history of the environment and vegetation]. In *Historia, Ecología y Vegetación* (pp. 1-411). Bogotá, Colombia: Corporación Colombiana para la Amazonía-Araraquara.
- Vianna, J. A., Ayerdi, P., Medina-Vogel, G., Mangel, J. C., Zeballos, H., Apaza, M., and Faugeron, S. (2010). Phylogeography of the marine otter (*Lontra felina*): Historical and contemporary factors determining its distribution. *Journal of Heredity*, 101, 676-689.
- Waits, L., Sullivan, J., O'Brien, S. J., and Ward, R. H. (1999). Rapid radiation events in the family Ursidae indicated by likelihood phylogenetic estimations from multiple fragments of mtDNA. *Molecular Phylogenetics and Evolution*, 13, 82-92.
- Walsh, P. S., Metzger, D. A., and Higuchi, R. (1991). Chelex 100 as a medium for simple extraction of DNA for PCR-based typing from forensic material. *BioTechniques*, 10, 506-513.
- Webb, S. D. (1985). Late Cenozoic mammal dispersals between the Americas. In F. G. Stehli and S. D. Webb (Eds.), *The great American biotic interchange*. (pp. 357-386). New York, New York: Plenum Press.

- Webb, S. D. (1997). The great American faunal interchange. In A. G. Coates (Ed.), *Central America: A Natural and Cultural History* (pp. 97-122). New Haven, Connecticut: Yale University Press.
- West, R. G. (1967). The Quaternary of the British Isles. The geologic systems. In Rankama K (Ed.). *The Quaternary*. Vol 2. Interscience, New York. Pp. 1-87.
- Wright, S. (1943). Isolation by distance. *Genetics*, 28, 114-138.

Chapter 69

THE PRECISION MEDICINE AND PRECISION PUBLIC HEALTH APPROACHES TO CANCER TREATMENT AND PREVENTION: A CROSS-COMPARISON

Stephen M. Modell^{1,*}, Sharon L. R. Kardia² and Toby Citrin¹

¹Department of Health Management and Policy,

²Department of Epidemiology

University of Michigan School of Public Health, Ann Arbor, MI, US

ABSTRACT

Like other forms of diagnostics, genetic testing comes with a retinue of costs and benefits. Significant benefits in terms of morbidity and mortality have accrued to individuals tested for more prevalent genetic conditions like cystic fibrosis and sickle cell disease, including persons seen in the emergency room or identified through public health surveillance. These benefits do not mitigate the drawbacks of genetic testing, false and missed diagnoses and sheer cost among them.

Both medicine and public health have aimed at means of maximizing genetic test benefits in the interventions that they apply. The President's Precision Medicine Initiative (PMI) holds promise in that its results could be used to tailor medical treatments to the individual characteristics of patients, "precision" implying a more accurate and precise regimen overall. The National Cancer Institute (NCI) has already launched the NCI-MATCH precision medicine trial, which assigns targeted treatments based on the genetic abnormalities in a tumor, regardless of cancer type. Other trials, such as the NCI Pediatric MATCH trial, are yet to happen. The efficacy of cancer treatments also intersects public health concerns. The Evaluation of Genomic Applications in Practice and Prevention (EGAPP) Working Group has evaluated the use of *UGT1A1* genotyping to determine the best dose of irinotecan to prevent side effects when treating patients for metastatic colorectal cancer. Analytic validity does not always equate with improved patient

* Corresponding Author's Email: mod@umich.edu (Center for Public Health and Community Genomics, University of Michigan School of Public Health, 4628 SPH Tower, 1415 Washington Hts., Ann Arbor, MI 48109-2029; Tel: (734) 615-3141; Fax: (734) 764-1357).

outcomes, however, thus the public health emphasis on development of a suitable evidence base for precision medical and public health efforts.

The public health approach to precision medicine, or “precision public health”, differs from the medical approach in several important ways: (1) population-based with attention to at-risk populations, as opposed to being strictly individualized; (2) focus on primary and secondary prevention, rather than frank disease (tertiary prevention); and (3) prioritizing interventions that have already demonstrated readiness for large-scale implementation, in contrast to the undertaking of novel clinical trials. Precision public health is exemplified in the Centers for Disease Control and Prevention’s emphasis on the implementation of Tier 1 genetic tests that have passed systematic review for analytic and clinical validity and utility – the use of family history for referral for hereditary breast and ovarian cancer genetic testing (*BRCA1/2* mutations), and hereditary nonpolyposis colorectal cancer cascade screening (Lynch syndrome *MLH1*, *MSH2*, *MSH6* mutations).

This paper will cross-compare the precision medical approach to cancer based on pharmacogenomic regimens using companion diagnostics, and the public health approach to precision management of hereditary cancer for 3 cancer types – lung, breast, and colorectal. It will describe methods of early detection and consider how lives can be saved through precise management – from predictive testing and cancer monitoring of the at-risk population, to tailored chemoprevention that fits the needs of the individual. In the population context, a cascade screening “multiplier effect” exists in that relatives can also be assessed and followed for mutations identified in the proband. Cost-benefit analyses (T4 translational research) of medical and public health approaches will be closely examined and compared. Points of commonality between the two approaches will also be discussed, since primary/secondary and tertiary disease prevention represent a continuum. These analyses point to the value of allocating resources towards the health of at-risk populations. Questions remain if particular forms of genetic testing are to become “universalized”, and if the needs of all at-risk groups, including racial-ethnic, are to be addressed.

Keywords: lung cancer, breast cancer, colorectal cancer, Lynch syndrome, genetic testing, cascade screening, universal screening, cost-effectiveness, precision medicine, pharmacogenomics, public health, race, ethnicity

FROM THE HUMAN GENOME PROJECT TO THE PRECISION MEDICINE INITIATIVE

If the twentieth century is known for its success in mapping the human genome, the twenty-first century is becoming equally well known for scientists’ attempts at making genetic interventions “precise” – appropriately chosen, and delivered to the right person and physical target within the human body. The Precision Medicine Initiative (PMI), launched by the Obama administration in January 2015 with a \$215 million outlay in the President’s 2016 Budget and continuing onward in the current administration, brought medicine closer than ever to the ability to tailor medical regimens to the needs of the individual patient [1]. An article by Francis Collins, Director of the U.S. National Institutes of Health (NIH), and Harold Varmus, former Director of the U.S. National Cancer Institute (NCI), which appeared shortly after the announcement, broke the Initiative into two stages: a near-term focus on cancers, and a longer-term aim to yield new knowledge applicable to the broader range of health and disease [2]. Muin Khoury, Director of the Office of Public Health Genomics at the U.S. Centers for Disease Control and Prevention (CDC-OPHG), and Sandro Galea, Dean of Boston University School

of Public Health, followed-up with the affirmation that the PMI could, in time, be used to develop, evaluate, and deliver health interventions with greater “precision” for both individuals and populations [3].

The distance between a strictly individualized approach to precision medicine and one that is population-oriented, fitting the right intervention to the right population, is transcended by a common denominator shared by medicine and public health – cost-effectiveness. Translational research spans several distinct territories, from basic genome-based discovery as it yields candidate health applications (e.g., new genetic tests and therapeutic interventions) - T1 translational research, to evaluation of real-world outcomes (e.g., morbidity and mortality, cost-effectiveness, and quality-of-life indicators) - T4 translational research [4]. In this piece we take the position that whatever the molecular genetic or behavioral approach used, it must make sense dollar-wise such that the interventions are being mustered in an effective and economical way. Our analysis will show that the criterion of cost-effectiveness naturally shifts the prospect of a national precision medicine effort in the population-oriented direction. Though the two may seem like strange bedfellows, both the pharmacogenomics industry and public health community are in agreement that the PMI must be a sustainable effort, so that the real question becomes, “What direction(s) can the PMI effectively take?”

DEFINITIONS OF PRECISION MEDICINE AND PRECISION PUBLIC HEALTH

The definition of a particular medical intervention illustrates both the basic actions to be taken and its scope. Jameson and Longo define “precision medicine” as “treatments targeted to the needs of individual patients on the basis of genetic, biomarker, phenotypic, or psychosocial characteristics that distinguish a given patient from other patients with similar clinical presentations” [5]. Implicit in this definition is the goal of improving clinical outcomes for individual patients, while avoiding unnecessary side-effects that could be incurred by ignoring patients’ individual characteristics. The authors admit that medicine to date has more or less employed such an approach. Hemophilia requires the administration of an appropriate clotting factor, be it factor VIII or factor IX (these days in recombinant form), to stop the bleeding. Clinicians need to undertake a thorough work-up in order to arrive at a precise diagnosis of the hemophilia, which will enable the appropriate therapy to be administered. Choice of antibiotic is another example. For the antibiotic to take hold, the right type of antibiotic must be given for the particular bacterial infection. The latter example suggests that targeted approaches, aimed at particular persons for specific conditions, could actually have population-level applicability, since antibiotics and vaccinations are given on a widespread basis. They are the very opposite of “orphan drugs”, designed to cater to the needs of very rare cases.

Shen and Hwang point out that despite the commonality with past precedent, a substantive shift in methodology between the old medicine and the new medicine is occurring [6]. The practice of medicine has so far remained largely “empirical”. Physicians typically rely on a combination of patient and occasionally family medical history, physical examination, and laboratory data to secure a diagnosis and choose a drug. Treatments are based on a provider’s experience with similar patients. Drugs administered are most often “blockbuster” drugs designed to accommodate the “typical” patient. If the wrong analgesic, antibiotic, or

antiarrhythmic is given, the patient will be weaned off the current drug, and a new one tested until the right drug and dosage are chosen. The idea behind precision medicine is to rely on new biomarkers and genomic tests to “deliver the right treatment to the right patient at the right time” [6]. It is a personally tailored, as opposed to “one size fits all” approach. The fit is “precise” – one person; one drug.

Classic pharmacogenomic examples readily demonstrate this novel approach. Warfarin (brand name Coumadin) dosing allows the frequently used anticoagulant to be titrated to the needs of the patient susceptible to clotting events associated with atrial fibrillation and deep vein thrombosis [7]. Two genes are known to be involved in warfarin therapeutic outcomes – *CYP2C9*, which codes for an enzyme primarily responsible for warfarin metabolism, and *VKORC1*, which codes for the warfarin drug target. Genotyping can yield information useful to guide a person’s initial warfarin dose and allow the clinician to readily stabilize his or her prothrombin measures, a process which usually takes several weeks. Cost-effectiveness studies of genotype-guided dosing have concluded that considerable potential exists for cost savings, but that it cannot be realized until test costs decrease and the uncertainty concerning effectiveness is reduced. The U.S. Centers for Medicare and Medicaid Services (CMS) has consequently adopted a provisional “coverage with evidence development” approach [7]. A physiologic measure, the ratio of the patient’s prothrombin time to a control or “normal” sample, continues to be the professional standard for warfarin dosing. Since the patient is followed with this measure, its use may be considered a tool in “personalized” or individually-tailored medicine.

Imatinib (Gleevec) for chronic myelogenous leukemia (CML) and gastrointestinal stromal tumors is another prime example of the tailoring in drug regimens that may take place. Medical researchers developed this drug over a multi-decade period to inhibit the function of a translocation-related “fusion” gene, *BCR-Abl*, which produces an abnormal tyrosine kinase that is not properly regulated [8]. In initial trials, all of 31 research participants experienced complete remission. The five year survival rate for CML has increased from 31% (1993) to 59% (2003 – 2009) [9]. The drug is administered to patients who are Ph+ (Philadelphia chromosome positive), with effectiveness monitored by white blood cell and platelet counts. Like warfarin, Gleevec targets a particular type of patient and has a very specific chemical target – the ATP-binding site on a particular kinase – and is a drug that can be closely monitored. Gleevec’s cost is about \$3,500 per month, which may evade some patients’ pocketbooks. However, it is covered by Medicare Part D and Medicare Advantage Plans. Both of the above examples describe “personalized medicine” – separating patients into different groups then individually tailoring the treatment to the patient.

Though many authors use the two terms interchangeably, Khoury distinguishes “personalized medicine” from “precision medicine” in that the latter inculcates multiple determinants of health, genetics being one rung, thus can absorb the notion of social determinants as easily as it can molecular determinants. The President’s PMI plan is data intensive in a way that would allow the recording of multiple determinants for large numbers of people, ostensibly through a planned million-person cohort:

Participants will be involved in the design of the Initiative and will have the opportunity to contribute diverse sources of data – including medical records; profiles of the patient’s genes, metabolites (chemical makeup), and microorganisms in and on the body; environmental and lifestyle data; patient generated information; and personal device and sensor data. [1]

Collins and Varmus write that the numerous clinical trials stemming from the PMI and its large-scale cohort will require the building of a “cancer knowledge network” to store all the resulting molecular and medical data in digital form and make it readily deliverable to providers and patients [2]. Simonds and Khoury illustrate this idea with the example of a cancer clinical trials and effectiveness research infrastructure developed by the H. Lee Moffitt Cancer Center. The system is part of its personalized cancer care initiative started in 2006 – Total Cancer Care. The program: (1) integrates data from multiple sources (electronic medical records, biospecimen databases, and molecular data); (2) makes resultant information available to patients by providing active feedback about their health and upcoming appointments and expanded electronic health record; and (3) affords data interfaces for researchers and clinicians [10].

It is to be hoped that the streamlining of cancer clinical trials and centralization of incoming data will reduce costs and increase efficiency over standard medical practices. According to Cancer Research U.K., between 2003 and 2007 cancer trials were accompanied by a 75% increase in administrative costs, a figure in need of remedy [11]. Cost-effectiveness and clinical utility will enter into assessments of the PMI. Public health efforts in the U.S. have to date assumed a twin duty – assessment of the cost-effectiveness of clinical programs, at the same time that a vision of precision medicine’s meaning in terms of population health is being formulated. This vision departs from the medical model of precision health in a number of important respects: (1) it is population-based, with attention to at-risk populations, as opposed to being strictly individualized; (2) the focus is on primary and secondary prevention, rather than frank disease (tertiary prevention); and (3) the emphasis is on interventions that have already demonstrated readiness for large-scale implementation, in contrast to the undertaking of novel clinical trials [12]. To glimpse the future, it is helpful to visit recent experience with precision medicine in the cancer arena. This inspection will take our trek from the individually-focused domain of clinical medicine to the population-oriented territory of public health. The next section will focus on three major cancer categories of mutual interest to clinical medicine and public health – lung, breast, and colorectal cancer – from the medical pharmacogenomics point of view.

PHARMACOGENOMIC INTERVENTIONS AND COST-EFFECTIVENESS

The Precision Medicine Approach to Lung Cancer

Lung cancer is the second most common cancer in both men and women, and is the leading cause of cancer-related death in both genders [13, 14]. Of note, the lung cancer incidence rate for black women is roughly equal to that of white women, despite the fact that they smoke fewer cigarettes [13]. The two major forms of lung cancer are non-small cell lung cancer (NSCLC), and small cell lung cancer (SCLC). NSCLC comprises ~85% of all lung cancers; SCLC ~10-15% [15]. The 5-year survival rate for NSCLC is 21%, suggesting progress that has been and that is yet to be made [16].

Targeted therapies aimed at cancers harboring very specific genetic alterations are becoming more and more common in oncogenomics. A 2012 review of the role of pharmacogenomics in moving genetic association studies from bench to bedside describes the

use of EGFR tyrosine kinase inhibitors (TKI) in the treatment of lung cancer and *HER2* (tyrosine kinase ERBB2)-directed therapies in the treatment of *HER2* (human epidermal growth factor 2)-positive early-stage breast cancer as prime examples of success in the area of cancer pharmacogenomics [17]. A 2016 review of precision medicine approaches in oncology cites ALK (anaplastic lymphoma kinase) fusion oncogene and EGFR (epidermal growth factor receptor (EGFR)) mutations as the main molecular predictive biomarkers supporting NSCLC treatment [15]. Molecular testing for ALK fusion genes has proven valuable. Abbott Molecular already offers a multiplexed assay, the Vysis ALK Break Apart FISH Probe Kit, endorsed by the 2016 National Comprehensive Cancer Network (NCCN) NSCLC practice guidelines [15]. Though ALK fusion gene rearrangements are relatively rare (< 5% of NSCLC cases), clinical responses to targeted inhibitors (e.g., crizotinib) can be quite dramatic [5]. In favor of the precision medicine approach, excluding patients without these mutations can also minimize the exposure of patients to costly and potentially toxic therapies unlikely to benefit them.

EML4-ALK is the specific biomarker used in determining patient efficacy for the choice of a TKI agent such as crizotinib in the tertiary or after-the-cancer-has-arisen management of NSCLC [5]. A 2014 cost-effectiveness study on the use of *EML4-ALK* fusion oncogene testing in first-line crizotinib treatment for patients with advanced NSCLC reveals the complexity inherent in this precision medicine approach [18, 19]. The investigative team found that *EML4-ALK* testing to govern therapeutic decisions improved patient outcomes by an average of 0.011 quality-adjusted life-years (QALYs) while adding extra costs of \$2,725 per patient, of which only \$60 was attributable to the molecular assay itself. The overall interpretation of the cost-benefit calculus changes dramatically, however, when the cost of the companion drug (crizotinib) is considered. The incremental cost-effectiveness ratio is defined as the difference in cost between two alternative interventions, divided by the difference in their effect or impact. The incremental cost effectiveness ratio for administration of the drug itself was \$250,632 per QALY gained. The investigators concluded that the regimen is “likely not considered cost-effective in the current setting” [19]. This assessment was unaltered when the model was subjected to a sensitivity analysis of alternative costs for the molecular testing. States one reviewer, “Where companion diagnostic precision medicine is considered, these assays are by nature tightly coupled to the cost of the specific associated drug” [18].

Assays for EGFR inhibition may be used when other TKI inhibitors, such as gefitinib and osimertinib, are being considered for NSCLC therapy [15]. The U.S. Food and Drug Administration (FDA) has approved two multiplex assay kits, also NCCN endorsed, for this purpose – the *therascreen* EGFR RGQ PCR Kit (for use with gefitinib), and the cobas EGFR Mutation Test (for use with osimertinib). A cost-effectiveness analysis performed by an Australian team compared the use of combined multiplex testing and targeted therapy with NSCLC chemotherapy without testing, and thirdly with best supportive care without testing [20]. The combined strategy resulted in an additional 0.009 life-years (LYs) gained, compared to 1.458 LYs gained in the case of each of the other two strategies. The combined strategy resulted in an incremental cost-effectiveness ratio of \$485,199 (Australian)/QALY comparing combined and best supportive care strategies, and \$489,338 (Australian)/QALY comparing combined and chemotherapy only strategies. Decreasing test and test interpretation costs by half reduced the ratios, but they still remained greater than \$200,000 (Australian)/QALY. The authors concluded that multiplex testing and targeted therapy is not cost-effective as a fourth-line treatment in metastatic lung cancer when first-line treatments such as chemotherapy without pharmacogenomics testing can be employed.

Other predictive biomarkers for NSCLC treatment are on the horizon. For example, *KRAS* mutations can suggest lack of therapeutic efficacy to EGFR targeted therapies. The FDA has cleared *KRAS* mutation detection assays for use in colorectal cancer, but such assays have not yet been approved for use in NSCLC [15].

The Precision Medicine Approach to Breast Cancer

Breast cancer is the most common cancer among American women. It is the second leading cause of cancer death in women, exceeded only by lung cancer [21]. About 85% of breast cancer cases occur in women with no family history of breast cancer. These cases are due to somatic cell mutations which occur as a result of aging, various exposures (such as pre- and postmenopausal hormone therapy), and other life events.

Precision medicine efforts for breast cancer due to somatic mutations fall into at least four categories, two of them – endocrine therapy and *HER2* therapy – being mainstay treatment categories. *HER2*-directed therapies, for which trastuzumab (Herceptin) is often used, are one of the two major pharmacogenomics successes in the cancer area cited by Ritchie [17]. Tamoxifen is a major drug of choice in the category of endocrine therapies, itself being a selective estrogen receptor modulator [15].

Herceptin has proven utility in reducing risk for cancer recurrence after surgery for early-stage *HER2*-positive (*HER2*+) breast cancer, and improving survival in late-stage (metastatic) *HER2*+ breast cancer, but it also poses serious side-effects such as heart damage. Herceptin therapy can also cost a sizable amount, up to \$50,000 per year [22]. Overexpression of the *HER2* gene (*HER2*+ status) is associated with rapid tumor growth and negative diagnostic and prognostic indicators. A systematic review and meta-analysis performed in Canada compared seven different strategies employing immunohistochemistry (IHC) and fluorescence in situ hybridization (FISH) to determine *HER2* status, thus appropriateness of using Herceptin [22]. Each strategy utilized an IHC score (0, 1+, 2+, 3+) alone or coupled with FISH confirmation to form this inference. The incremental cost-effectiveness ratio was lowest when cases with an IHC score of 2+ or 3+ (as opposed to 0 or 1+) were confirmed by FISH, which yielded a ratio of \$3,351 (Canadian) (minimum) to \$12,230 (Canadian) (maximum) per accurately determined case. The cost-effectiveness analysis is favorable given that accurate assessment of *HER2* status is capable of reducing the cost of Herceptin therapy by \$0.6 million per year, and saving \$12 million per year in women who are *HER2*-, thus can be kept off Herceptin [22].

Estrogen-focused therapies have been a part of standard care for more than thirty years, and have displayed an evolution in policy analysts' thoughts about the gold standard for precision management [15]. The Evaluation of Genomic Applications in Practice and Prevention (EGAPP) Working Group was launched by the CDC Office of Public Health Genomics in 2004 to conduct systematic, evidence-based reviews of burgeoning genetic tests and other applications of genetic technologies in transition from research to clinical and public health practice. EGAPP has produced two systematic reviews of the use of gene expression profiling to improve therapeutic outcomes in women with breast cancer. EGAPP's 2009 review examined the validity and utility of three tests – Oncotype DX, MammaPrint, and the H:I (normalized gene expression) ratio [23]. These tests have been designed to go beyond the standard estrogen/progesterone receptor status indicator to predict tumor recurrence risk for women on tamoxifen, for whom alternative therapies might be considered. The EGAPP

Working Group found adequate evidence from the NSABP B-14 randomized controlled trial to support the association between Oncotype DX recurrence scores and 10-year distant metastases in estrogen receptor positive (ER+) patients, and adequate evidence to support the association between RS score and overall chemotherapy benefit, particularly for patients in the high risk category [23]. Recent publication of interim results for the TAILORx study has shown a further association between recurrence score and 5-year disease-free survival and distant recurrence in patients at low risk [24].

A follow-up systematic review by the EGAPP Working Group published in 2016 confirmed its previous findings but also noted the lack of direct evidence that the use of Oncotype DX improves clinical outcomes [25]. It also highlighted contradictory cost-effectiveness results in studies performed in the U.S. (for which \$2,000 in cost savings per patient were due to a decrease in post-testing chemotherapy use) and the U.K. (for which an incremental cost-effectiveness ratio of 26,940 £/QALY gained were due to an increase in chemotherapy use) [25]. The U.K. National Institute for Health and Care Excellence (NICE) Diagnostics Advisory Committee concluded that, under the assumption of equal chemotherapy benefit for all Oncotype DX risk categories, the test is not cost-effective at its current pricing level. In the first EGAPP review, data were adequate to support an association between the MammaPrint Index and 5- to 10-year metastasis rates, but the efficacy relative to classical clinical factors was unclear. More conclusive results await the completion of the MINDACT trial. For the H:I ratio test, populations studied were quite heterogeneous, and the test's commercial offering was based on a single study in women with primary, untreated breast cancer.

The Precision Medicine Approach to Colorectal Cancer

Colorectal cancer (CRC) is the third most common cancer in both men and women in the U.S. [26]. It is the third leading cause of cancer deaths when men and women are considered separately, and the second leading cause in both sexes combined.

Considerable effort has been placed into targeted therapies and companion diagnostics for CRC. About 70-80% of patients present with resectable localized disease, treated by surgery and often followed by adjuvant therapy [27]. CRC patients with advanced disease may receive first-line chemotherapy, or chemotherapy and radiation before surgery is considered. The EGAPP Working Group published in 2009 a systematic review of one first-line chemotherapeutic agent, irinotecan, which may be accompanied by *UGT1A1* genotyping to check for ability to adequately clear the drug [27]. Such genotyping aids decisions to either increase drug dose for more aggressive therapy, or switch to common alternate drugs, bevacizumab or cetuximab, for instance, in individuals with reduced clearance at risk for adverse events. EGAPP had two major findings: (1) unless patients receive a certain threshold dose of irinotecan, the increase in risk for toxicity is not significant, thus testing may not be warranted; and (2) reducing the dose of irinotecan (personalized dosing) to avoid adverse events may lead to more cases of unresponsive tumors than instances of adverse events avoided. It is inconclusive that benefits outweigh harms in the application of a precision approach here.

The alternate drugs mentioned above are positioned among the three main classes of targeted therapies approved for metastatic CRC: (1) multikinase inhibitors; (2) angiogenic inhibitors; and (3) anti-EGFR antibodies [15]. Up to 50% of CRC patients respond to anti-

EGFR therapy, which includes cetuximab [28]. *KRAS* wild-type status is considered an important factor in achieving a clinical response from this category of therapy [29]. However, 40-60% of patients with wild-type status do not respond to such therapy. Data suggest that *BRAF* gene status also plays a role in anti-EGFR response, and that the *BRAF* V600E mutation, present in 5-10% of CRC tumors, can lead to a positive response. To date the FDA has cleared two genetic testing kits, the *therascreen* *KRAS* kit from Qiagen, and the cobas *KRAS* Mutation Test from Roche [15]. The American Society for Clinical Oncology (ASCO) and three other professional organizations released new consensus guidelines in 2015 strongly recommending *KRAS* mutational testing for patients being considered for anti-EGFR therapy, and that *BRAF* V600 testing also be considered.

Table 1. Pharmacogenomic/Companion diagnostic approach – 3 cancers

Condition	Therapy	Biomarkers	Cost-Effectiveness	References
Non-small cell lung cancer (NSCLC)	Tyrosine kinase inhibitors	<i>EML4-ALK</i>	\$255,970/QALY gained	Djalalov et al. [19]
		EGFR	\$489,338 (Austr.)/QALY gained	Doble et al. [20]
HER2+ Breast cancer	Herceptin	IHC, FISH	\$3,351 – 12,230 (Can.) saved per case	Dendukuri et al. [22]
ER+ Breast cancer	Endocrine	21-gene assay	26,940 £ (Brit.) cost/QALY gained	EGAPP [25]; Ward et al. [33]
Metastatic colorectal cancer	Topoisomerase 1 inhibitor (TOP1) – irinotecan	<i>UGT1A1</i> genotyping	ambivalent results – reducing toxicity vs. reducing tumor burden	EGAPP [27]
	Anti-EGFR	<i>KRAS</i>	\$7,500 – 12,400 saved per case	EGAPP [30]; Vijayaraghavan et al. [31]
			\$180,000/QALY gained	EGAPP [30]; Shiroiwa et al. [32]

The EGAPP Working Group conducted a systematic review of companion diagnostic use of *KRAS* and *BRAF* mutation testing in anti-EGFR therapy in 2013 [30]. The systematic review looked at multiple pooled assessments, each consisting of up to seven studies, which together indicated statistically significant increased response rates to cetuximab and other anti-EGFR drugs, reduced risk of disease progression, and enhanced overall survival with *KRAS* wild-type status. EGAPP cited a cost-effectiveness study of *KRAS* testing in metastatic CRC patients in the U.S. and Germany [30, 31]. For the U.S. patients, use of *KRAS* testing to select patients for EGFR inhibitor therapy saved \$7,500 to \$12,400 per case. A second cited study from Japan displayed an incremental cost-effectiveness ratio for cetuximab with *KRAS* testing to be \$180,000 per QALY gained compared to therapy without testing [32]. Due to this cost, the investigators concluded that the protocol was not cost-effective, even when treatment was limited to patients with wild-type *KRAS*. Two of three studies looking at *BRAF* testing found associations similar to the *KRAS* studies in terms of progression-free and overall survival, but these findings, however, were not contingent on whether or not cetuximab was included in the combination therapy. Thus, while the review supported recommendations for a precision

approach using *KRAS* mutation testing, it found insufficient evidence to link *BRAF* V600E mutation testing with treatment response independent of prognostic association.

Table 1 summarizes the cost-effectiveness and descriptive findings for the pharmacogenomics or companion diagnostic precision medicine approach to the three cancers.

TAKING PRECISION MEDICINE BEYOND THE PERSONALIZED LEVEL

One argument against the informativeness of the above studies is that they represent a personalized medicine approach, but not a precision medicine approach in its full capacity [3]. The electronic storage of information, which can be multifactorial, including relevant lifestyle, would seem to be crucial to maximizing benefits and minimizing cost. The investigative team looking at the infrastructure characteristics of the Lee Moffit Cancer Center “Total Cancer Care” program also examined six other major healthcare programs engaged in cancer care comparative effectiveness research in genomic and precision medicine [10]. Four of the researched programs – at Duke University, Kaiser Permanente, University of Pennsylvania, and the Fred Hutchinson Cancer Center – had established a complex infrastructure (with data and biospecimen registries and multidisciplinary research teams), were engaged in “knowledge generation” (via randomized controlled trials and observational studies), and had reached the stage of “knowledge synthesis” (horizon scanning, evidence synthesis, and decision modeling). These programs display a depth of information and range of multidisciplinary expertise that is reflective of the type of knowledge synthesis of which the PMI’s million person cohort will be capable.

The Duke University and Kaiser Permanente teams noted a number of strategic challenges to the amassing of cancer study data – limited data quality, large variation in genomic methodology used, and poor demonstration of clinical utility for the genomic tests supplying the data [10]. These observations point out challenges that could foreseeably face PMI investigators attempting to collate information from the expansive precision medicine cohort. The Kaiser Permanente group was able to show that in screening for Lynch syndrome, a hereditary form of CRC, microsatellite instability testing (MSI) was preferable compared to immunohistochemical staining (IHC). In the treatment phase, the investigators found that screening for *KRAS* and *BRAF* mutations improved the cost-effectiveness of anti-EGFR therapy, but that the cost of the therapy surpassed generally accepted cost-effectiveness thresholds of \$100,000/quality-adjusted life-year. These findings allude to the capability of the PMI to develop new data on useful oncogenomic screening, mutational testing, and therapeutic procedures, and to the correspondingly likely possibility that many of the discoveries, while being individually beneficial, could elude effectiveness for the clinical population as a whole.

The PMI will no doubt be carried in new directions that have not yet been quantified, however. NCI is engaged a new type of clinical trial called “NCI-MATCH” [34]. In this innovative program, adult cancer patients are assigned to targeted treatments based on the genetic abnormalities in their tumors, regardless of the type of cancer they have. This concept represents the therapeutic end of what has been happening with diagnostics. *KRAS* and EGFR mutation testing is now occurring for multiple cancer types. Why should cancer therapies be restricted to just one cancer type? New management models may also evolve. The data

infrastructure supporting precision medicine can and is being used to develop procedural algorithms that combine genetic testing with genetically targeted therapies. These algorithms, if tailored to the lay person, can then be used by both medical providers and patients in collaboration, improving the effectiveness of the prescribed regimen. Shen and Hwang use *CYP2C9* (they cite *CYP450*) testing for warfarin dose performed at home first by nurses then by patients as an example [6]. This linkage of “big” or “rich” data and the development of new, useful algorithms is cited by multiple authors [2, 5, 35]. The issue is whether providers of various types, and consumers, can understand the test results and appreciate the connection to one-size-does-not-fit-all therapeutic management [36].

Authors discussing the translation of big data into usable guidelines and algorithms are not just limiting the payoff to individualized treatment, however. Jameson and Longo, for instance, speak of not just a pharmacogenomic future, but one in which “guideline-based screening”, e.g., colonoscopy, can be targeted on the basis of age and family history [5]. Family health history is one of several tools, including individualized genetic testing and family cascade testing, fueling the public health drive towards precision interventions [37, 38].

The public health approach to precision medicine, or “precision public health”, is highly evidence-based. Precision public health is exemplified in the Centers for Disease Control and Prevention’s emphasis on the implementation of Tier 1 genetic tests that have passed systematic review for analytic and clinical validity and utility – family history-based hereditary breast and ovarian cancer (HBOC) genetic testing (*BRCA1/2* mutations in relatives), and hereditary nonpolyposis colorectal cancer genetic testing (Lynch syndrome *MLH1*, *MSH2*, *MSH6*, *PMS2*, *EPCAM* mutations) [39]. Indeed, genetic counseling and testing for HBOC are already incorporated into healthcare reform as services not requiring co-pay for individuals deemed at risk by their providers [40]. The PMI promises much more, however, and part of the vision of public health is that precision techniques can be used to direct genetic preventive strategies to those subsets of the population that will derive maximal benefit [41].

PUBLIC HEALTH INTERVENTIONS AND COST-EFFECTIVENESS

The Precision Public Health Approach to Lung Cancer

It has been remarked that “although personalized treatments can help save the lives of sick people, prevention applies to all” [3]. This comment applies especially to lung cancer, which can be tackled as we have seen individually and after it has manifested, or by using a preventive, population-based approach. Initial genome-wide association studies (GWAS) published by several investigative teams in 2008 were suggestive of lung cancer susceptibility genes being situated on the long arm of chromosome 15 [42]. The studies were all large (3,500 to 14,000 participants) and replicated, yielding strong evidence for an association between SNP variations at 15q24/15q25.1 and lung cancer. Thorgeirsson et al. found a highly significant association ($P = 1.5 \times 10^{-8}$) between a common variant in the nicotinic acetylcholine receptor gene cluster on chromosome 15q24 and smoking frequency, with an odds ratio of 1.31 (1.19 – 1.44, 95% C.I.) between nicotine dependent cases and low quantity smokers plus population controls [43]. Studies by Hung et al. [44] and Amos et al. [45] found odds ratios between 1.21 and 1.77 for associations between the nicotinic acetylcholine receptor regions 15q25 and 15q25.1 and lung

cancer in ever smokers, the former accounting for 14% (attributable risk) of the lung cancer cases in the first study. The second study examined 2,724 NSCLC cases, the same type of cancer being treated in later stages by pharmacogenomics regimens.

More recent GWAS in never-smoking Asian females have pointed out genetic associations independent of smoking status. Case-control studies of never-smoking Asian females funded through the National Institutes of Health [46] and the Mayo Foundation [47] have identified genetic variants in the 3q28 and 13q31.3 regions associated with risk for lung cancer (N = 754 to 7,254 participants). These associations show both statistical significance ($P = 10^{-6}$ to 10^{-8}) in terms of odds ratios and allelic risk, and biological plausibility (association with the regulation of cell proliferation and division). These two lines of discovery – lung cancer in connection with nicotine dependence and independent of it – could beneficially lead to both personalized smoking-cessation interventions, and to increased screening of people at risk for lung cancer [41]. The difficulty is that the association studies are at the primary research (T1) stage and do not yet imply clinical validity.

Precision medicine in public health terms involves targeting groups at risk. Various professional societies have developed guidelines for lung cancer screening, generally beginning at age 55 [48]. The NCCN lung cancer screening guidelines recommend screening individuals age 50-55 years, those who have between 20 and 30 pack-years of exposure, and who exhibit one additional risk factor, such as family history. The relative risk of developing lung cancer is 1.8 if the individual has at least one first-degree relative with lung cancer, and 3.0 given two first-degree relatives with the condition [49]. The positive and negative predictive values for lung cancer appearing in a proband's first-degree relatives are 89.9% and 99.1%, respectively [50]. In the Utah Family High Risk Program, the cost of taking a family health history varied between \$10 and \$25 depending upon receipt of follow-up educational interventions [51]. These figures indicate cost-effectiveness for the use of family history in risk assessment for lung cancer.

Public health programs are especially concerned with the rights and welfare of underserved groups, an aim that is built into public health codes of ethics [52]. Surprisingly, of the 58,160 lung disease studies published between 1993 and 2013, less than 5% reported the inclusion of minority participants [53]. NIH is presently consulting researchers adept at recruiting under-represented groups into studies as part of PMI Research Cohort formation. An admixture study of 1812 African Americans performed by the Karmanos Cancer Institute in Detroit, MI demonstrates what can come out of the use of the PMI Cohort. Excess African ancestry was observed on chromosome 3q among ever smokers with NSCLC, a chromosomal region identified by previous studies with mostly persons of European ancestry [54].

The Precision Public Health Approach to Breast Cancer

About 10 to 15% of women diagnosed with breast cancer have germline mutations in the *BRCA1* or 2 genes. Between 10 and 30% of women under age 60 diagnosed with triple-negative breast cancer (cancer which does not have receptors for estrogen, progesterone, or *HER2/neu*) display a *BRCA1/2* mutation [55]. Ashkenazi Jewish ancestry confers an increased risk, though such mutations are by no means relegated to just one group. Like lung cancer, the public health approach to hereditary breast and ovarian cancer (HBOC) focuses on primary prevention before disease has appeared. Primary prevention can be conducted through a variety of means, several

of which fit under a public health precision model. A study out of the Cleveland Clinic Genomic Medicine Institute compared two methods for cancer risk assessment – “family history-based risk assessment” and “DTC [Direct-to-Consumer] personal genomic screening,” the latter using a variety of risk alleles [56]. Of 22 high risk females appearing in clinic, family history classified eight individuals as being at risk for breast cancer, but only one of the eight was classified as high-risk through personal genomic screening method.

Family history is a quick way of identifying risk, and has value in this instance as it does with lung cancer. Valdez et al. note that the relative risk of developing breast cancer is 1.8 and ovarian cancer 2.9 if the individual has at least one first-degree relative with these conditions [49]. It is 3.0 and 14.7 given two affected first-degree relatives. The positive and negative predictive values for these cancers appearing in a proband’s first-degree relatives are 89.1% and 98.9% for breast cancer, and 76.1% and 99.3% for ovarian cancer [50]. The U.S. Preventive Services Task Force (USPSTF) has conducted evidence reviews of genetic testing for the key mutations involved in HBOC, *BRCA1* and 2. It recommends:

Primary care providers screen women who have family members with breast, ovarian, tubal, or peritoneal cancer with 1 of several screening tools designed to identify a family history that may be associated with an increased risk for potentially harmful mutations in breast cancer susceptibility genes (*BRCA1* or *BRCA2*). Women with positive screening results should receive genetic counseling and, if indicated after counseling, *BRCA* testing. [57]

The CDC-OPHG classifies the use of family history of known breast/ovarian cancer with deleterious *BRCA* mutations as Tier 1 [39]. CDC defines Tier 1 genetic interventions as those supported by clinical practice guidelines based on thorough systematic review. These modalities are ready for implementation not just on the individual but on the level of the at-risk population as well [38].

Family history, however, is part of a train of diagnostic interventions, including genetic counseling and genetic testing. The latter can lead to annual screening via MRI or to surgery. A 2012 cost-effectiveness analysis of *BRCA1/2* testing of women ≥ 35 years at elevated risk of carrying a mutation, considering the eventual use of these MRI and surgery, determined genetic testing to be cost-effective if testing cost were $\leq \$8,948$ [58]. Currently targeted *BRCA1/2* mutation testing ranges from \$4,500 to \$650 [58, 59].

Economies of scale operate when identifying persons at risk for cancer. Identifying single individuals can be expected to be less efficient than pinpointing families at risk, just as using information reflecting population risk data can be expected to be more efficient than utilizing separate family histories. Cascade screening is the testing of successive test positive family members starting with a positive index case. Two studies delineate that the success of *BRCA1/2* cascade screening in affected families, as measured by actual uptake of genetic testing, is a function of ability of family members to communicate with one another, and to attend a clinical session [60, 61]. Uptake rates of 54% (Utah study) and 73% (Manchester portion of U.K. study) were achieved when these conditions were satisfied. The London portion of the U.K. study, which covered a more dispersed population, achieved 62% for those clinically seen. George et al. indicate that the cost of *BRCA* testing in 30 at-risk relatives can be brought down from a minimum of \$66,000 to \$12,000 once the responsible mutation has been identified in a family member [62].

On a population level, it is possible to gather information on early-stage breast and ovarian cancer, and to report aggregate information back to participating institutions. A breast cancer incidence study utilizing the Connecticut Tumor Registry found that the ability to detect cases depends on the relative population size of the groups being assessed (here white versus nonwhite) [63]. Population-based risk assessment has the advantage of detecting *BRCA1/2* carriers with a negative family history. Clinical validity goes up with the prevalence of a disorder in a given population. Rubinstein et al. performed a decision analysis of *BRCA1/2* testing in American Ashkenazi Jewish women aged 35-55 years [64]. At a cost of \$460 for founder mutation testing, the investigators concluded that such a program is cost-effective, amounting to \$8,300/QALY gained.

The PMI, like state cancer registries, is especially geared towards targeting at-risk individuals and populations. The CDC notes that state health departments have engaged in bidirectional reporting, i.e., identification of cases from the state tumor registry, and reporting back of information to participating facilities. The Michigan Department of Health and Human Services and Connecticut Department of Health have both identified thousands of cases of breast and ovarian cancer suggesting risk for HBOC, and returned back facility-specific data [65]. In the case of Connecticut, facility-specific reports on numbers of breast and ovarian cancers, compiled from 3,792 cases in the state cancer registry, were reported back to providers to alert them that patients might be at an increased risk for HBOC. Staff at 70% of the 32 involved hospitals also requested and received an HBOC training session.

In the age of electronic health records (EHRs), systems can be designed to provide physicians with decision support tools that alert them to the need for genetic testing, and assist with interpretation of results and potential impact on patients and their families [11]. It is also possible with EHR systems to flag patients who are members of high-risk families to make them aware of their risk status and available options [66, 67].

The Precision Public Health Approach to Colorectal Cancer

The hereditary form of CRC that has received the most public health and medical attention is hereditary nonpolyposis colorectal cancer (HNPCC) or Lynch syndrome (LS), which produces only a small number of polyps or none at all. LS is the most common heritable cause of CRC, representing 1 in 35 cases [66]. Whereas the lifetime risk for developing CRC is 2 to 5% in the general population, it is 80% in those with LS.

Family history as an instrument for performing cancer risk assessment in first- and second-degree relatives is even more accurate for CRC than for the other two families of cancers. Valdez et al. report that the relative risk for CRC is 2.2 given a first-degree relative with the condition, and 4.0 given at least two first-degree relatives with CRC [49]. Lynch syndrome is notoriously underdiagnosed, however. The vast majority of families with a history of CRC do not know they may have LS, or even that a genetic test is available [66]. Nevertheless, the EGAPP Working Group, using a chain of evidence methodology comparing four preliminary testing (MSI, IHC, and *BRAF*) and mismatch repair (MMR) gene strategies leading to medical and surgical management, found that adequate evidence exists that an appropriate testing strategy can improve clinical outcomes for patients and their families [68].

A precision public health approach to Lynch syndrome entails spotting cases before the management is made more complex by disease advancement, coupled with identifying at-risk

relatives free of disease. A 2014 consensus statement by the U.S. Multi-Society Task Force on Colorectal Cancer looked at initial approaches for index case identification. Use of clinical criteria (e.g., the Revised Bethesda Guidelines) and a CRC risk assessment tool using family history – “selective approaches” used to identify a range of patients – both showed adequate sensitivity in identifying individuals with germline mutations, but up to 28% of LS patients can be missed with a liberal interpretation of the revised guidelines [69]. A more “universal approach” assessing LS risk based on IHC testing of all incoming cases of CRC was found to have greater sensitivity than selective strategies including the use of the Bethesda guidelines. Such an approach is still shy of being considered population-wide risk assessment in that testing begins with confirmed colorectal cancer cases. Multiple studies have shown that: (1) the systematic application of testing among patients with newly diagnosed CRC could provide substantial clinical benefits at acceptable costs; and (2) adopting a “universal” approach towards CRC genetic testing is cost-effective [69]. Palomaki et al., for example, found that a strategy employing IHC and *BRAF* preliminary testing followed by MMR mutation testing in *BRAF* negative cases would cost an average of \$18,863 per LS case detected [70].

Having identified an index case, the public health approach is then to move on to relatives to identify those at risk for LS, and those who are not. The EGAPP Working Group cites seven studies showing how uptake of colonoscopy by relatives with LS ranged from 53 to 100% [68]. Sharaf et al., in a sub-analysis of four select articles from a broader LS studies review, concluded that 52% or less of the first-degree relatives received LS genetic testing [71]. This number was critiqued by Jasperson, who noted that the Sharaf analysis excluded a large study containing 466 family members composed almost exclusively of first-degree relatives at high-risk of developing LS [72]. The latter study found 75% of the members to have been tested for LS mutations.

The cost of cascade testing is also a consideration. The per mutation cost of MMR gene testing for LS can decrease from \$1,000 to \$350 once a specific familial mutation has been identified in the proband [59]. The consequence is that the cost per LS index case can decrease from \$18,863 to \$13,000 [68]. In the most comprehensive cost-effectiveness assessment of universal and “near-universal” genetic screening to date, covering seven LS studies that also looked at medical/surgical follow-up and testing of relatives, Grosse found that all except one reported incremental cost-effectiveness ratios less than the threshold \$100,000 per life-year or quality-adjusted life-year gained [73].

As with HBOC, collection and reporting of LS data has occurred at the state level. Michigan identified 10,000 CRC cases from its cancer registry, and returned facility-specific data and educational materials to 145 reporting institutions [74]. Connecticut reported back information on 3,517 CRC cases in its cancer registry to targeted practitioners at acute care hospitals. These bidirectional reporting efforts show that health-related information may be stored *en masse* for both practical and research purposes (as in the PMI Cohort), and can be reissued for clinical and public health purposes.

The Colorado Department of Public Health and Environment also targeted educational outreach relating to hereditary CRC to 430 medical providers and 200 at-risk cases [74]. A survey was conducted which showed that 98% of provider and patient respondents thought that the information was clear and useful, many patients having indicated they engaged in further dialogue with their physician, a genetic counselor, or a family member. About 1/3 of the patient respondents communicated that they planned to have a risk assessment as a result of receiving the materials. The Colorado educational efforts were conducted by mail. An extensive, 33-

study article review by Naylor et al. found that further patient education involving phone or in-person contact combined with patient navigation of resources can lead to improvements on the order of 15% in colorectal screening rates in minority populations [75].

Table 2 outlines the array of precision public health efforts, which involves targeting of screening, genetic testing, and education for the three cancer conditions.

Table 2. Public health precision approach – 3 cancers

Condition	Genetic test	Cost-effectiveness	References
Familial lung cancer	Family history (FH)	*FP rate: 1.9 (female) – 6.1 (male) FN rate: 29.1 (female) – 37.5 (male) **PPV: 81.0 (68.6 – 90.1)*** NPV: 97.1 (96.0 – 98.0) Cost \$10 – 25/FH taken	Ziogas and Anton-Culver [50] Johnson et al. [51]
Hereditary breast and ovarian cancer (HBOC)	<i>BRCA1/2</i>	Cost-effectiveness threshold \$8,948 (current testing \$650 – 4,500) \$8,300/QALY gained (U.S. Ashkenazi population)	Goodman [58] Rubinstein et al. [64]
Hereditary nonpolyposis colorectal cancer (Lynch syndrome – LS)	Multi-gene testing (<i>MLH1</i> , <i>MSH2</i> , <i>MSH6</i> , <i>PMS2</i> , <i>EPCAM</i>)	\$18,863 cost/index case \$13,000 cost/relative \$29,600 – 63,900/QALY gained (“universal” testing)	EGAPP [68]; Palomaki et al. [70] Grosse [73]

* False positive (FP) and false negative (FN) rates.

** Positive predictive value (PPV) and negative predictive value (NPV).

*** 95% Confidence interval.

CONCLUSION: CONTINUITY AND EMERGENCE IN PRECISION HEALTH APPROACHES

A major point brought out by the two tables is that pharmacogenomic regimens cost considerably more than primary prevention approaches to cancer management. From the articles reviewed herein, the former often exceed cost-effectiveness thresholds, though this conclusion need not necessarily be true for the regimens that are being offered. The cancer drug being prescribed is typically the main contributor to the cost, as opposed to the companion diagnostic, which can actually bring the price down. One can then view this conclusion in three ways: (1) optimistically, economies of scale or institutional policy changes might ultimately bring the cost of the pharmacogenomic regimen down; (2) the value of a person’s life is to be held paramount, thus the costs incurred, from a private perspective, are worth it; or (3) the value of the primary prevention approach, due to its cost-effectiveness, is to be emphasized.

The medical and public health approaches are not to be viewed as mutually exclusive. Their territories are more like the circles of a Venn diagram which intersect with plenty of shared space in the middle. Both the CDC and its EGAPP program have devoted considerable attention to companion diagnostics and the attempt to prove that they have overall utility, including cost-effectiveness. Public health has an evaluative role that dips into hospital-based interventions. The national view is more one of mergence [3]. The fact that *BRCA1/2* genetic testing takes off where chemotherapeutic approaches to breast cancer leave off, in terms of the host cancer's hormone receptor status, indicates that the domains of the primary/secondary and tertiary prevention camps are complementary. It would be idealistic to assume that predictive genetic testing plus monitoring could eliminate all tumors before they spread. Both ends of the prevention continuum are needed, and the more targeted they are, the better. It is also important to recognize that the medical and public health approaches both defy stereotyping. The use of *KRAS* testing in anti-EGFR therapy of metastatic colorectal cancer can be cost-saving. The value of a public health approach genetic to lung cancer is limited by the current state of GWAS findings and the heavy influence of gene-environment interactions in the genesis of lung cancer. Indeed, some would argue that federal and state policy approaches towards smoking and lung cancer should take center stage.

Policymakers have a role in deciding allocation of healthcare dollars. In providing arguments for and against utility in this paper, we are not advocating a withdrawal of dollars from either medical or public health oncogenomic approaches. The case has been made for a more detailed consideration of the proportion of dollars that go towards prevention, however. The paper has not dealt just with chemotherapy, or just with Tier 1 genetic testing, however – emphasis has been placed on a precision approach for both. In public health, this nuance translates into tailoring the intervention towards communities at-risk or in-need. It will be more cost-effective thereby, and will achieve the ethical standard of addressing the health of all by attending to those most in need.

A final point should be made about the medical and public health precision approaches that are emerging. The medical approach is showing healthy signs of growth. The PMI Research Cohort will grant it the ability to host more clinical trials with suitable cohorts of participants, and to speed up the process of conducting clinical trials. As NCI plans, some of these trials will cross typological borders and target treatments based on somatic abnormalities in the tumors regardless of cancer type [34]. Public health research will also benefit from the PMI Cohort, especially in instances where major germline mutations have not turned up and GWAS are actively identifying lower risk alleles having a cumulative effect. Given the thought being given to participating groups, the discovery of new risk alleles will most certainly benefit diverse communities. The promise of tapping into lifestyle and environmental factors as data on the cohort participants builds will strengthen prevention approaches even further. The word “precision” also applies to the educational component of health programs. Educational efforts can become more tailored to the needs of at-risk groups as the data accumulates. Precision public health does not end with the target molecule, but with the whole person and the community to which he or she belongs.

REFERENCES

- [1] The White House, Office of the Press Secretary. *FACT SHEET: President Obama's Precision Medicine Initiative*. Washington, D.C. 2015. <https://obamawhitehouse.archives.gov/the-press-office/2015/01/30/fact-sheet-president-obama-s-precision-medicine-initiative>.
- [2] Collins, F.S., Varmus, H. A new initiative on precision medicine. *N. Engl. J. Med.* 2015; 372(9): 793-795.
- [3] Khoury, M.J., Galea, S. *Precision public health: more precision ahead for individual and population interventions*. Atlanta, GA. 2016. https://blogs.cdc.gov/genomics/2016/09/07/precision_public_health.
- [4] Khoury, M.J., Gwinn, M., Yoon, P.W., Dowling, N., Moore, C.A., Bradley, L. The continuum of translation research in genomic medicine: how can we accelerate the appropriate integration of human genome discoveries into health care and disease prevention? *Genet. Med.* 2007; 9(10): 665-674.
- [5] Jameson, J.L., Longo, D.L. Precision medicine – personalized, problematic, and promising. *N. Engl. J. Med.* 2015; 372(23): 2229-2234.
- [6] Shen, B., Hwang, J. The clinical utility of precision medicine: properly assessing the value of emerging diagnostic tests. *Clin. Pharmacol. Ther.* 2010; 88(6): 754-756.
- [7] Institute of Medicine (U.S.) Roundtable on Translating Genomic-Based Research for Health. *The value of genetic and genomic technologies: workshop summary*. Board on Health Sciences Policy. Washington, D.C.: The National Academies Press; 2010.
- [8] Pray, L.A. Gleevec: the breakthrough in cancer treatment. *Nature Educ.* 2008; 1(1): 37.
- [9] Cancer.Net. *Leukemia – chronic myeloid – CML: statistics*. 2015. <http://www.cancer.net/cancer-types/leukemia-chronic-myeloid-cml/statistics>.
- [10] Simonds, N.I., Khoury, M.J., Schully, S.D., Armstrong, K., Cohn, W.F., Fenstermacher, D.A., Ginsburg, G.S., Goddard, K.A.B., et al. Comparative effectiveness research in cancer genomics and precision medicine: current landscape and future prospects. *J. Natl. Cancer Inst.* 2013; 105(13): 929-936.
- [11] Mirnezami, R., Nicholson, J., Darzi, A. Preparing for precision medicine. *N. Engl. J. Med.* 2012; 366(6): 489-491.
- [12] Khoury, M.J., Gwinn, M.L., Glasgow, R.E., Kramer, B.S. A population approach to precision medicine. *Am. J. Prev. Med.* 2012; 42(6): 639-645.
- [13] American Cancer Society. *Key statistics for lung cancer*. 2016a. <https://www.cancer.org/cancer/non-small-cell-lung-cancer/about/key-statistics.html>.
- [14] Cancer Research Institute. *Lung cancer*. 2016. <http://www.cancerresearch.org/cancer-immunotherapy/impacting-all-cancers/lung-cancer?gclid=CM3a26TUqNACFQIDaQod5nsDaw>.
- [15] Millner, L.M., Strotman, L.N. The future of precision medicine in oncology. *Clin. Lab. Med.* 2016; 36(3): 557-573.
- [16] Cancer.Net. *Lung cancer – non-small cell – statistics*. 2016. <http://www.cancer.net/cancer-types/lung-cancer-non-small-cell/statistics>.
- [17] Ritchie, M.D. The success of pharmacogenomics in moving genetic association studies from bench to bedside: study design and implementation of precision medicine in the post-GWAS era. *Hum. Genet.* 2012; 131(10): 1615-1626.

- [18] Brunstein, J. Cost-effectiveness considerations with molecular diagnostics in oncology. *MLO Med. Lab. Obs.* 2016; 48(5): 30-31.
- [19] Djalalov, S., Beca, J., Hoch, J.S., Krahn, M., Tsao, M.-S., Cutz, J.-C., Leighl, N.B. Cost effectiveness of EML4-ALK fusion testing and first-line crizotinib treatment for patients with advanced ALK-positive non-small-cell lung cancer. *J. Clin. Oncol.* 2014; 32(10): 1012-1019.
- [20] Doble, B., John, T., Thomas, D., Fellowes, A., Fox, S., Lorgelly, P. Cost-effectiveness of precision medicine in the fourth-line treatment of metastatic lung adenocarcinoma: an early decision analytic model of multiplex targeted sequencing. *Lung Cancer* 2016; doi: 10.1016/j.lungcan.2016.05.024. [Epub ahead of print]
- [21] American Cancer Society. *What are the key statistics about breast cancer?* 2016b. <http://www.cancer.org/cancer/breastcancer/detailedguide/breast-cancer-key-statistics>.
- [22] Dendukuri, N., Khetani, K., McIsaac, M., Brophy, J. Testing for HER2-positive breast cancer: a systematic review and cost-effectiveness analysis. *Can. Med. Assoc. J.* 2007; 176(10): 1429-1434.
- [23] Evaluation of Genomic Applications in Practice and Prevention (EGAPP) Working Group. Recommendations from the EGAPP Working Group: can tumor gene expression profiling improve outcomes in patients with breast cancer? *Genet. Med.* 2009a; 11(1): 66-73.
- [24] Sparano, J.A., Gray, R.J., Makower, D.F., Pritchard, K.I., Albain, K.S., Hayes, D.F., Geyer, C.E., Jr., Dees, E.C., et al. Prospective validation of a 21-gene expression assay in breast cancer. *N. Engl. J. Med.* 2015; 373(21): 2005-2014.
- [25] EGAPP Working Group. Recommendations from the EGAPP Working Group: does the use of Oncotype DX tumor gene expression profiling to guide treatment decisions improve outcomes in patients with breast cancer? *Genet. Med.* 2016; 18(8): 770-779.
- [26] American Cancer Society. *Key statistics for colorectal cancer.* 2016c. www.cancer.org/cancer/colon-rectal-cancer/about/key-statistics.html.
- [27] EGAPP Working Group. Recommendations from the EGAPP Working Group: can *UGT1A1* genotyping reduce morbidity and mortality in patients with metastatic colorectal cancer treated with irinotecan? *Genet. Med.* 2009b; 11(1): 15-20.
- [28] Markman, M. *Colorectal cancer and KRAS/BRAF.* 2016. [http:// emedicine.medscape.com/article/1690010-overview](http://emedicine.medscape.com/article/1690010-overview).
- [29] Wang, G., Kelley, R.K., and GAPPNet. KRAS mutational analysis for colorectal cancer. *PLoS Curr.* 2010; 1: doi: 10.1371/currents.RRN1175.
- [30] EGAPP Working Group. Recommendations from the EGAPP Working Group: can testing of tumor tissue for mutations in EGFR pathway downstream effector genes in patients with metastatic colorectal cancer improve health outcomes by guiding decisions regarding anti-EGFR therapy? *Genet. Med.* 2013; 15(7): 517-527.
- [31] Vijayaraghavan, A., Efrusy, M.B., Göke, B., Kirchner, T., Santas, C.C., Goldberg, R.M. Cost-effectiveness of KRAS testing in metastatic colorectal cancer patients in the United States and Germany. *Int. J. Cancer* 2012; 131(2): 438-445.
- [32] Shiroyiwa, T., Motoo, Y., Tsutani, K. Cost-effectiveness analysis of KRAS testing and cetuximab as last-line therapy for colorectal cancer. *Mol. Diagn. Ther.* 2010; 14(6): 375-384.
- [33] Ward, S. Scope, A., Rafia, R., Pandor, A., Harnan, S., Evans, P., Wyld, L. Gene expression profiling and expanded immunohistochemistry tests to guide the use of

- adjuvant chemotherapy in breast cancer management: a systematic review and cost-effectiveness analysis. *Health Technol. Assess.* 2013; 17(44): 1-302.
- [34] National Cancer Institute. *NCI and the Precision Medicine Initiative*®. Bethesda, MD. 2016. <https://www.cancer.gov/research/key-initiatives/precision-medicine>.
- [35] Shaikh, A.R., Butte, A.J., Schully, S.D., Dalton, W.S., Khoury, M.J., Hesse, B.W. Collaborative biomedicine in the age of big data: the case of cancer. *J. Med. Internet Res.* 2014; 16(4): e101.
- [36] Graber, C. The problem with precision medicine. *The New Yorker* Feb 5, 2015. <http://www.newyorker.com/tech/elements/problem-precision-medicine>.
- [37] Marcus, P.M., Pashayan, N., Church, T.R., Doria-Rose, V.P., Gould, M.K., Hubbard, R.A., Marrone, M., Miglioretti, D.L., et al. Population-based precision cancer screening: a symposium on evidence, epidemiology, and next steps. *Cancer Epidemiol. Biomarkers Prev.* 2016; 25(11): 1449-1455.
- [38] Green, R.F., Dotson, W.D., Kolor, K., Khoury, M.J. Genomics in public health: perspective from the Office of Public Health Genomics at the Centers for Disease Control and Prevention (CDC). *Healthcare* 2015; 3(3): 830-837.
- [39] Centers for Disease Control and Prevention, Office of Public Health Genomics. *Public Health Genomics Knowledge Base* (v1.3). Atlanta, GA. 2016. <https://phgkb.cdc.gov/GAPPKB/topicStartPage.do>.
- [40] Centers for Medicare & Medicaid Services. *Affordable Care Act Implementation FAQs – Set 12*. Baltimore, MD. 2013. https://www.cms.gov/CCIIO/Resources/Fact-Sheets-and-FAQs/aca_implementation_faqs12.html.
- [41] Khoury, M.J., Evans, J.P. A public health perspective on a national precision medicine cohort: balancing long-term knowledge generation with early health benefit. *J. Natl. Med. Assoc.* 2015; 313(21): 2117-2118.
- [42] Chanock, S.J., Hunter, D.J. Genomics: when the smoke clears.... *Nature* 2008; 452(7187): 537–538.
- [43] Thorgeirsson, T.E., Geller, F., Sulem, P., Rafner, T., Wiste, A., Magnusson, K.P., Manolescu, A., Thorleifsson, G., et al. A variant associated with nicotine dependence, lung cancer and peripheral arterial disease. *Nature* 2008; 452(7187): 638-642.
- [44] Hung, R.J., McKay, J.D., Gaborieau, V., Boffetta, P., Hashibe, M., Zaridze, D., Mukeria, A., Szeszenia-Dabrowska, N., et al. A susceptibility locus for lung cancer maps to nicotinic acetylcholine receptor subunit genes on 15q25. *Nature* 2008; 452(7187): 633-637.
- [45] Amos, C.I., Wu, X., Broderick, P., Gorlov, I.P., Gu, J., Eisen, T., Dong, Q., et al. Genome-wide association scan of tag SNPs identifies a susceptibility locus for lung cancer at 15q25.1. *Nat. Genet.* 2008; 40(5): 616-622.
- [46] Hosgood, H.D., Wang, W.C., Hong, Y.C., Wang, J.C., Chen, K., Chang, I.S., Chen, C.J., Lu, D., et al. Genetic variant in TP63 on locus 3q28 is associated with the risk of lung adenocarcinoma among never-smoking females in Asia. *Hum. Genet.* 2012; 131(7): 1197-1203.
- [47] Li, Y., Sheu, C.C., Ye, Y., de Andrade, M., Wang, L., Chang, S.C., Aubry, M.C., Aakre, J.A., et al. Genetic variants and risk of lung cancer in never smokers: a genome-wide association study. *Lancet Oncol.* 2010; 11(4): 321-330.
- [48] Marcus, P.M., Freedman, A.N., Khoury, M.J. targeted cancer screening in average-risk individuals. *Am. J. Prev. Med.* 2015; 49(5): 765-771.

- [49] Valdez, R., Yoon, P.W., Qureshi, N., Green, R.F., Khoury, M.J. Family history in public health practice: a genomic tool for disease prevention and health promotion. *Annu. Rev. Public Health* 2010; 31: 69-87.
- [50] Ziogas, A., Anton-Culver, H. Validation of family history data in cancer family registries. *Am. J. Prev. Med.* 2003; 24(2): 190-198.
- [51] Johnson, J., Giles, R.T., Larsen, L., Ware, J., Adams, T., Hunt, S.C. Utah's Family High Risk Program: bridging the gap between genomics and public health. *Prev. Chronic Dis.* 2005; 2(2): A24.
- [52] Thomas, J.C., Irwin, D.E., Zulker, E.S., Millikan, R.C. Genomics and the Public Health Code of Ethics. *Am. J. Public Health* 2005; 95(12): 2139-2143.
- [53] Reardon, S. U.S. tailored-medicine project aims for ethnic balance. *Nature* 2015; 523(7561): 391-392.
- [54] Schwartz, A.G., Wenzlaff, A.S., Bock, C.H., Ruterbusch, J.J., Chen, W., Cote, M.L., Artis, A.S., Van Dyke, A.L., et al. Admixture mapping of lung cancer in 1812 African-Americans. *Carcinogenesis* 2011; 32(3): 312-317.
- [55] Cancer.Net. *Hereditary breast and ovarian cancer*. 2015. [http://www. cancer.net/ cancer-types/hereditary-breast-and-ovarian-cancer](http://www.cancer.net/cancer-types/hereditary-breast-and-ovarian-cancer).
- [56] American Society of Human Genetics. *New research validates clinical use of family health history as the 'gold standard' for assessing personal disease risk*. Bethesda, MD. 2010. [http://www.ashg.org/pdf/PR_ FamilyHealthHistory_110510.pdf](http://www.ashg.org/pdf/PR_FamilyHealthHistory_110510.pdf).
- [57] U.S. Preventive Services Task Force. *Final recommendation statement: BRCA-related cancer: risk assessment, genetic counseling and genetic testing*, December 2013. Rockville, MD. 2013. [https://www. uspreventiveservicestaskforce.org/Page/Document/ RecommendationStatementFinal/brca-related-cancer-risk-assessment-genetic-counseling-and-genetic-testing](https://www.uspreventiveservicestaskforce.org/Page/Document/RecommendationStatementFinal/brca-related-cancer-risk-assessment-genetic-counseling-and-genetic-testing).
- [58] Goodman, A. BRCA1/2 genetic testing found cost-effective in current era. *Am. J. Manag. Care* 2012; 18(3 Spec. No.): SP133.
- [59] American Public Health Association. *Advancing cancer genomics in public health*. Washington, D.C. 2013. [http://www.apha.org/policies-and-advocacy/public-health-policy-statements/policy-database/2014/07/ 23/10/25/advancing-cancer-genomics-in-public-health](http://www.apha.org/policies-and-advocacy/public-health-policy-statements/policy-database/2014/07/23/10/25/advancing-cancer-genomics-in-public-health).
- [60] Brooks, L., Lennard, F., Shenton, A., Lalloo, F., Ambus, I., Ardern-Jones, A., Belk, R., Kerr, B., et al. BRCA1/2 predictive testing: a study of uptake in two centres. *Eur. J. Hum. Genet.* 2004; 12(8): 654-662.
- [61] Botkin, J.R., Smith, K.R., Croyle, R.T., Baty, B.J., Wylie, J.E., Dutson, D., Chan, A., Hamann, H.A., et al. Genetic testing for a BRCA1 mutation: prophylactic surgery and screening behavior in women 2 years post testing. *Am. J. Med. Genet.* 2003; 118A(3): 201-209.
- [62] George, R., Kovak, K., Cox, S.L. Aligning policy to promote cascade genetic screening for prevention and early diagnosis of heritable diseases. *J. Genet. Counsel.* 2015; 24(3): 388-399.
- [63] Crabbe, J.C.F., Gregorio, D.I., Samociuk, H., Swede, H. Secular trends, race, and geographic disparity of early-stage breast cancer incidence: 25 years of surveillance in Connecticut. *Am. J. Public Health* 2015; 105(Suppl. 3): e64-e70.

- [64] Rubinstein, W.S., Jiang, H., Dellefave, L., Rademaker, A.W. Cost-effectiveness of population-based BRCA1/2 testing and ovarian cancer prevention for Ashkenazi Jews: a call for dialogue. *Genet. Med.* 2009; 11(9): 629-639.
- [65] Centers for Disease Control and Prevention, Office of Public Health Genomics. *Genomics implementation: more detailed information on key Tier 1 applications – hereditary breast and ovarian cancer (HBOC)*. Atlanta, GA. 2015. https://www.cdc.gov/genomics/implementation/toolkit/hboc_1.htm.
- [66] Modell, S.M., Greendale, K., Citrin, T., Kardia, S.L. Expert and advocacy group consensus findings on the horizon of public health genetic testing. *Healthcare* 2016; 4(1): pii: E14. doi: 10.3390/healthcare4010014.
- [67] Center for Public Health and Community Genomics, Genetic Alliance. *Priorities for public health genomics 2012-2017: stakeholder consultation/priorities conference report, September 2011*. Ann Arbor, MI. 2011. <https://stacks.cdc.gov/view/cdc/11738>.
- [68] EGAPP Working Group. Recommendations from the EGAPP Working Group: genetic testing strategies in newly diagnosed individuals with colorectal cancer aimed at reducing morbidity and mortality from Lynch syndrome in relatives. *Genet. Med.* 2009c; 11(1): 35-41.
- [69] Giardiello, F.M., Allen, J.I., Axilbund, J.E., Boland, C.R., Burke, C.A., Burt, R.W., Church, J.M., Dominitz, J.A., et al. Guidelines on genetic evaluation and management of Lynch syndrome: a consensus statement by the US Multi-Society Task Force on Colorectal Cancer. *Am. J. Gastroenterol.* 2014; 109(8): 1159-1179.
- [70] Palomaki, G.E., McClain, M.R., Melillo, S., Hampel, H.L., Thibodeau, S.N. EGAPP supplementary evidence review: DNA testing strategies aimed at reducing morbidity and mortality from Lynch syndrome. *Genet. Med.* 2009; 11(1): 42-65.
- [71] Sharaf, R.N., Myer, P., Stave, C.D., Diamond, L.C., Ladabaum, U. Uptake of genetic testing by relatives of Lynch syndrome probands: a systematic review. *Clin. Gastroenterol. Hepatol.* 2013; 11(9): 1093-1100.
- [72] Jaspersion, K. Cascade genetic testing in Lynch syndrome: room for improvement. *Nat. Rev. Gastroenterol.* 2013; 122: 505-508.
- [73] Grosse, S.D. When is genomic testing cost-effective? Testing for Lynch syndrome in patients with newly-diagnosed colorectal cancer and their relatives. *Healthcare* 2015; 3(4): 860-878. doi: 10.3390/healthcare 3040860.
- [74] Centers for Disease Control and Prevention, Office of Public Health Genomics. *Lynch syndrome phase 1*. Atlanta, GA. 2014. https://www.cdc.gov/genomics/implementation/toolkit/lynch_2.htm.
- [75] Naylor, K., Ward, J., Polite, B.N. Interventions to improve care related to colorectal cancer among racial and ethnic minorities: a systematic review. *J. Gen. Intern. Med.* 2012; 27(8): 1033-1046.

Chapter 70

NEUROPSYCHOLOGICAL PROFILE OF PEOPLE WITH WILLIAMS SYNDROME (WS)

Anne-Sophie Pezzino, Nathalie Marec-Breton and Agnès Lacroix*

Center for the Psychology of Cognition, Communication and Behavior
University of Rennes 2, Rennes, France

ABSTRACT

Williams syndrome (WS) is a genetic neurodevelopmental disorder (prevalence close to 1 in 20,000-30,000 births) resulting from the deletion of 16-25 genes on the long arm of Chromosome 7 (Scherer & Osborne, 2006). Individuals with WS have an intelligence quotient of 40-70 (Howlin, Davies, & Udwin, 1998). They have a unique neuropsychological profile, characterized by an apparent dissociation between cognition and language, as language is relatively well preserved, compared with other cognitive skills (Karmiloff-Smith, et al., 2004; Martens, Wilson, & Reutens, 2008). However, a more complex profile is now emerging, with good lexical, short-term memory (especially auditory-verbal) and face processing skills, but visuospatial (especially local processing of information), executive (planning and inhibition), memory (working memory and long-term) and attentional deficits (Bellugi, Lichtenberger, Jones, Lai, & George, 2000; Schmitt, Eliez, Warsofsky, Bellugi, & Reiss, 2001; Fayasse & Thibaut, 2003; Menghini, Addona, Costanzo, & Vicari, 2010; Costanzo et al., 2013; Dessalegn, Landau & Rapp, 2013).

Individuals with WS also have specific auditory-perceptual cognitive skills (hyperacusis), category-specific perception of speech sounds (Majerus, Palmisano, Van Der Linden, Barisnikov, & Poncelet, 2001; Majerus, Barisnikov, Vuillemin, Poncelet, & Linden, 2003), and musical skills that exceed their cognitive level (Levitin, Cole, Lincoln, & Bellugi, 2005). Other behavioral characteristics include hypersociability (Jones et al., 2000; Mervis, Morris, Klein-Tasman, et al., 2003).

In this chapter, we provide a review of the literature on the specific neuropsychological profile of individuals with WS, from a cognitive-behavioral and neuroanatomical point of view. Early studies of the neuropsychological profile in WS focused on the dissociation between cognition and language. Since then, research has shown this syndrome to be more complex.

* Corresponding Author's Email: agnes.lacroix@univ-rennes2.fr.

Our aim is to highlight the heterogeneity of the cognitive profiles observed in this syndrome, and to identify the factors that might explain this heterogeneity. The complexity and specific features of the neuropsychological profile in WS need to be understood in order to develop therapeutic and learning methods adapted to the developmental pace of individuals with WS.

Keywords: genetic syndrome, Williams syndrome, neuropsychological profile, learning abilities

INTRODUCTION

Williams syndrome (WS) is a rare genetic disease (one case per 20,000 births) caused by a microdeletion on the long arm of Chromosome 7 (7q11.23) leading to the loss of 16-25 genes (Scherer & Osborne, 2006). At the cognitive level, individuals with WS have an intelligence quotient (IQ) of about 40-70 (Howlin et al., 1998; Mervis, Morris, Bertrand, & Robinson, 1999). Their unique neuropsychological profile is characterized by a dissociation between oral language (relatively preserved) and other (impaired) cognitive abilities (Hoffman, Landau, & Pagani, 2003; Tager-Flusberg, Plesa-Skwerer, Faja, & Joseph, 2003; Karmiloff-Smith et al., 2004; Reiss, Hoffman, & Landau, 2005; Meyer-Lindenberg, Mervis, & Berman, 2006; Martens et al., 2008; O'Hearn, Courtney, Street, & Landau, 2009). Neuroanatomical studies suggest that these cognitive profiles can be explained by brain defects: preservation of the frontotemporal lobe (oral language), but parietal lobe (visuospatial skills) deficiency (Bellugi et al., 2000; Martens et al., 2008; Eisenberg, Jabbi, & Berman, 2010; Thomas, Purser, & Van Herwegen, 2012).

In this chapter, we explore the complex neuropsychological profile of individuals with WS, which goes far beyond a straightforward dissociation between language and other cognitive skills. Research on the neuropsychological profile of individuals with WS has been based on four major domains: intellectual ability, perception (auditory and visual), language and, more recently, memory and executive functions. Results suggest specificity and considerable variability in cognitive abilities (Bellugi et al., 2000; Martens et al., 2008). The purpose of this chapter is to analyze more recent studies, in order to find explanations for the heterogeneity of the neuropsychological profile in WS. From a developmental perspective, we attempt to understand the complexity and characteristics of this profile.

INTELLECTUAL ABILITY

Historically, studies have relied on the IQ of individuals with WS to characterize their cognitive profile (Howlin et al., 1998; Mervis et al., 1999; Martens et al., 2008). Research has highlighted a syndrome-specific dissociation between language skills and other cognitive abilities (Mervis & Klein-Tasman, 2000; Martens et al., 2008). Studies show that verbal IQ is higher than performance IQ among individuals with WS (Boddaert et al., 2006; Searcy et al., 2004). It should be noted that there is less heterogeneity in subtests measuring performance IQ (Searcy et al., 2004).

Studies have yielded discrepant findings on the IQ of children with WS across development. Some studies have demonstrated a decline in IQ (Gosh & Pankau, 1994), while others have found an increase of 3-17 points (Udwin, Davies, & Hosylin, 1996).

There therefore seems to be considerable variability in results on intellectual ability and, consequently, in behavioral abilities. For example, studies have shown that teenagers with WS have difficulty with Piagetian tests of number, weight and substances that children are normally able to perform successfully by the age of 8 years (Dehaene, 1997). However, while some adults with WS have difficulty with arithmetic, others are able to master basic operations such as addition, subtraction and division (Howlin et al., 1998).

More recently, research has suggested that intellectual level is relatively stable (Searcy et al., 2004), with most studies ruling out a link between intellectual ability and the development of cognitive processes underlying the learning behavior of individuals with WS (Majerus et al., 2003; Menghini, Verucci, & Vicari, 2004; Steele, Scerif, Cornish, & Karmiloff-Smith, 2013). It is now acknowledged that full-scale IQ alone does not reflect the variability in cognitive skills and behaviors of people with WS. For example, although a longitudinal study by Udwin, Davies, and Howlin (1996) of 23 participants with WS (IQ = 70) revealed weaknesses that could be correlated with IQ (WAIS-R; 1981) in reading, spelling and arithmetic tasks, other results suggest a lack of connection between intellectual ability and the acquisition of reading skills (Majerus et al., 2003; Menghini et al., 2004; Steele et al., 2013).

To sum up, previous findings highlight the importance of describing the cognitive skills elicited by the tasks administered to individuals with WS. In the rest of this section, we therefore describe the cognitive skills of individuals with WS, including visual and auditory perception, oral language, memory, and executive functions.

PERCEPTUAL SKILLS (VISUAL AND AUDITORY)

One way of exploring the marked heterogeneity of individuals with WS is to study their perceptual abilities, both auditory (categorical and allophonic perception) and visual (visuospatial and face processing).

Auditory processing performances seem atypical in WS (Majerus et al., 2003). The hyperacusis that is frequently mentioned in studies of WS may induce categorical and allophonic perception of speech sounds, preventing the formation of stable phonological representations (Majerus et al., 2003; Nazzi, Paterson, & Karmiloff-Smith, 2003; Eckert et al., 2006; Bogliotti, Serniclaes, Messaoud-Galusi & Sprenger-Charolles, 2008; Martens et al., 2008; Majerus D'Argembeau, Martinez Perez et al., 2010).

These auditory peculiarities could explain other cognitive and behavioral specificities (Carrasco, Castillo, Aravena, Rothhammer, & Aboitiz, 2005; Levitin et al., 2005). Surprisingly, despite their lower pain threshold and fearful reactions to specific sounds, some people with WS spontaneously exhibit a strong interest in music (Carrasco et al., 2005; Levitin et al., 2005). For example, 85% of individuals with WS display high musical creativity and high emotional reactivity to music, not to mention perfect and relative pitch (Lenhoff, 1998), compared with mental age-matched controls (Martens et al., 2008). Other studies have reported performances equivalent to those of chronological and mental age-matched controls on melodies and phrasing, but lower performances on pitch measurements, rhythm, musical interpretation and

directional tone (Deruelle, Schön, Rondan, & Mancini, 2005). These results are corroborated by neuroanatomical data indicating intact limbic brain structures (emotional functions; Reiss et al., 2000), reduced temporal and perisylvian cortical thickness, and activation of the right tonsil (music; Thompson et al., 2005). Individuals with WS therefore appear to have preserved musical ability, despite a certain modularity in this cognitive domain.

Furthermore, studies exploring visual processing have revealed impaired performances on tests featuring visuospatial components (Farran & Jarrold, 2003; Farran, Jarrold, & Gathercole, 2003; Hoffman et al., 2003; Vicari, Bellucci, & Carlesimo, 2003; Farran, 2005; Landau, Hoffman, & Kurz, 2006; Dilks, Landau, & Hoffman, 2008; O'Hearn et al., 2011). Individuals with WS perform similarly to chronological and mental age-matched controls on visual perception tasks (length estimation; Vicari, Bellugi, & Carlesimo, 2006), but less well on spatial tasks. They also perform better on visual imagery tasks than on visuospatial tasks (block tapping, copying, drawing, hierarchical forms and line orientation; Bellugi et al., 2000; Hoffman et al., 2003; Tager-Flusberg et al., 2003; Karmiloff-Smith et al., 2004; Vicari, Bellugi, & Carlesimo, 2006; Landau, 2011).

This visuospatial deficit seems to be related to the development of the spatial lexicon (Bellugi et al., 2000). A dissociation can be observed between impaired visuospatial representations and preserved language skills (Bellugi et al., 2000; Schmitt et al., 2001; Vicari et al., 2003). Despite relatively well preserved linguistic development, studies have shown that the comprehension of individuals with WS and the language (spatial prepositions) they use to describe spatial locations are below the level expected for their intellectual ability (Bellugi et al., 2000; Jarrold, Baddeley, Hewes, & Phillips, 2001; Vicari et al., 2003; Searcy et al., 2004).

In another cognitive domain, namely face processing, individuals with WS appear to have good face recognition, face discrimination, and recall of both familiar and unknown faces, despite their visuospatial deficits (Bellugi et al., 2000; Mervis, Robinson, Bertrand et al., 2000; Grice et al., 2001; Karmiloff-Smith, Brown, Grice & Paterson, 2003; Carrasco et al., 2005; Farran & Jarrold, 2003; Tager-Flusberg, Plesa-Skwerer, Faja, & Joseph, 2003; Karmiloff-Smith et al., 2004; Martens et al., 2008; O'Hearn et al., 2009). Researchers have investigated the visuoconstructive skills of individuals with WS in order to better understand the atypical cognitive processes involved in face processing (Bellugi et al., 2000). Results indicate that individuals with WS are better able to extract the characteristics of human faces than those of geometric forms (Farran, 2005; Reiss et al., 2005; Porter & Coltheart, 2006; Atkinson et al., 2006; Landau et al., 2006; Musolino & Landau, 2010). For example, individuals with WS display deficits in judgement of line orientation and localization tasks, but not in face recognition (Bellugi et al., 2000). However, these studies suggest that the presence of hair on these faces could facilitate face recognition processes in individuals with WS. It should be noted that in neuroimaging, no difference is observed in the behavior of individuals with WS performing recognition versus localization tasks (Meyer-Lindenberg, Mervis, & Berman, 2006). Research therefore points to incomplete modularization or another specific form of face processing in people with WS (Grice et al., 2001; Martens et al., 2008).

A specific form of face processing does indeed seem to take place in people with WS. Studies have revealed the use of mainly componential (local) processes in face processing/recognition, whereas controls favor global processes (Farran, 2005; Reiss et al., 2005; Atkinson et al., 2006; Landau et al., 2006; Porter & Coltheart, 2006; Musolino & Landau, 2010). A bias toward global rather than local processing of the information therefore appears to be involved in the visuospatial deficits (visuomotor integration, copy tests) observed in

individuals with WS (Bellugi et al., 2000; Farran & Jarrold, 2003; Fayasse & Thibaut, 2003; Martens et al., 2008; Majerus et al., 2010). More specifically, studies suggest difficulty alternating between the underlying global and local processing strategies (Mervis & Klein-Tasman, 2000; Meyer-Lindenberg et al., 2004; Porter & Coltheart, 2006). Similar results have been yielded by neuroimaging studies, which have revealed atrophy of the parietal (deterioration of the dorsal stream) and temporal cortices, and deterioration of the dorsal spatial stream (intraparietal sulcus) compared with the ventral visual stream (Boddaert et al., 2006; Eckert et al., 2006; Meyer-Lindenberg, Buckholz et al., 2006; Van Essen et al., 2006; Martens et al., 2008; Sarpal et al., 2008; Eisenberg et al., 2010; O'Hearn et al., 2011; Thomas et al., 2012). It is important to put these results into perspective, as there may be a local bias in tasks involving drawing when these are administered to individuals with WS (Farran & Jarrold, 2003). Furthermore, the latter's performances on low-level perceptual tasks requiring global processing are similar to those of mental age-matched controls (Mervis et al., 1999; Georgopoulos, Georgopoulos, Kuz, & Landau, 2004). From a functional point of view, there is also a deficit in the saccadic eye movements involved in visuospatial processes (Scerif, Cornish, Wilding, Driver, & Karmiloff-Smith, 2004).

ORAL LANGUAGE

The linguistic development of individuals with WS is often claimed in the literature to be largely unimpaired (Karmiloff-Smith et al., 2003; Carrasco et al., 2005). Their language skills certainly seem to be relatively protected (linguistic age: 8-15 years), compared with their other cognitive skills (Bellugi et al., 2000; Porter & Coltheart, 2005; Alloway & Gathercole, 2006; Porter & Coltheart, 2006; Brock, 2007; Martens et al., 2008; Rhodes, Riby, Fraser, & Campbell, 2011a; Rowe & Mervis, 2012). We find the same dissociation in neuroimaging, between the relative preservation of the temporal lobes (oral language) and a deterioration in the parietal and frontal lobes (visuospatial processing, attention, and executive functions; Reiss et al., 2005; Landau et al., 2006; Meyer-Lindenberg, Mervis, & Berman, 2006; Musolino & Landau, 2010). More specifically, there is a dissociation between the relatively preserved level of vocabulary and difficulties in the syntactic, morphosyntactic and lexical-semantic production domains, grammatical understanding, gender agreement, pragmatics, and oral expression among individuals with WS (Karmiloff-Smith et al., 2003; Carrasco et al., 2005). There is now a large body of published research on oral language, and it is possible to distinguish between the different language skills of individuals with WS, by looking at their phonological, syntactic and semantic levels, their pragmatic skills, and their narrative productions.

Early language development seems delayed by approximately 2 years in individuals with WS, compared with their chronological age-matched peers, but follows a similar trajectory to that of their mental age-matched peers (Bellugi et al., 2000; Laing et al., 2002; Martens et al., 2008). One explanation for this delay concerns the beginning of communication. Studies show that they make less use of gestural language (self-referencing behaviors), which is supposed to be a precursor of language development, than mental age-matched controls (Laing et al., 2002). Moreover, individuals with WS exhibit category-specific perception of speech sounds, indicating a possible deficit in early phonological processing (Karmiloff-Smith, Scerif, & Thomas, 2002; Nazzi et al., 2003).

Furthermore, the appearance of first words is delayed (at around 30-40 months) in WS (Martens et al., 2008). Mervis and Robinson (2000) noticed that 2-year-olds with WS exhibit a smaller lexical repertoire and poorer (expressive) vocabulary skills than their chronological age-matched peers. Nevertheless, the lexical store expands at around 11 years and the complexity of the sentences that are produced increases across development (Bellugi et al., 2000; Volterra, Caselli, Capirci, Tonucci, & Vicari, 2003). It should be noted that adults with WS have a higher vocabulary level than their chronological age-matched peers (animal naming tests; Bellugi et al., 2000). The basic morphosyntactic structures develop quickly among individuals with WS (Martens et al., 2008). Moreover, they exhibit metalinguistic abilities even before they have mastered grammar (Bellugi et al., 2000). However, compared with their mental age-matched counterparts, individuals with WS exhibit delays in terms of mean length of utterance (whose articles and prepositions substitution) (Levy, 2004) and the ability to classify objects (Nazzi, Gopnik, & Karmiloff-Smith, 2005). Karmiloff-Smith et al. (1997) suggested that individuals with WS initially lack the semantic information they need to integrate into syntactic processing.

Apart from lagging behind their chronological age-matched peers, individuals with WS appear to display typical development of linguistic skills (Karmiloff-Smith et al., 2003; Martens et al., 2008; Lacroix, Stojanovik, & Lukács, 2009). Research therefore demonstrates that phonology, vocabulary, syntax, morphosyntax and semantics are protected but delayed (Karmiloff-Smith et al., 2003; Alloway & Gathercole, 2006). Young people with WS nevertheless produce more atypical morphosyntactic errors (Martens et al., 2008) and take longer to resolve them (until age 14-16 years) than their mental age-matched peers do (until age 8-9 years; Bellugi et al., 2000). It should be noted that this gap between the appearance and development of language in WS could be explained by an imbalance between phonological and semantic skills, leading to fragile semantic representations (Mervis & Bertrand, 1997). Naming precedes identification in WS (Jarrold et al., 2001), but both skills are delayed with regard to those of chronological and mental age-matched children (Laing, Hulme, Grant, & Karmiloff-Smith, 2001). There therefore seem to be two dissociations in oral and semantic fluencies in WS (Temple, Almazan, & Sherwood, 2002; Vicari, Bates, Caselli, Pasqualetti et al., 2004; Stojanovik, 2006): greater ease performing semantic fluency tasks than naming ones, as mentioned earlier; and an intact computational component (form of linguistic expressions) versus an impaired lexical one (shape and meaning of parts of speech).

Beyond language development, several studies have investigated the lexical and morphological, syntactic and grammatical, semantic and pragmatic skills of individuals with WS (Brock, 2007; Martens et al., 2008). Research on lexical skills indicates that vocabulary (expressive aspect), verbal fluency, and grammatical (oral fluency, irregular form of past tense, plural, word roots/suffixes), syntactic, and semantic skills are relatively protected, albeit delayed (Bellugi et al., 2000; Thomas et al., 2001; Zukowski, 2005; Martens et al., 2008).

At first sight, individuals with WS use more sophisticated words than their chronological age-matched peers (idiomatic and stereotypical linguistic expressions), despite some errors of contextual use (Udwin & Yule, 1991). The grammatical and syntactic abilities of adults with WS seem equivalent to those of mental age-matched individuals (i.e., 5-10 years; Losh, Bellugi, & Anderson, 2001), and poorer than those of chronological age-matched ones (Karmiloff-Smith et al., 1997). Measures of grammar, syntax (grammatical understanding, morphosyntax, gender agreement), more complex semantics (many words produced but not necessarily understood) and pragmatics indicate atypical skills and delays among individuals

with WS (Bellugi et al., 2000; Temple et al., 2002; Vicari et al., 2004; Lacroix & Bernicot, 2006; Stojanovik, 2006). More specifically, just like their mental age-matched peers, they are capable of using a variety of complex grammatical forms (passive, anaphoric and relative sentences, conditional forms, irregular past tense), despite some morphosyntactic errors, during oral production (Bellugi et al., 2000; Mervis & Klein-Tasman, 2000; Karmiloff-Smith et al., 2003; Ring & Clahsen, 2005). Other studies have reported poorer performances than those of mental age-matched controls when it comes to finding complex similarities (semantic cues; Klein & Mervis, 1999; Mervis & Klein-Tasman, 2000).

The narrative productions of individuals with WS are poorer than those of mental age-matched controls on some tests (e.g., sentence-repetition task; Bellugi et al., 2000). Their expressive linguistic skills are better than their receptive ones, even if they spend more time on the pronunciation and articulation of words than their verbal ability-matched peers (around 8 years) do (Bellugi et al., 2000; Losh et al., 2001; Jarrold, Cowan, Hewes, & Riby, 2004b; Stojanovik, 2006). Furthermore, teenagers with WS display preserved affective language (expression and prosody) in their narratives (Losh et al., 2001) despite their difficulty with mutual communication (Stojanovik, 2006).

These results point to heterogeneous oral language development in WS (Temple et al., 2002; Vicari et al., 2004; Porter & Coltheart, 2005; Stojanovik, 2006; Martens et al., 2008). Furthermore, Stojanovik, Perkins, and Howard (2001) have suggested that the language skills of individuals with WS should be regarded as a relative, but not necessarily effective, cognitive strength.

MEMORY ABILITIES

The past few years have witnessed an increase in research on memory in people with WS. More specifically, since the 1990s, there has been a surge of interest in the underlying processes that are necessary for learning. In this section, we review these studies of short-term memory (including working memory) and long-term memory, which show the memory skills of individuals with WS to be fragile and unloss-making (Jarrold et al., 2004b; Martens et al., 2008; Rhodes et al., 2011a).

Early studies focused on short-term memory, neglecting the executive components of working memory. They failed to reach a consensus, mainly because of evaluation differences (Klein & Mervis, 1999; Mervis & Klein-Tasman, 2000). Some studies reported preserved phonological short-term memory in comparison with mental age-matched controls (Bellugi et al., 2000; Mervis & Klein-Tasman, 2000; Jarrold et al., 2001; Laing et al., 2005; Sampaio, Sousa, Fernandez, Henriques, & Goncalves, 2008), while others compared these performances with those of verbal ability-matched controls (Laing et al., 2001; Majerus et al., 2003; Jarrold, Baddeley, Hewes, Leeke, & Phillips, 2004a; Sampaio et al., 2008; Lacroix, Stojanovik, & Lukács, 2009). Individuals with WS appear to have greater short-term memory difficulties for recall (explicit episodic encoding) than for recognition (O'Hearn et al., 2009; Rhodes, Riby, Park, Fraser, & Campbell, 2010) leading to peculiarities in learning information (Baddeley, 2000). Other studies report poorer performances than those of chronological and mental age-matched individuals on spatial (location) as opposed to visual (recognition) short-term memory tasks (Vicari, Bellugi, & Carlesimo, 2006).

As for working memory skills, they seem fragile because of genuine difficulties in data accumulation (Jarrold et al., 2004b; O'Hearn et al., 2009; Menghini et al., 2010; Rhodes et al., 2010). Scores on working memory tasks (digit and word spans) are therefore lower than those of chronological and mental age-matched controls (Vicari, Bellugi, & Carlesimo, 2006).

Furthermore, there are executive dissociations between verbal and nonverbal working memory among individuals with WS. For example, the verbal/spatial dissociation seems to extend to the functioning of working memory. Individuals with WS display executive deficits in both verbal and spatial working memory (manipulating and updating information), associated with preserved short-term maintenance of spatial, but not verbal, information (Menghini et al., 2010; Rhodes et al., 2011a).

Other results point to specific deficits in spatial versus verbal working memory (Alloway & Gathercole, 2006; Vicari, Bellugi, & Carlesimo, 2006; Sampaio et al., 2008; Rhodes et al., 2011a; Rowe & Mervis, 2012). Neuroimaging studies have shown that a deficit in the dorsal stream (parietal regions) is behind the specific impairment of memory for spatial information (Vicari et al., 2003; Sarpal et al., 2008). These spatial specificities seem to be observed in more complex tasks involving high levels of processing and/or of the handling of a large volume of spatial information (O'Hearn et al., 2009; Menghini et al., 2010; Rhodes et al., 2010; Rhodes et al., 2011a).

Another dissociation concerns visual/visuospatial working memory. Individuals with WS exhibit relatively preserved maintenance of visual information (recognition and visual span tests), but a deficit in handling spatial information, or more specifically, connecting visual and spatial information (Corsi Block localization task; Vicari et al., 2003; Vicari, Bellugi, & Carlesimo, 2006; Jarrold, Phillips, & Baddeley, 2007). For example, it has been reported that, compared with mental age-matched individuals, individuals with WS exhibit a general localization deficit, rather than a specific one for face or house identification (Vicari et al., 2003; Meyer-Lindenberg, Mervis, & Berman, 2006; Vicari, Bellugi, & Carlesimo, 2006; Jarrold et al., 2007; Sarpal et al., 2008; Vicari O'Hearn et al., 2009). One possible explanation for the specificities of their visuospatial working memory is that they have a visuospatial sketchpad deficit that is independent of their mental ability (Munir, Cornish, & Wilding, 2000; Jarrold et al., 2007; O'Hearn et al., 2009; Menghini et al., 2010; Rhodes et al., 2010; Rhodes et al., 2011a). Research indicates a possible dissociation between the processes linked to the visuospatial sketchpad, as visual information seems protected, but not spatial information (Benton Facial Recognition Test; Vicari et al., 2003; Kittler, Krinsky-McHale, & Devenny, 2008; Sampaio et al., 2008). For example, it has been suggested that visuospatial working memory tasks may require mental rotation (Luzzatti, Vecchi, Agazzi, Cesa-Bianchi, & Vergani, 1998). Another explanation therefore concerns deficits in mental rotation, rather than visual abilities, in WS.

Numerous studies of WS have focused on one of the most homogeneous cognitive skills, namely phonological memory. Results indicate relative preservation of phonological short-term memory (Majerus et al., 2003; Laing et al., 2005; Porter & Coltheart, 2005; Sampaio et al., 2008; Lacroix, Stojanovik, & Lukács, 2009). Despite the multiplicity of tests used (digit and word span, nonword repetition) the phonological memory performances of individuals with WS are better than those of their mental age-matched peers (Danielsson, Henry, Messer, Carney, & Rönnerberg, 2016). Certain studies indicate poorer performances on phonological memory tasks with regard to a verbal age of 8 years (Jarrold et al., 2004b), contrary to other studies (Majerus et al., 2003). There is a degree of variability in short-term phonological

memory abilities (Mervis & Klein-Tasman, 2000; Jarrold et al., 2004b). One possible explanation is that the articulatory flow is slower (more time needed for subvocal repetition, planning and pausing), which requires information to be retained for longer, and therefore causes difficulties with short-term recall (Jarrold et al., 2004b; Laing et al., 2005). These articulatory specificities may reflect phonological loop deficits (Kittler et al., 2008; Sampaio et al., 2008).

These data suggest that the short-term memory deficits lead to association difficulties during the learning of new information (Baddeley, 2000). We therefore postulate that the link between short-term memory and long-term episodic learning (episodic buffer; Baddeley, 2000) is impaired. The long-term memory weakness is therefore secondary to greater deficits in short-term memory, particularly if long-term compensation strategies are used (Baddeley et al., 2000). Individuals with WS perform more poorly on long-term memory tasks (recall and recognition) than chronological age-matched controls. More specifically, they display a greater deficit on recall tasks than on recognition (Jarrold et al., 2007). For example, findings point to a visual deterioration in long-term memory for delayed recall rather than a recognition difficulty (fewer elements recalled than copied in the Rey-Osterreith Complex Figure test) in WS (Vicari et al., 1996). We also find this verbal/spatial dissociation in long-term memory performances, with the preservation of verbal rather than visuospatial information, in comparison with the performances of mental age-matched controls (Jarrold et al., 2007).

All these results demonstrate that individuals with WS have weaknesses and atypical verbal and nonverbal memory development, with regard to their mental age-matched peers (Jarrold et al., 2007; Vicari, Verucci, & Carlesimo, 2007; Sampaio et al., 2008). Furthermore, their deficits in visuospatial and phonological short-term memory may be secondary to more general problems in visuospatial and phonological processing. In short, some memory deficits seem to be more generally connected to learning difficulties (Jarrold et al., 2007; Kittler et al., 2008).

Furthermore, memory abilities must be studied in relation to executive and attentional functions, as both these areas play a fundamental role in the cognitive and behavioral development of individuals with WS.

EXECUTIVE AND ATTENTIONAL FUNCTIONS

Children with WS exhibit attentional and executive deficits that decrease in adolescence (Farran & Jarrold, 2003; Carrasco et al., 2005; Willcutt, Sonuga-Barke, Nigg, & Sergeant, 2008; Rhodes et al., 2010). Neuroimaging studies show a deficit in the frontal and parietal lobes for particular attentional tasks (Atkinson & Braddick, 2010; Rhodes et al., 2010). More specifically, the performances of individuals with WS on selective, divided and sustained attention tasks are poorer than those of their chronological age-matched peers, but equivalent to those of mental age-matched controls (Farran & Jarrold, 2003; Menghini et al., 2010; Costanzo et al., 2013; Greer, Riby, Hamilton & Riby, 2013). Individuals with WS also perform better on tasks requiring sustained rather than selective attention, compared with their mental age-matched peers (Atkinson & Braddick, 2010).

Furthermore, there seem to be verbal/visuospatial dissociations in selective and sustained attention tasks. For example, individuals with WS perform equally well on both visual sustained

and verbal selective attention tasks, but more poorly compared with chronological or mental age-matched controls (Scerif et al., 2004; Menghini et al., 2010; Costanzo et al., 2013).

More generally, people with WS have deficits in flexibility and attentional set-shifting (pointing/counter-pointing, face fixation) that extend to the visuomotor domain (Mervis, Robinson, Rowe, Becerra, & Klein-Tasman, 2003; Willcutt et al., 2008; Menghini et al., 2010; Rhodes et al., 2010; Hocking et al., 2013). More specifically, individuals with WS have impaired performances on attentional switching tasks (Rhodes et al., 2010). Analysis of these performances has revealed a dissociation whereby verbal performances are better than visuospatial ones (Carney, Brown & Henry, 2013; Menghini et al., 2010). One possible explanation is an inhibition deficit in verbal skills (Davidson, Amso, Anderson, & Diamond, 2006).

The executive function planning is also impaired in individuals with WS, compared with chronological and mental age-matched controls (Vicari, Bellugi, & Carlesimo, 2006; Willcutt et al., 2008; Menghini et al., 2010; Rhodes et al., 2010; Cowie, Braddick & Atkinson, 2012; Costanzo et al., 2013; Greer et al., 2013). These results have been confirmed by neuroimaging studies showing a deficit in the dorsal and frontoparietal circuits involved in planning, motor execution, and inhibition (Campbell et al., 2009; Faria et al., 2011). However, when individuals with WS are given more time to specify their answers, their planning performances seem relatively protected (Rhodes et al., 2010; Menghini et al., 2010). This points to cognitive slowing rather than an executive deficit in WS.

Inhibition functions are impaired in individuals with WS (Vicari, Bellugi, & Carlesimo, 2006; Jarrold et al., 2007; Porter, Coltheart, & Langdon, 2007; Kittler et al., 2008; Sampaio et al., 2008; O'Hearn et al., 2009; Menghini et al., 2010; Rhodes et al., 2010; Osório et al., 2012; Costanzo et al., 2013; Hocking et al., 2013). Individuals with WS take longer and make more errors (inhibition deficits) than chronological or mental age-matched controls (Greer et al., 2013). More specifically, individuals with WS exhibit poorer visual inhibition in the visuospatial domain (Costanzo et al., 2013). Carney, Brown, and Henry (2013) observed poorer performances compared with those of chronological and mental age-matched participants. Neuroimaging studies indicate deficits in the frontal lobe and frontostriatal networks that could account for failures in the inhibition process and thus for inappropriate social behaviors (hypersociability, social disinhibition, and emotional problems; Porter et al., 2007; Rhodes et al., 2010; Rhodes, Riby, Matthews, & Coghill, 2011b; Greer et al., 2013; Hocking et al., 2013; Little et al., 2013).

Compared with chronological age-matched controls, individuals with WS therefore display deficits in a range of executive functions (attention, memory, problem solving, planning and inhibition; Vicari, Bellugi, & Carlesimo, 2006; Jarrold et al., 2007; Porter et al., 2007; Kittler et al., 2008; Sampaio et al., 2008; Willcutt et al., 2008; O'Hearn et al., 2009; Lanfranchi, Jerman, Dal Pont, Alberti, & Vianello, 2010; Rhodes et al., 2010; Osório et al., 2012). These executive deficits may also contribute to the social phenotype (behavior and adaptive functioning), resulting in hypersociability, inattention, distractibility, low frustration threshold, anxiety, and poor social understanding (Carrasco et al., 2005; Meyer-Lindenberg, Mervis, & Berman, 2006; Martens et al., 2008; Willcutt et al., 2008; Campbell et al., 2009; Martens, Wilson, Dudgeon, & Reutens, 2009; Menghini et al., 2010; Rhodes et al., 2010; Faria et al., 2011; Rhodes et al., 2011b; Meda, Pryweller & Thornton-Wells, 2012; Greer et al., 2013). Howlin and colleagues (1998) found that adults with WS (mean age = 27 years) exhibit socio-adaptive behavior equivalent to that of 6-year-olds. Neuroimaging studies indicate differences

in the left temporal lobe (hyperactivity; Campbell et al., 2009) and a deficit in the frontal and parietal lobes (hypersociability) among children with WS (Thompson et al., 2005; Boddaert et al., 2006; Eckert et al., 2006; Campbell et al., 2009; Faria et al., 2011; Osório et al., 2012).

These results underscore the importance of examining interdomain interactions from a developmental point of view, as cognitive strengths and weaknesses beyond the domain of interest can have a considerable impact on the behavioral phenotype (Karmiloff-Smith, 2009; Karmiloff-Smith, 2012).

DISCUSSION AND PERSPECTIVES

Approximately 95% of individuals with WS display considerable learning difficulties (Paterson, Girelli, Butterworth, & Karmiloff-Smith, 2006). However, despite their intellectual deficiency and neuropsychological profile, individuals with WS appear to be capable of learning more than the basic skills of reading, spelling and mathematics at school (Howlin et al., 1998; Bellugi et al., 2000). For example, those aged between 6 and 16 years have the reading, spelling, arithmetic and social adaptation levels of 6- to 8-year-olds (Howlin et al., 1998; Bellugi et al., 2000).

None of these studies simultaneously considered all the cognitive processes underlying the unique neuropsychological profile responsible for these learning difficulties. Instead, most focused on two or three cognitive-behavioral or neuroanatomical features. In this chapter, we have therefore tried to provide an overview of the specific neuropsychological profile of individuals with WS, which might explain their learning difficulties. Although they exhibit considerable heterogeneity in the cognitive, linguistic, perceptual, memory and executive functions involved in the behavioral phenotypes (Majerus et al., 2001; Majerus et al., 2003; Porter & Coltheart, 2005), we can nevertheless identify universal cognitive characteristics

At the cognitive level, the development of cognitive functions seems typical, albeit delayed, in WS (Majerus et al., 2001; Porter & Coltheart, 2005). The study by Bellugi, Lichtenberger, Jones, Lai, and St. George (2000) was the first to establish clear dissociations in the cognitive architecture of individuals with WS. Certain skills seem relatively protected (oral language, lexical level, short-term memory, face processing) while others seem impaired (visuospatial processing, planning, inhibition, attention, long-term and working memory; Fayasse & Thibaut, 2003; Karmiloff-Smith et al., 2003; Martens et al., 2008; Menghini et al., 2010; Costanzo et al., 2013; Dessalegn et al., 2013). Several explanations have been put forward for these deficits, which include a linguistic imbalance between semantics and phonology leading to poor semantic representations (Laing et al., 2002; Nazzi et al., 2003), the use of local rather than overall processing of visuospatial information and difficulty alternating between strategies (Meyer-Lindenberg, Mervis, & Berman, 2006; Porter & Coltheart, 2006; Dessalegn et al., 2013), working memory deficits concerning the visuospatial sketchpad and/or phonological loop (Kittler et al., 2008; Sampaio et al., 2008; O'Hearn et al., 2009; Menghini et al., 2010; Rhodes et al., 2010; Rhodes et al., 2011a), and deficits in the dorsal frontoparietal circuit involved in inhibition, planning and social behavior (Campbell et al., 2009; Rhodes et al., 2011b; Faria, Landau, O'Hearn, et al., 2012; Hocking et al., 2013; Greer et al., 2013; Little et al., 2013). However, the results of some studies suggest that these dissociative profiles are not perceptible in all individuals with WS (Porter & Coltheart, 2005; Vicari, Bellugi, &

Carlesimo, 2006; Sampaio et al., 2008; Rhodes et al., 2011a). This heterogeneity of results may stem from experimental artefacts (use of different tools), but may also reveal the existence of developmental subprofiles.

At the experimental level, there seem to be disparities in results depending on the test, control group, age range, size and developmental specificities of the sample (Vicari, Bellugi, & Carlesimo, 2006; Jarrold et al., 2007; Van der Molen, Van Luit, Jongmans, & Van der Molen, 2007; Martens et al., 2008; Sampaio et al., 2008; Schuchardt, Mäehler, & Hasselhorn, 2011; Greer et al., 2013). For example, floor and ceiling effects have been found for visuospatial processing, depending on the choice of test and type of control (e.g., mental, chronological, verbal or vocabulary age-matched, other pathologies; Farran & Jarrold, 2003). Some studies that did not include a control group based on verbal ability have reported good performances on verbal memory (Sampaio et al., 2008), in contrast to other studies (Jarrold et al., 2004a). Furthermore, 68% of the studies of language or visuospatial skills included samples that varied from 1 to 54 individuals with WS (Martens et al., 2008).

At the behavioral level, a direct window on the initial state (Karmiloff-Smith, 1998), individuals with WS exhibit hypersociability, hypersensitivity, anxiety, specific phobias and socially inappropriate language (Martens et al., 2008; Carrasco et al., 2005). These behavioral characteristics seem to persist with age (Dykens, 2003; Rosner, Hodapp, Fidler, Sagun, & Dykens, 2004). However, this behavioral change in individuals with WS seems to take place later than it does in controls (see research on depression: Meyer-Lindenberg, Mervis, & Berman, 2006). It should be noted that most behavioral research has used two types of assessment: self-report questionnaire and/or parental questionnaire. It is thus advisable to temper results on hyperactivity, anxiety and attentional difficulties among individuals with WS and/or parental fears (Dykens, 2003) owing to potential biases.

At the neuroanatomical and functional levels, results clarify the specificities of the cognitive profile. Limbic structures (hypersensitivity) are preserved (Reiss et al., 2000), as is the lower temporal lobe and the activation of the right tonsil (language and music) (Thompson et al., 2005), but there are deficits in the intraparietal/occipitoparietal sulcus (visuospatial deficits and hypersociability), and reduced connectivity between the tonsil and orbitofrontal cortex (hypersociability and anxiety) (Van Essen et al., 2006; Marenco et al., 2007; Dilks et al., 2008; Sarpal et al., 2008; O'Hearn et al., 2009; Faria et al., 2011).

All these results therefore suggest that there is a specific and heterogeneous cognitive pattern in WS. This cognitive specificity seems stable across development (Karmiloff-Smith & al., 2002; Dykens, 2003; Rosner et al., 2004). Moreover, the numerous cognitive dissociations may explain this considerable heterogeneity. It is therefore important to understand the trajectories of developmental disorders better by favoring longitudinal studies (Karmiloff-Smith et al., 2002). We dispute the usefulness of nativist hypotheses or modular continuity for estimating cognitive profiles from adult models, that is, without studying developmental trajectories (Karmiloff-Smith et al., 2002). For example, comparisons of individuals with WS versus Down syndrome (DS) reveal different cognitive profiles in adulthood, despite resemblances during development (Klein & Mervis, 1999).

Intersyndrome comparisons can improve current understanding of the cognitive specificities of individuals with WS. For example, studies indicate similar IQ scores across WS and DS (Klein & Mervis, 1999; Mervis & Klein-Tasman, 2000), and show that good face processing skills are not specific to individuals with WS. They report similar abilities to distinguish between faces across WS and other types of intellectual deficiency or learning

disorder (Tager-Flusberg & Sullivan, 2000; Martens et al., 2008). By contrast, good verbal abilities and verbal working memory seem to be specific to WS. Several studies indicate better verbal skills (vocabulary and grammar) in WS compared with DS for the same vocabularylexicon age?? and mental ages (Bellugi et al., 2000) or children with a mental retardation of mixed etiology or a learning disorder (Udwin & Yule, 1991). Moreover, verbal working memory skills (digit and word spans) are better in WS than in DS (Klein & Mervis, 1999; Edgin, Pennington, & Mervis, 2010) but equivalent in disorders of mixed etiology (Devenny, Krinsky-McHale, Kittler, et al., 2004). It should be noted that we find this dissociation between the verbal and visuospatial domains. data indicate lower spatial associative and short-term memory performances in WS than in DS (Klein & Mervis, 1999). Finally, we can put the visuo-spatial deficits displayed by individuals with WS into perspective: performances on visuospatial tasks are better in WS than in learning disorders (Udwin & Yule, 1991) or DS (Klein & Mervis, 1999).

Studying atypical populations such as individuals with WS allows us to gain a better understanding of typical learning development. If we are to provide appropriate treatment, it is vital that we identify the cognitive specificities of individuals with WS and their developmental trajectory. Questions about the adaptability of tests to individuals with intellectual deficiency, the processes that are studied, and the generalization of the results require particular attention. For instance, results can only be generalized if the sample size is sufficiently large and there is a reasonable age range between the participants. It is also crucial to take account of the trajectory and developmental specificities of samples with neurodevelopmental disorders, such as participants with WS (Karmiloff-Smith, 1998). Longitudinal studies of individuals with WS are therefore needed in order to carry out a more thorough examination of their relatively typical cognitive abilities, albeit delayed with regard to controls (Karmiloff-Smith et al., 2002; Martens et al., 2008).

REFERENCES

- Alloway, T. P., Gathercole, S. E. and Pickering, S. J. (2006). Verbal and visuospatial short-term and working memory in children: Are they separable? *Child development*, 77 (6): 1698-1716.
- Atkinson, J. and Braddick, O. (2010). From genes to brain development to phenotypic behavior: "Dorsal-stream vulnerability" in relation to spatial cognition, attention, and planning of actions in Williams syndrome (WS) and other developmental disorders. *Progress in brain research*, 189: 261-283.
- Atkinson, J., Braddick, O., Rose, F. E., Searcy, Y. M., Wattam-Bell, J. and Bellugi, U. (2006). Dorsal-stream motion processing deficits persist into adulthood in Williams syndrome. *Neuropsychologia*, 44 (5): 828-833.
- Baddeley, A. D. (2000). Short-term and working memory. In E. Tulving and F. I. M. Craik (Eds.), *The Oxford handbook of memory* (pp. 77-92). New York: Oxford University Press.
- Bellugi, U., Lichtenberger, L., Jones, W., Lai, Z. and George, M. S. (2000). I. The neurocognitive profile of Williams Syndrome: A complex pattern of strengths and weaknesses. *Journal of cognitive neuroscience*, 12 (Supplement 1): 7-29.

- Boddaert, N., Mochel, F., Meresse, I., Seidenwurm, D., Cachia, A., Brunelle, F., ... Zilbovicius, M. (2006). Parieto-occipital grey matter abnormalities in children with Williams syndrome. *NeuroImage*, 30 (3): 721-725.
- Bogliotti, C., Serniclaes, W., Messaoud-Galusi, S. and Sprenger-Charolles, L. (2008). Discrimination of speech sounds by children with dyslexia: Comparisons with chronological age and reading level controls. *Journal of experimental child Psychology*, 101 (2): 137-155.
- Brock, J. (2007). Language abilities in Williams syndrome: A critical review. *Development and psychopathology*, 19 (01): 97-127.
- Campbell, L. E., Daly, E., Toal, F., Stevens, A., Azuma, R., Karmiloff-Smith, A. ... Murphy, K. C. (2009). Brain structural differences associated with the behavioural phenotype in children with Williams syndrome. *Brain Research*, 1258: 96-107.
- Carney, D. P. J., Brown, J. H. and Henry, L. A. (2013). Executive function in Williams and Down syndromes. *Research in developmental disabilities*, 34 (1): 46-55.
- Carrasco, X., Castillo, S., Aravena, T., Rothhammer, P. and Aboitiz, F. (2005). Williams syndrome: Pediatric, neurologic, and cognitive development. *Pediatric neurology*, 32 (3): 166-172.
- Cashon, C. H., Ha, O. R., Allen, C.L., Graf, K. M., Saffran, J. R. and Mervis, C. B. (2009). 9- to 20-month-olds with Williams syndrome are linguistic statistical learners. Paper presented at the biennial meeting of the Society for Research in Child Development, Denver, CO.
- Costanzo, F., Varuzza, C., Menghini, D., Addona, F., Gianesini, T. and Vicari, S. (2013). Executive functions in intellectual disabilities: A comparison between Williams syndrome and Down syndrome. *Research in developmental disabilities*, 34 (5): 1770-1780.
- Cowie, D., Braddick, O. and Atkinson, J. (2012). Visually guided step descent in children with Williams syndrome. *Developmental science*, 15 (1): 74-86.
- Danielsson, H., Henry, L., Messer, D., Carney, D. P. and Rönnerberg, J. (2016). Developmental delays in phonological recoding among children and adolescents with Down syndrome and Williams syndrome. *Research in developmental disabilities*, 55: 64-76.
- Davidson, M. C., Amso, D., Anderson, L. C. and Diamond, A. (2006). Development of cognitive control and executive functions from 4 to 13 years: Evidence from manipulations of memory, inhibition, and task switching. *Neuropsychologia*, 44 (11): 2037-2078.
- Dehaene, S. (1997). *The number sense: How the mind creates mathematics*. New York: Oxford University Press.
- Deruelle, C., Schön, D., Rondan, C. and Mancini, J. (2005). Global and local music perception in children with Williams syndrome. *NeuroReport*, 16 (6): 631-634.
- Dessalegn, B., Landau, B. and Rapp, B. (2013). Consequences of severe visual-spatial deficits for reading acquisition: Evidence from Williams syndrome. *Neurocase*, 19 (4): 328-347.
- Devenny, D. A., Krinsky-McHale, S. J., Kittler, P. M., Flory, M., Jenkins, E. and Brown, W. T. (2004). Age-associated memory changes in adults with Williams syndrome. *Developmental neuropsychology*, 26 (3): 691-706.
- Dilks, D., Landau, B. and Hoffman, J. E. (2008). Vision for perception and vision for action: Normal and unusual development. *Developmental Science*, 11 (4): 474-486.
- Dykens, E. M. (2003). Anxiety, fears, and phobias in persons with Williams syndrome. *Developmental neuropsychology*, 23 (1-2): 291-316.

- Eckert, M. A., Galaburda, A. M., Karchemskiy, A., Liang, A., Thompson, P., Dutton, R. A. ... Rose, F. E. (2006). Anomalous sylvian fissure morphology in Williams syndrome. *NeuroImage*, 33 (1): 39-45.
- Edgin, J. O., Pennington, B. F. and Mervis, C. B. (2010). Neuropsychological components of intellectual disability: The contributions of immediate, working, and associative memory. *Journal of Intellectual Disability Research*, 54 (5): 406-417.
- Eisenberg, D. P., Jabbi, M. and Berman, K. F. (2010). Bridging the gene-behavior divide through neuroimaging deletion syndromes: Velocardiofacial (22q11. 2 Deletion) and Williams (7q11. 23 Deletion) syndromes. *NeuroImage*, 53 (3): 857-869.
- Faria, A., Landau, B., O'Hearn, K. M., Li, X., Jiang, H., Oishi, K., Zhang, J., Miller, M. I. and Mori, S. (2011). Quantitative analysis of gray and white matter in Williams syndrome. Unpublished manuscript, Johns Hopkins University.
- Farran, E. K. (2005). Perceptual grouping ability in Williams syndrome: Evidence for deviant patterns of performance. *Neuropsychologia*, 43 (5): 815-822.
- Farran, E. K. and Jarrold, C. (2003). Visuospatial cognition in Williams syndrome: Reviewing and accounting for the strengths and weaknesses in performance. *Developmental neuropsychology*, 23 (1-2): 173-200.
- Farran, E. K., Jarrold, C. and Gathercole, S. E. (2003). Divided attention, selective attention and drawing: Processing preferences in Williams syndrome are dependent on the task administered. *Neuropsychologia*, 41 (6): 676-687.
- Fayasse, M. and Thibaut, J. P. (2003). Visuo-construction impairments in Williams syndrome. *L'année psychologique*, 103 (4): 695-727.
- Georgopoulos, M. A., Georgopoulos, A. P., Kuz, N. and Landau, B. (2004). Figure copying in Williams syndrome and normal subjects. *Experimental brain research*, 157 (2): 137-146.
- Gosh, A. and Pankau, R. (1994). Social-emotional and behavioral adjustment in children with Williams-Beuren syndrome. *American journal of medical genetics*, 53: 335-339.
- Greer, J., Riby, D. M., Hamilton, C. and Riby, L. M. (2013). Attentional lapse and inhibition control in adults with Williams Syndrome. *Research in developmental disabilities*, 34 (11): 4170-4177.
- Grice, S. J., Spratling, M. W., Karmiloff-Smith, A., Halit, H., Csibra, G., de Haan, M. and Johnson, M. H. (2001). Disordered visual processing and oscillatory brain activity in autism and Williams syndrome. *NeuroReport*, 12 (12): 2697-2700.
- Hocking, D. R., Thomas, D., Menant, J. C., Porter, M. A., Smith, S., Lord, S. R. and Cornish, K. M. (2013). The interplay between executive control and motor functioning in Williams syndrome. *Developmental science*, 16 (3): 428-442.
- Hoffman, J. E., Landau, B. and Pagani, J. (2003). Spatial breakdown in spatial construction: Evidence from eye movements in children with Williams syndrome. *Cognitive Psychology*, 46: 260-301.
- Howlin, P., Davies, M., & Udwin, O. (1998). Cognitive functioning in adults with Williams syndrome. *Journal of child psychology and psychiatry*, 39 (2): 183-189.
- Jarrold, C., Baddeley, A. D., Hewes, A. K., Leeke, T. C. and Phillips, C. E. (2004a). What links verbal short-term memory performance and vocabulary level? Evidence of changing relationships among individuals with learning disability. *Journal of memory and language*, 50 (2): 134-148.

- Jarrold, C., Baddeley, A. D., Hewes, A. K. and Phillips, C. (2001). A longitudinal assessment of diverging verbal and non-verbal abilities in the Williams syndrome phenotype. *Cortex*, 37 (3): 423-431.
- Jarrold, C., Baddeley, A. and Phillips, C. (2007). Long-term memory for verbal and visual information in Down syndrome and Williams syndrome: Performance on the Doors and People test. *Cortex*, 43(2): 233-247.
- Jarrold, C., Cowan, N., Hewes, A. K. and Riby, D. M. (2004b). Speech timing and verbal short-term memory: Evidence for contrasting deficits in Down syndrome and Williams syndrome. *Journal of memory and language*, 51 (3): 365–380.
- Jones, W., Bellugi, U., Lai, Z., Chiles, M., Reilly, J., Lincoln, A. and Adolphs, R. (2000). II. Hypersociability in Williams syndrome. *Journal of cognitive neuroscience*, 12 (Supplement 1): 30-46.
- Karmiloff-Smith, A. (1998). Development itself is the key to understanding developmental disorders. *Trends in cognitive sciences*, 2 (10): 389-398.
- Karmiloff-Smith, A. (2009). Nativism versus neuroconstructivism: Rethinking the study of developmental disorders. *Developmental psychology*, 45 (1): 56.
- Karmiloff-Smith, A. (2012). Brain: The neuroconstructivist approach. In E. K. Farran and A. Karmiloff-Smith (Eds.), *Neurodevelopmental disorders across the lifespan* (pp. 37-58). New York: Oxford University Press.
- Karmiloff-Smith, A., Brown, J. H., Grice, S. and Paterson, S. (2003). Dethroning the myth: Cognitive dissociations and innate modularity in Williams syndrome. *Developmental Neuropsychology*, 23 (1-2): 227–242.
- KarmiloffSmith, A., Grant, J., Berthoud, I., Davies, M., Howlin, P. and Udwin, O. (1997). Language and Williams syndrome: How intact is “intact”? *Child development*, 68 (2): 246-262.
- KarmiloffSmith, A., Scerif, G. and Thomas, M. (2002). Different approaches to relating genotype to phenotype in developmental disorders. *Developmental psychobiology*, 40 (3): 311-322.
- Karmiloff-Smith, A., Thomas, M., Annaz, D., Humphreys, K., Ewing, S., Brace, N. ... Campbell, R. (2004). Exploring the Williams syndrome face-processing debate: The importance of building developmental trajectories. *Journal of child psychology and psychiatry*, 45 (7): 1258-1274.
- Kittler, P. M., Krinsky-McHale, S. J. and Devenny, D. A. (2008). Dual-task processing as a measure of executive function: A comparison between adults with Williams and Down syndromes. *American journal on mental retardation*, 113 (2): 117-132.
- Klein, B. P. and Mervis, C. B. (1999). Contrasting patterns of cognitive abilities of 9- and 10-year-olds with Williams syndrome or Down syndrome. *Developmental neuropsychology*, 16 (2): 177-196.
- Kuhl, P. (2004). Early language acquisition: Cracking the speech code. *Nature Reviews Neuroscience*, 5: 831–843.
- Lacroix, A. and Bernicot, J. (2006). *French-speaking children and adolescents with Williams syndrome: Pragmatics, language and sociability*. Paper presented at the 3rd Williams syndrome workshop, Reading, UK, 6-7 July.
- Lacroix, A., Stojanovik, V. and Lukacs, A. (2009). What do we learn on language acquisition from Williams’ syndrome? In [editors], *Language and linguistics: Emerging trends* (pp. 177-192). Lieu: Nova Publishers.

- Laing, E., Butterworth, G., Ansari, D., Gsödl, M., Longhi, E., Panagiotaki, G. ... Karmiloff-Smith, A. (2002). Atypical development of language and social communication in toddlers with Williams syndrome. *Developmental science*, 5 (2): 233-246.
- Laing, E., Hulme, C., Grant, J. and Karmiloff-Smith, A. (2001). Learning to read in Williams syndrome: Looking beneath the surface of atypical reading development. *Journal of Child Psychology and Psychiatry*, 42 (6): 729-739.
- Laing, E., Grant, J., Thomas, M., Parmigiani, C., Ewing, S., & Karmiloff-Smith, A. (2005). Love is... an abstract word: The influence of lexical semantics on verbal short-term memory in Williams syndrome. *Cortex*, 41 (2): 169-179.
- Landau, B. (2011). The organization and development of spatial cognition: Insights from Williams syndrome. In J. Burack, R. M. Hodapp, G. Iarocci and E. Zigler (Eds.), *The Oxford handbook of intellectual disabilities and development*, 2nd Ed. New York: Oxford University Press.
- Landau, B., Hoffman, J. and Kurz, N. (2006). Object recognition with severe spatial deficits in Williams syndrome: Sparing and breakdown. *Cognition*, 100 (3): 483-510.
- Lanfranchi, S., Jerman, O., Dal Pont, E., Alberti, A. and Vianello, R. (2010). Executive function in adolescents with Down syndrome. *Journal of intellectual disability research*, 54 (4): 308-319.
- Levitin, D. J., Cole, K., Lincoln, A. and Bellugi, U. (2005). Aversion, awareness, and attraction: Investigating claims of hyperacusis in the Williams syndrome phenotype. *Journal of child psychology and psychiatry*, 46 (5): 514-523.
- Levitin, D. J., Menon, V., Schmitt, J. E., Eliez, S., White, C. D., Glover, G. H. ... Reiss, A. L. (2003). Neural correlates of auditory perception in Williams syndrome: An fMRI study. *NeuroImage*, 18 (1): 74-82.
- Levy, Y. (2004). A longitudinal study of language development in two children with Williams syndrome. *Journal of child language*, 31 (02): 287-310.
- Little, K., Riby, D. M., Janes, E., Clark, F., Fleck, R. and Rodgers, J. (2013). Heterogeneity of social approach behaviour in Williams syndrome: The role of response inhibition. *Research in developmental disabilities*, 34 (3): 959-967.
- Losh, M., Bellugi, U. and Anderson, J. D. (2001). Narrative as a social engagement tool: The excessive use of evaluation in narratives from children with Williams syndrome. *Narrative inquiry*, 10 (2): 265-290.
- Luzzatti, C., Vecchi, T., Agazzi, D., Cesa-Bianchi, M. and Vergani, C. (1998). A neurological dissociation between preserved visual and impaired spatial processing in mental imagery. *Cortex*, 34 (3): 461-469.
- Majerus, S., Barisnikov, K., Vuillemin, I., Poncelet, M. and Linden, M. V. D. (2003). An investigation of verbal short-term memory and phonological processing in four children with Williams syndrome. *Neurocase*, 9 (5): 390-401.
- Majerus, S., D'Argembeau, A., Martinez Perez, T., Belayachi, S., Van der Linden, M., Collette, F., Salmon, E., Seurinck, R., Fias, W. and Maquet, P. (2010). The commonality of neural networks for verbal and visual short-term memory. *Journal of cognitive neuroscience*, 22: 2570-2593.
- Majerus, S., Palmisano, I., Van Der Linden, M., Barisnikov, K. and Poncelet, M. (2001). An investigation of phonological processing in Williams syndrome [Abstract]. *Twenty-ninth annual INS meeting abstracts*, 7: 153.

- Marler, J. A., Elfenbein, J. L., Ryals, B. M., Urban, Z. and Netzloff, M. L. (2005). Sensorineural hearing loss in children and adults with Williams syndrome. *American journal of medical genetics part A*, 138 (4): 318-327.
- Marenco, S., Siuta, M. A., Kippenhan, J. S., Grodofsky, S., Chang, W., Kohn, P., ... Meyer-Lindenberg, A. (2007). Genetic contributions to white matter architecture revealed by diffusion tensor imaging in Williams syndrome. *Proceedings of the National Academy of Sciences*, 104 (38): 15117-15122.
- Martens, M. A., Wilson, S. J., Dudgeon, P. and Reutens, D. C. (2009). Approachability and the amygdala: Insights from Williams syndrome. *Neuropsychologia*, 47 (12): 2446-2453.
- Martens, M. A., Wilson, S. J., & Reutens, D. C. (2008). Williams syndrome: A critical review of the cognitive, behavioral, and neuroanatomical phenotype. *Journal of child psychology and psychiatry*, 49 (6): 576-608.
- Meda, S. A., Pryweller, J. R. and Thornton-Wells, T. A. (2012). Regional brain differences in cortical thickness, surface area and subcortical volume in individuals with Williams syndrome. *PLoS One*, 7 (2): e31913.
- Menghini, D., Addona, F., Costanzo, F. and Vicari, S. (2010). Executive functions in individuals with Williams syndrome: Executive functions in Williams syndrome. *Journal of intellectual disability research*, 54 (5): 418-432.
- Menghini, D., Verucci, L. and Vicari, S. (2004). Reading and phonological awareness in Williams syndrome. *Neuropsychology*, 18 (1): 29.
- Mervis, C. B. and Bertrand, J. (1997). Developmental relations between cognition and language: Evidence from Williams syndrome. In: Adamson LB, Ronski MA, editors. *Communication and language acquisition: Discoveries from atypical development*. New York: Brookes; 1997. pp. 75-106.
- Mervis, C. B. Developmental relations between cognition and language: Evidence Klein-Tasman, B. P. (2000). Williams syndrome: Cognition, personality, and adaptive behavior. *Mental retardation and developmental disabilities research reviews*, 6 (2): 148-158.
- Mervis, C. B., Morris, C. A., Bertrand, J. and Robinson, B. F. (1999). Williams syndrome: Findings from an integrated program of research. In Tager-Flusberg H. (ed), *Neurodevelopmental disorders: Developmental cognitive neuroscience* (pp.65-110). Cambridge, MA: MIT Press.
- Mervis, C. B., Morris, C. A., Klein-Tasman, B. P., Bertrand, J., Kwitny, S., Appelbaum, L. G. and Rice, C. E. (2003). Attentional characteristics of infants and toddlers with Williams syndrome during triadic interactions. *Developmental neuropsychology*, 23 (1-2): 243-268.
- Mervis, C. B. and Robinson, B. F. (2000). Expressive vocabulary ability of toddlers with Williams syndrome or Down syndrome: A comparison. *Developmental neuropsychology*, 17 (1): 111-126.
- Mervis, C. B., Robinson, B. F., Bertrand, J., Morris, C. A., Klein-Tasman, B. P. and Armstrong, S. C. (2000). The Williams syndrome cognitive profile. *Brain and cognition*, 44 (3): 604-628.
- Mervis, C. B., Robinson, B. F., Rowe, M. L., Becerra, A. M. and Klein-Tasman, B. P. (2003). Language abilities of individuals with Williams syndrome. *International review of research in mental retardation*, 27: 35-81.
- Meyer-Lindenberg, A., Buckholtz, J. W., Kolachana, B., Hariri, A. R., Pezawas, L., Blasi, G., ... Egan, M. (2006). Neural mechanisms of genetic risk for impulsivity and violence in humans. *Proceedings of the National Academy of Sciences*, 103 (16): 6269-6274.

- Meyer-Lindenberg, A., Mervis, C. B. and Berman, K. F. (2006). Neural mechanisms in Williams syndrome: a unique window to genetic influences on cognition and behaviour. *Nature reviews neuroscience*, 7 (5): 380-393.
- Munir, F., Cornish, K. M. and Wilding, J. (2000). A neuropsychological profile of attention deficits in young males with fragile X syndrome. *Neuropsychologia*, 38 (9): 1261-1270.
- Musolino, J. and Landau, B. (2010). When theories don't compete: Response to Thomas, Karaminis, and Knowland's commentary on Musolino, Chunyo, and Landau. *Language learning and development*, 6 (2): 170-178.
- Nazzi, T., Gopnik, A. and Karmiloff-smith, A. (2005). Asynchrony in the cognitive and lexical development of young children with Williams syndrome. *Journal of child language*, 32 (02): 427-438.
- Nazzi, T., Paterson, S. and Karmiloff-Smith, A. (2003). Early word segmentation by infants and toddlers with Williams syndrome. *Infancy*, 4 (2): 251-271.
- Newman, R., Ratner, N. B., Jusczyk, A. M., Jusczyk, P. W. and Dow, K. A. (2006). Infants' early ability to segment the conversational speech signal predicts later language development: A retrospective analysis. *Developmental psychology*, 42 (4): 643.
- O'Hearn, K., Courtney, S., Street, W. and Landau, B. (2009). Working memory impairment in people with Williams syndrome: Effects of delay, task and stimuli. *Brain and cognition*, 69 (3): 495-503.
- O'Hearn, K., Landau, B. and Hoffman, J. E. (2005). Multiple object tracking in people with Williams syndrome and in normally developing children. *Psychological science*, 16 (11): 905-912.
- O'Hearn, K., Roth, J. K., Courtney, S. M., Luna, B., Street, W., Terwillinger, R. and Landau, B. (2011). Object recognition in Williams syndrome: Uneven ventral stream activation. *Developmental science*, 14 (3): 549-565.
- Osório, A., Cruz, R., Sampaio, A., Garayzábal, E., Martínez-Regueiro, R., Gonçalves, Ó. F. ... Fernández-Prieto, M. (2012). How executive functions are related to intelligence in Williams syndrome. *Research in developmental disabilities*, 33 (4): 1169-1175.
- Paterson, S. J., Girelli, L., Butterworth, B. and Karmiloff-Smith, A. (2006). Are numerical impairments syndrome specific? Evidence from Williams syndrome and Down's syndrome. *Journal of child psychology and psychiatry*, 47 (2): 190-204.
- Porter, M. and Coltheart, M. (2005). Cognitive heterogeneity in Williams syndrome. *Developmental neuropsychology*, 27 (2): 275-306.
- Porter, M. A. and Coltheart, M. (2006). Global and local processing in Williams syndrome, autism, and Down syndrome: Perception, attention, and construction. *Developmental neuropsychology*, 30 (3): 771-789.
- Porter, M. A., Coltheart, M. and Langdon, R. (2007). The neuropsychological basis of hypersociability in Williams and Down syndrome. *Neuropsychologia*, 45 (12): 2839-2849.
- Reiss, A. L., Eliez, S., Schmitt, J. E., Straus, E., Lai, Z., Jones, W. and Bellugi, U. (2000). IV. Neuroanatomy of Williams syndrome: A high-resolution MRI study. *Journal of cognitive neuroscience*, 12 (Supplement 1): 65-73.
- Reiss, J. E., Hoffman, J. E. and Landau, B. (2005). Motion processing specialization in Williams syndrome. *Vision research*, 45 (27): 3379-3390.
- Rhodes, S. M., Riby, D. M., Fraser, E. and Campbell, L. E. (2011a). The extent of working memory deficits associated with Williams syndrome: Exploration of verbal and spatial domains and executively controlled processes. *Brain and cognition*, 77 (2): 208-214.

- Rhodes, S. M., Riby, D. M., Matthews, K. and Coghill, D. R. (2011b). Attention-deficit/hyperactivity disorder and Williams syndrome: Shared behavioral and neuropsychological profiles. *Journal of clinical and experimental neuropsychology*, 33 (1): 147–156.
- Rhodes, S. M., Riby, D. M., Park, J., Fraser, E. and Campbell, L. E. (2010). Executive neuropsychological functioning in individuals with Williams syndrome. *Neuropsychologia*, 48 (5): 1216–1226.
- Ring, M. and Clahsen, H. (2005). Distinct patterns of language impairment in Down's syndrome and Williams syndrome: The case of syntactic chains. *Journal of neurolinguistics*, 18 (6): 479–501.
- Rosner, B. A., Hodapp, R. M., Fidler, D. J., Sagun, J. N. and Dykens, E. M. (2004). Social competence in persons with Prader-Willi, Williams and Down's Syndromes. *Journal of applied research in intellectual disabilities*, 17 (3): 209–217.
- Rossen, M., Klima, E. S., Bellugi, U., Bihrl, A. and Jones, W. (1996). Interaction between language and cognition: Evidence from Williams syndrome. In J. H. Beitchman, N. Cohen, M. Konstantareas and R. Tannock (Eds.), *Language, learning, and behavior disorders: Developmental, biological, and clinical perspectives* (pp. 367–392). New York: Cambridge University Press.
- Rowe, M. L. and Mervis, C. B. (2012). 12 Working memory in Williams syndrome 1. *Working memory and neurodevelopmental disorders*, 267.
- Sampaio, A., Sousa, N., Fernández, M., Henriques, M. and Gonçalves, O. F. (2008). Memory abilities in Williams syndrome: Dissociation or developmental delay hypothesis? *Brain and cognition*, 66 (3): 290–297.
- Sarpal, D., Buchsbaum, B. R., Kohn, P. D., Kippenhan, J. S., Mervis, C. B., Morris, C. A. ... Berman, K. F. (2008). A genetic model for understanding higher order visual processing: Functional interactions of the ventral visual stream in Williams syndrome. *Cerebral cortex*, 18 (10): 2402–2409.
- Scerif, G., Cornish, K., Wilding, J., Driver, J. and Karmiloff-Smith, A. (2004). Visual search in typically developing toddlers and toddlers with Fragile X or Williams syndrome. *Developmental science*, 7 (1): 116–130.
- Scherer, S.W. and Osborne, L.R. (2006). Williams-Beuren syndrome. In P. T. Stankiewicz and J. R. Lupski (Eds.), *Genomic disorders: The genomic basis of disease* (pp. 221–236). Totowa, NJ: Humana Press.
- Schmitt, J. E., Eliez, S., Warsofsky, I. S., Bellugi, U. and Reiss, A. L. (2001). Corpus callosum morphology of Williams syndrome: Relation to genetics and behavior. *Developmental medicine & child neurology*, 43 (3): 155–159.
- Schuchardt, K., Mäehler, C. and Hasselhorn, M. (2011). Functional deficits in phonological working memory in children with intellectual disabilities. *Research in developmental disabilities*, 32 (5): 1934–1940.
- Searcy, Y. M., Lincoln, A. J., Rose, F. E., Klima, E. S., Bavar, N. and Korenberg, J. R. (2004). The relationship between age and IQ in adults with Williams syndrome. *American journal on mental retardation*, 109 (3): 231–236.
- Serniclaes, W., Ventura, P., Morais, J. and Kolinsky, R. (2005). Categorical perception of speech sounds in illiterate adults. *Cognition*, 98 (2): B35–B44.

- Steele, A., Scerif, G., Cornish, K. and Karmiloff-Smith, A. (2013). Learning to read in Williams syndrome and Down syndrome: Syndrome-specific precursors and developmental trajectories. *Journal of child psychology and psychiatry*, 54 (7): 754-762.
- Stevens, K. N. (2005). Features in speech perception and lexical access. In D. Pisoni and E. Remez (Eds.), *The handbook of speech perception* (pp. 125–155). Oxford: Blackwell.
- Stojanovik, V. (2006). Social interaction deficits and conversational inadequacy in Williams syndrome. *Journal of neurolinguistics*, 19 (2): 157-173.
- Stojanovik, V., Perkins, M. and Howard, S. (2001). Language and conversational abilities in Williams syndrome: How good is good? *International journal of language & communication disorders*, 36 (sup1): 234-239.
- Tager-Flusberg, H. and Sullivan, K. (2000). A componential view of theory of mind: Evidence from Williams syndrome. *Cognition*, 76 (1): 59-90.
- Temple, C. M., Almazan, M. and Sherwood, S. (2002). Lexical skills in Williams syndrome: A cognitive neuropsychological analysis. *Journal of neurolinguistics*, 15: 463-495.
- Thomas, M. S. C., Purser, H. R. and Van Herweger, J. (2012). Cognition: The developmental trajectory approach. In E. K. Farran and A. Karmiloff-Smith (Eds.), *Neurodevelopmental disorders across the lifespan* (pp. 13-36). New York: Oxford University Press.
- Thompson, P. M., Lee, A. D., Dutton, R. A., Geaga, J. A., Hayashi, K. M., Eckert, M. A. ... Toga, A. W. (2005). Abnormal cortical complexity and thickness profiles mapped in Williams syndrome. *The journal of neuroscience*, 25 (16): 4146-4158.
- Udwin, O., Davies, M. and Hosylin, P. (1996). A longitudinal study of cognitive abilities and educational attainment in Williams syndrome. *Developmental medicine & child neurology*, 38 (11): 1020-1029.
- Udwin, O. and Yule, W. (1991). A cognitive and behavioral phenotype in Williams syndrome. *Journal of clinical and experimental neuropsychology*, 13: 232–244.
- Vaissière, J. (2005). Perception of intonation. In D.B. Pisoni and R.E. Remez., *The handbook of speech perception* (pp. 236-263). Oxford: Blackwell.
- Van Essen, D. C., Dierker, D., Snyder, A. Z., Raichle, M. E., Reiss, A. L. and Korenberg, J. (2006). Symmetry of cortical folding abnormalities in Williams syndrome revealed by surface-based analyses. *Journal of neuroscience*, 26 (20): 5470–5483.
- Vicari, S., Bates, E., Caselli, M. C., Pasqualetti, P., Gagliardi, C., Tonucci, F. and Volterra, V. (2004). Neuropsychological profile of Italians with Williams syndrome: An example of a dissociation between language and cognition? *Journal of the international neuropsychological society*, 10 (06): 862-876.
- Vicari, S., Bellucci, S. and Carlesimo, G. A. (2003). Visual and spatial working memory dissociation: Evidence from Williams syndrome. *Developmental medicine & child neurology*, 45 (4): 269–273.
- Vicari, S., Bellucci, S. and Carlesimo, G. A. (2006). Evidence from two genetic syndromes for the independence of spatial and visual working memory. *Developmental medicine & child neurology*, 48 (02): 126–131.
- Vicari, S., Carlesimo, G., Brizzolara, D. and Pezzini, G. (1996). Short-term memory in children with Williams syndrome: A reduced contribution of lexical-semantic knowledge to word span. *Neuropsychologia*, 34: 919–925.
- Vicari, S., Verucci, L. and Carlesimo, G. A. (2007). Implicit memory is independent from IQ and age but not from etiology: Evidence from Down and Williams syndromes. *Journal of intellectual disability research*, 51 (12): 932-941.

- Volterra, V., Caselli, M. C., Capirci, O., Tonucci, F. and Vicari, S. (2003). Early linguistic abilities of Italian children with Williams syndrome. *Developmental neuropsychology*, 23 (1-2): 33-58.
- Willcutt, E., Sonuga-Barke, E., Nigg, J. and Sergeant, J. (2008). Recent developments in neuropsychological models of childhood psychiatric disorders. In T. Banachewski & L.A. Rohde (eds), *Biological child psychiatry: Recent trends and developments* (pp. 195-226). Basel: Karger Publishers.

Chapter 71

PREIMPLANTATION GENETIC DIAGNOSIS (PGD) FOR CHROMOSOME REARRANGEMENTS

Chun Kyu Lim

Laboratory of Reproductive Medicine, Cheil General Hospital and Women's Healthcare
Center, Dankook University College of Medicine, Seoul, South Korea

INTRODUCTION

Chromosome rearrangements are the most common genetic abnormalities in humans. Abnormal chromosomal configurations are formed among non-homologous chromosomes to allow full synapsis of homologous chromosomes at meiosis I of chromosome rearrangement carriers. These abnormal chromosomal configurations result in the production of gametes with various chromosomal complements due to malsegregation of derivative chromosomes or recombination. Most of the gametes have unbalanced chromosomal complements and only a small number of gametes have normal or balanced chromosomal complements. Carriers of balanced chromosome rearrangements are phenotypically normal but they are at an increased risk of abnormal pregnancies due to the unbalanced gametes. However, chromosome rearrangement carriers who cannot have babies or experience repeated abortions or abnormal pregnancies can have healthy babies after introduction of PGD. Balanced reciprocal translocation is the most common chromosome rearrangement. Thirty-two types of gametes can be produced in meiosis of reciprocal translocation carrier. According to the results that analyze the meiotic segregation of embryos from PGD cycles of reciprocal translocation carriers, 2:2 segregation is the main segregation mode. The meiotic segregation might be affected by the gender of carriers. The frequency of balanced embryos was not different between female and male carriers. However, the frequencies of 2:2 segregation, especially adjacent-1 segregation, 3:1 and 4:0 segregation were significantly different between female and male carriers. Robertsonian translocation is also common chromosome rearrangement in humans. Gametes with eight different chromosomal complements can be produced in Robertsonian translocation carriers. The frequency of balanced embryos was higher in male carriers than in female carriers. Carriers of complex chromosome rearrangements (CCR), very rare chromosome rearrangements, can achieve pregnancy by PGD although the number of

normal or balanced embryos was extremely low. Although it is very difficult to estimate the rate of normal or balanced embryos, the rate is estimated to be less than 10% in PGD for CCR carriers. The 3:3 segregation and chaotic segregation (meiotic segregation whose meiotic segregation cannot be defined) are prevalent segregation modes in carriers of three-way translocation. In PGD cycles of CCR carriers, cycle cancellation is very frequent due to the absence of the normal or balanced embryos. Therefore, it is important that a large number of embryos are obtained in one cycle. Occasionally, abnormalities of chromosome rearrangement-unrelated chromosomes are observed in embryos or abortuses although chromosome rearrangement-related chromosomes are normal or balanced. At present, PGD that diagnose all 24 chromosomes using array-CGH, SNP array or NGS is carried out worldwide. In those PGD cycles, the risk that embryo transfer is cancelled might be increased. However, the rate of normal or balanced embryos is unexpectedly increased and pregnancy rate is improved. Surely, PGD is the very effective assisted reproductive technique in achieving pregnancy of chromosome rearrangement carriers by preventing repeated abortions that chromosome rearrangement carriers usually experience. In the field of PGD for chromosome rearrangements, the next stage will be the development of technique that can discriminate between normal embryos and balanced embryos.

Chromosome rearrangements, such as translocation (reciprocal or Robertsonian), inversion and complex chromosome rearrangements, are the most common genetic abnormalities in humans. Chromosome rearrangements are classified as balanced or unbalanced (Nussbaum et al., 2016). Carriers of balanced chromosome rearrangements have the normal chromosomal complements but carriers of unbalanced chromosome rearrangements have additional or missing chromosomal material. Generally, the incidence of balanced chromosome rearrangements is about 0.19% of newborns (Jacobs, et al., 1974; Hamerton, et al., 1975). Carriers of balanced chromosome rearrangements are phenotypically normal because they have all the genetic materials. However, they are at an increased risk of implantation failure, repeated abortions or birth of chromosomally unbalanced offspring. Of course, not all carriers of balanced chromosome rearrangements experience the abnormal pregnancies. The abnormal pregnancies are resulted from the unbalanced gametes generated during meiosis of chromosome rearrangement carriers. Abnormal chromosomal configurations are formed to allow full synapses of homologous chromosomes during meiosis of balanced chromosome rearrangement carriers. The unbalanced gametes are produced as a results of malsegregation of these abnormal chromosomal configuration. Most of the gametes produced during meiosis of chromosome rearrangement carriers have unbalanced chromosomal complements. These unbalanced gametes result in implantation failure, repeated abortions or birth of chromosomally unbalanced offspring.

After the first successful clinical application of preimplantation genetic diagnosis (PGD) in 1990 (Handyside et al., 1990), PGD has been widely used worldwide to select normal or balanced embryos in *in vitro* fertilization and embryo transfer (IVF-ET) programs of balanced chromosome rearrangement carriers. Cleavage-stage fluorescence in situ hybridization (FISH) has been widely used to PGD for carriers of balanced chromosome rearrangements till the early 2010s (Van Assche, et al. 1999; Otani, et al. 2006; Bernicot, et al. 2012; Ko, et al. 2013) and PGD based on array comparative genomic hybridization (CGH) (Alfarawati, et al. 2011; Fiorentino, et al., 2011; Colls, et al. 2012; Huang, et al. 2013; Tan, et al. 2013) or next generation sequencing (NGS) (Fiorentino, et al. 2014a, b; Huang, et al. 2014; Wells, et al. 2014; Tan et al., 2014; Lukaszuk et al., 2015) is widely applied nowadays. The only abnormalities of

rearranged chromosomes can be diagnosed in FISH-based PGD and the abnormalities of other chromosomes cannot be diagnosed. However, abnormalities of all 24 chromosomes can be diagnosed after the application of array CGH or NGS into PGD.

EMBRYO BIOPSY

To perform PGD, one cell or a few cells have to be removed from oocytes or embryos. Cells necessary for PGD can be removed at three different stages, polar body from the oocytes and/or the zygotes, one or two blastomeres from cleavage stage embryos or trophectoderm (TE) from the blastocysts. Embryo biopsy is mainly performed at cleavage stage but trophectoderm biopsy is gradually increasing these days (Moutou et al., 2014; Cimadomo et al., 2016).

A small hole has to be made within the zona pellucid to remove polar body or blastomere(s). This zona breaching can be conducted by one of following methods; mechanical, laser-assisted or acidified Tyrode's drilling methods. Recently, laser-assisted zona breaching is most popularly used (Moutou et al., 2014). Clinical outcomes are not different among these three methods (De Vos and Van Steirteghem, 2001; Jones et al., 2006; Geber et al., 2011; Eldar-Geva et al., 2014).

Cleavage stage biopsy is performed on day 3 embryos. One or two blastomeres with one distinct nucleus are removed from embryos with at least 6 cells. Blastomere biopsy is usually conducted in $\text{Ca}^{2+}/\text{Mg}^{2+}$ -free medium in order to facilitate removal of blastomere from embryos. However, there are controversies about the effect of Ca^{2+} depletion on embryo development (Sefton et al., 1996; Dumoulin et al., 1998; Pey et al., 1998). Preimplantation embryos show high chromosomal mosaicism and some of these embryos can develop to blastocysts (Wells and Delhanty, 2000; Bielanska et al., 2002). Embryonic mosaicism cannot be identified by single-blastomere biopsy. Therefore, two-blastomere biopsy can be performed to compensate the problems caused by single-blastomere biopsy. However, embryonic development can be affected by two-blastomere biopsy because too much embryonic mass is removed from embryos (Cohen et al., 2007). ESHRE guidelines suggested that two-blastomere biopsy could be safely applied to PGD when embryos with ≥ 6 cells and $\leq 30\%$ of fragmentation (Goosens et al., 2008).

Polar body biopsy can be performed as an alternative to blastomere biopsy. In some countries, polar body biopsy is mainly performed since blastomere biopsy is legally prohibited. Zona breaching is performed with the same method as blastomere biopsy. Chromosomal abnormalities from mitotic errors or paternally derived aneuploidies cannot be detected when polar body biopsy is adopted. Nowadays the use of polar biopsy is decreasing.

Blastocyst stage preimplantation genetic screening (PGS) is emerging as the most promising approach to select euploid embryos. Several cells (five to ten cells) are removed from the blastocyst on day 5 or 6. PGD coupled with blastocyst biopsy is barely affected by biopsy operators (Capalbo et al., 2015) and embryonic mosaicism that is common in preimplantation embryos, because PGD is carried out using several cells. Therefore, the application of blastocyst biopsy is increasing and gradually replacing both polar body biopsy and blastomere biopsy. One of two different blastocyst biopsy methods is carried out currently. The first method is that a few cells herniated from blastocyst are removed using laser on day 5 or 6 through the small hole that is made in the zona pellucid on day 3 (McArthur et al., 2005).

Embryonic development might be affected by zona breaching and some of inner cell mass (ICM) may be biopsied together with trophoctoderm cells because the position that ICM is formed is unpredictable. The second method is that TE cells are removed simultaneously with zona breaching on day 5 or 6 (Capalbo et al., 2014). According to this method, only TE cells can be isolated from blastocyst and embryonic development is not disturbed since the biopsy operator can choose the site of biopsy. However, high standards of embryo culture and cryopreservation techniques have to be established for blastocyst biopsy to be carried out.

PREGNANCY HISTORY OF CHROMOSOME REARRANGEMENT CARRIERS

In general, chromosome rearrangement carriers have poor pregnancy history before PGD. They spend several years (mean of 3.2 years) to have a baby (Chang et al., 2012). And most of them experienced repeated miscarriages or live births of babies with multiple congenital anomalies due to unbalanced karyotypes. More than 90% of spontaneous pregnancies resulted in miscarriages (Munne et al., 2000a; Lim et al. 2008a; Keymolen et al., 2012; Liao et al., 2014). Most couples with chromosome rearrangements choose PGD to overcome the repeated miscarriage and have healthy babies. Actually, miscarriage rate decrease significantly when they become pregnant after PGD. Therefore, PGD can be a useful option that chromosome rearrangement carriers can choose to have healthy babies (Munne et al. 2000a; Lim et al., 2004).

PGD FOR BALANCED RECIPROCAL TRANSLOCATION

Balanced reciprocal translocation is the most common structural abnormalities of chromosomes with an incidence of about 0.16% in livebirth (Van Dyke et al., 1983). The risk that reciprocal carriers conceive a chromosomally abnormal embryo varies from 20% to 80%, depending on types of translocation, chromosomes involved in the translocation, position of the breakpoints, and gender of the carrier (Goldman and Hulten, 1993; Martin and Hulten, 1993). However, the risk of abnormal pregnancy or pregnancy loss can reduced by PGD in reciprocal translocation carriers (Munne et al., 2000a, 2002; Lim et al., 2004; Grace et al., 2006). During meiosis I of reciprocal translocation carriers, a quadrivalent is formed to make synapses between homologous chromosomes. And then, the quadrivalent segregate according to one of five segregation modes; 2:2 segregation (alternate, adjacent-1 and adjacent-2), 3:1 segregation and 4:0 segregation. Crossing-over between the centromere and the breakpoint or anaphase II non-disjunction result in gametes with unbalanced chromosomal complements. In theory, thirty-two different gametes with different chromosome complements can be generated at the end of meiosis I (Scriven et al., 1998). Only two among the gametes have normal chromosomal complements and the other gametes have partial aneuploidy. Uncommon viable conceptuses with abnormal chromosomal complement can be resulted from the gametes with partial aneuploidy.

Table 1. PGD outcomes of reciprocal translocation carriers

References	Biopsy stage	% of normal or balanced embryos	% of clinical pregnancy	% of abnormal pregnancy*	% of ET cancel**	PGD techniques
Gianaroli et al. (2002)	Cleavage stage	11.5 (7/61)	25.0		66.7	FISH
Lim et al. (2004)	Cleavage stage	23.1 (164/710)	35.1	5.6	9.3	FISH
Lim et al. (2008)	Cleavage stage	23.2 (114/508)	38.6	6.8	13.7	FISH
Ko et al. (2010)	Cleavage stage	18.7 (282/1,508)	28.4	7.8	12.8	FISH
Fiorentino et al. (2011)	Cleavage stage	11.8 (16/136)	63.6	0	38.9	aCGH
Keymolen et al. (2012)	Cleavage stage	20.8 (323/1,553)	26.7	0	44.6 (51.9)	FISH
Tan et al. (2013)	Cleavage stage	19.9 (449/2,258)	38.9	6.3	31.9	FISH
	Blastocyst stage	35.5 (177/499)	73.8	8.2	29.9	SNP array
Xiong et al. (2014)	Blastocyst stage	35.5 (177/499)			29.9	SNP array
Tobler et al. (2014)	Cleavage stage + Blastocyst stage	45.0 (226/498)	61.3		14.7	SNP array + aCGH
Tan et al. (2014)	Blastocyst stage	34.4 (84/244)				NGS
		35.6 (216/606)				SNP array
Idowu et al. (2015)	Cleavage stage + Blastocyst stage	19.0	59.3	7.4	38.0	SNP array

* Percent per embryos transfer.

** Embryo transfer was canceled because of the absence of normal/balanced embryos or insufficient quality of embryos for transfer.

Reciprocal translocation carrier couples can achieve successful pregnancies through PGD and the pregnancy rate of them is comparable to that of couples with normal karyotype. PGD results can be obtained in more than 90% of diagnosed embryos and about 11 ~ 45% of the embryos are identified as normal or balanced and transferable embryos (Table 1). Biopsy stage of embryos can affect the rate of normal or balanced embryos. The rate of normal or balanced embryos is higher when embryos are biopsied at blastocyst stage than cleavage stage (Xiong et al., 2014; Tobler et al., 2014; Idowu et al., 2015). Compared to blastomere biopsy, the fewer embryos are available for biopsy in blastocyst biopsy. The unbalanced embryos can develop to the blastocyst stage but it seems that the developmental potential of unbalanced embryos is inferior to that of balanced embryos. Therefore, the rate of normal or balanced embryos is higher when biopsy is carried out in blastocysts (Tan et al., 2013).

Embryo transfers are occasionally cancelled due to the absence of normal or balanced embryos in several cycles, ranging from 9.3 to 51.3% of started cycles. The main reason for the absence of normal or balanced embryos is that the number of embryos available for PGD is small. Although rare, embryo transfer might be cancelled due to the absence of normal or balanced embryos when a large number of embryos are available for PGD. And among pregnancies achieved by PGD, several pregnancies are aborted within the first trimester of

gestation. Aneuploidies of translocation-unrelated chromosomes are frequently observed in embryos or abortuses. Aneuploidies of translocation-unrelated chromosomes are observed in 12 ~ 26% of diagnosed embryos (Xiong et al., 2014; Idwou et al., 2015). It is known that the incidence of aneuploidy or mosaicism is very high in human preimplantation embryos. Chromosomal aneuploidy is increased in translocation carriers (Blanco et al., 2000; Morel et al., 2001; Oliver-Bonet et al., 2001) and about 50% of the embryos diagnosed as normal-balanced are aneuploid (Pujol et al., 2006). It has been suggested that translocations might disturb the meiotic disjunction of the translocation-unrelated chromosome pairs resulting in non-disjunction. Diagnoses for translocation-unrelated chromosomes are not possible when PGD for chromosome rearrangement is performed using FISH. However, abnormalities of translocation-unrelated chromosomes can be diagnosed since PGD for 24 chromosomes using array CGH or NGS is currently performed (Xiong et al., 2014; Tan et al., 2014; Idwou et al., 2015; Gui et al. 2016). PGD for 24 chromosomes has the merit of diagnosing aneuploidies of translocation-unrelated chromosomes together with translocation-related chromosomes. However, at the same time, the risk of embryo transfer cancellation is high due to the aneuploidies of translocation-unrelated chromosomes (Fiorentino et al., 2011; Tan et al., 2013; Xiong et al., 2014).

According to meiotic segregation analysis of embryos from reciprocal translocation carriers, 2:2 segregation is the most prevalent segregation mode. The 2:2 segregation is identified in more than half of the embryos from reciprocal translocation carriers (Scriven et al., 2000; Lim et al., 2008a; Ko et al., 2010). Among 2:2 segregation modes, the incidence of alternate segregation is similar that of adjacent-1 segregation and the incidence of adjacent-2 segregation is lower than that of alternate or adjacent-1 segregation. The incidence of 3:1 segregation mode is similar to that of alternate or adjacent-1 segregation and 4:0 segregation is observed although the frequency is low. And meiotic segregation cannot be determined in about 15% of diagnosed embryos. Meiotic segregation can be affected by gender of carrier, translocation-related chromosomes or the location of breakpoint. The frequency of normal or balanced embryos is not different between male and female carriers. However, the incidences of each segregation mode are significantly different between male and female carriers. The incidence of 2:2 segregation, especially adjacent-1 segregation, is higher in male carriers (60.8% vs. 52.7%, $P < 0.05$) and the incidences of 3:1 segregation or 4:0 segregation are significantly higher ($P < 0.05$) in female carriers (28.4% vs. 20.5% and 3.2% vs. 1.6%, respectively) (Ko et al., 2010). The incidence of alternate segregation is low when acrocentric chromosomes (chromosome 13, 14, 15, 21 or 22) are involved in translocation (Lim et al., 2008; Ye et al., 2012). And the incidence of normal or balanced embryos is lower in translocation carriers who have terminal breakpoint than carriers whose breakpoint is not terminal (Ye et al., 2012). It is known that the acrocentric chromosomes are less stable during meiosis or mitosis than metacentric or submetacentric chromosomes. During meiosis, translocated acrocentric chromosomes may not form a typical quadrivalent as observed in reciprocal translocations between metacentric and/or submetacentric chromosomes. This phenomenon may be related to predisposing of adjacent-2 or 3:1 segregation (Jalbert and Sele, 1979).

PGD FOR ROBERTSONIAN TRANSLOCATION

Centric fusion of two acrocentric chromosomes results in Robertsonian translocation, which is one of the most common chromosomal rearrangements (Nielsen and Wohler, 1991). Like balanced reciprocal translocation carriers, Robertsonian translocation carriers are phenotypically normal but they are also at an increased risk of chromosomally abnormal pregnancies (Scriven et al., 1998). Robertsonian translocation carriers who experienced repeated abortions or birth of babies with congenital anomalies or mental retardations can achieve normal pregnancy by PGD (Munne et al., 1998; Fischer et al., 2010). During meiosis of Robertsonian translocation carriers, trivalent is formed and gametes with eight different chromosomal complements are produced. Only two gametes produced by alternate segregation have normal or balanced chromosomal complements and the others have unbalanced chromosomal complements.

Pregnancy outcome of Robertsonian translocation carriers is similar to that of balanced reciprocal translocation carriers. More than 90% of biopsied embryos are diagnosed successfully and about 30 ~ 50% of diagnosed embryos are available to embryo transfer (Table 2). The rate of transferable embryos is higher in male carriers than in female carriers (Ko et al., 2013; De Rycke et al., 2015). Embryo transfer is cancelled due to the absence of normal or balanced embryos in 10~30% of oocyte retrieval cycles. Similar to reciprocal translocation carriers, several pregnancies after PGD are aborted within the first trimester of gestation. However, miscarriage rate is significantly lower after PGD than before PGD. Various karyotypes are observed in the cytogenetic analysis results of the abortuses: normal, balanced karyotype or aneuploidies of translocation-unrelated chromosomes.

Table 2. PGD outcomes of Robertsonian translocation carriers

References	Biopsy stage	% of normal or balanced embryos	% of clinical pregnancy	% of abnormal pregnancy*	% of ET cancel**	PGD techniques
Gianaroli et al. (2002)	Cleavage stage	23.4 (26/111)	61.5		43.5	FISH
Lim et al. (2004)	Cleavage stage	22.3 (29/135)	10.0	0	9.1	FISH
Fiorentino et al. (2011)	Cleavage stage	27.5 (14/51)	83.3	0	40	aCGH
Ko et al. (2013)	Cleavage stage	26.5 (331/1,247)	28.3	7.1	5.8	FISH
Tan et al. (2013)	Cleavage stage	36.1 (435/1,204)	38.4	6.4	16.1	FISH
	Blastocyst stage	57.8 (126/218)	69.4	8.3	9.6	SNP array
Xiong et al. (2014)	Blastocyst stage	57.8 (126/218)			9.6	SNP array
Tan et al. (2014)	Blastocyst stage	53.5 (31/58)				NGS
		52.4 (120/229)				SNP array
Idowu et al. (2015)	Cleavage stage + Blastocyst stage	37.0	55.6	3.7	19.0	SNP array

* Percent per embryos transfer.

** Embryo transfer was canceled because of the absence of normal/balanced embryos or insufficient quality of embryos for transfer.

Alternate or adjacent segregation are the dominant segregation modes according to meiotic segregation analysis of embryos from PGD for Robertsonian translocation carriers. There are some differences in segregation patterns between female and male carriers (Munne et al., 2000b; Ko et al., 2013). The rate of embryos from alternate segregation is significantly higher in male than in female carriers (43.9% vs. 29.9%, $P < 0.01$) and the rate of embryos whose meiotic segregation cannot be determined is significantly higher in female than male carriers (15.8% vs. 8.9%, $P < 0.01$). In (13;14) translocation carriers, the most common Robertsonian translocation, the rate of embryos from alternate embryos is significantly higher in male than in female carriers (44.0% vs. 29.2%, $P < 0.01$) and the rate of embryos from adjacent segregation is significantly higher in female than male carriers (47.9% vs. 37.1%, $P < 0.01$). However, the information on meiotic segregation of Robertsonian translocation carriers, especially female carriers, is insufficient. More studies are needed. Aneuploidies of translocation-unrelated chromosomes might increase in unbalanced embryos. In Robertsonian translocation carriers, aneuploidies of chromosome 18 is significantly higher ($P < 0.001$) in embryos from 3:0 segregation or embryos whose meiotic segregation cannot be determined (chaotic segregation) than in embryos from 2:1 segregation (alternate segregation or adjacent segregation) (Ko et al., 2013). Among embryos whose translocation-related chromosomes are diagnosed as normal or balanced, several embryos cannot be transferred due to aneuploidies of chromosome 18. It is not known what cause the increase of chromosome 18 aneuploidy in embryos from 3:0 or chaotic segregation. However, in those embryos, factors that cause malsegregation of translocation-related chromosomes seem to increase aneuploidy of chromosome 18 and these factors need to be identified. Nowadays, aneuploidies of translocation-unrelated chromosomes can be diagnosed concurrently together with abnormalities of translocation-related chromosomes. In PGD cycles of Robertsonian translocation carriers where 24 chromosomes are diagnosed, a high clinical pregnancy rate (83.3%) and implantation rate (66.7%) are achieved in carriers of Robertsonian translocation (Fiorentino et al., 2011). However, aneuploidies of translocation-unrelated chromosomes are observed in more than half of the normal or balanced embryos (Fiorentino et al., 2011; Rius et al., 2011). Therefore, the number of transferable embryos might be decreased in those PGD cycles.

PGD FOR COMPLEX CHROMOSOME REARRANGEMENT (CCR)

Complex chromosome rearrangements (CCRs) are very rare events in human. CCRs are generally defined as balanced or unbalanced structural chromosomal abnormalities involving more than two breakpoints and a simple exchange of genetic material between two chromosomes (Pai et al., 1980). Most familial CCRs are of maternal origin (Batista et al., 1994; Madan et al., 1997) and most *de novo* CCRs are of paternal origin (Batista et al., 1993, 1994). There are few studies on the incidence of CCRs in the general population. However, the incidence of CCRs has been estimated to be around 0.1% among infertile individuals (Mau-Holzmann, 2005). Like other chromosome rearrangement carriers, CCR carriers are also phenotypically normal. However, they are estimated to have a 50% risk of miscarriage and a 20% risk of having a child with an unbalanced karyotype (Gorski et al., 1988; Madan et al., 1997). However, these figures cannot be applied to individual CCR carriers as a wide variety

of gametes can be produced depending on the number of chromosomes and the number of breakpoints involved in CCRs (Kausch et al., 1988; Loup et al., 2010). The risk of phenotype abnormality increases with the number of chromosomes and the number of breakpoints involved in CCRs (Ruiz et al., 1996; Madan et al., 1997; Giardino et al., 2006). Estimating risk for spontaneous abortions or phenotypic abnormalities of CCR carriers is very difficult and remain empirical, ranging from 50 to 100% for spontaneous abortion (Batista et al., 1994) and from 20 to 90% for phenotypic abnormalities (Gorski et al., 1988). Analyses of meiotic segregation of CCRs are also very difficult due to the nature of CCRs and the number of chromosomes involved in CCRs. An abnormal configuration of chromosomes is formed during meiosis I of CCR carriers. The abnormal configuration causes either malsegregation of derivative chromosomes or generation of a recombinant chromosome (Pellestor et al., 2011b). For example, a hexavalent configuration is formed to allow full synapses of homologous segments during meiosis of three-way translocation, the most common type of CCR (Saadallah and Hulten, 1985). Theoretically, 64 different chromosomal combinations can be formed at the end of meiosis I and only two combinations are balanced. And empirically, the incidence of balanced embryos is very low in CCR carriers, ranging from 9.1 to 16.2% (Escudero et al., 2008; Lim et al., 2008b; Scriven et al., 2014). Since its introduction, PGD has contributed to the achievement of pregnancy and prevention of miscarriage in IVF-ET program of chromosome rearrangement carriers and it has been applied to IVF-ET program of CCR carriers.

Compared to PGD cycles of other chromosome rearrangement such as reciprocal translocation, Robertsonian translocation or inversion, the cancellation rate of embryo replacement is higher (25.0 ~ 64.7%) due to the absence of normal or balanced embryos in PGD cycle of CCR carriers (Table 3). The incidence of normal or balanced embryos is very low (4.9 ~ 5.6%) and therefore, the average number of transferred embryos is also small (1.7 ± 1.1 . per embryo replacement, unpublished data).

As mentioned above, meiotic segregation analysis of CCRs is very difficult due to the nature of CCR and its rarity. Although few studies have been conducted, meiotic segregation of three-way translocation, the most common CCR, is analyzed in several studies. The 3:3, 4:2 and 5:1 segregations are observed in embryos from three-way translocation carriers (34.1%, 20.7% and 2.2%, respectively). Among 3:3 segregations, the rates of alternate, adjacent 1 and adjacent 2 segregation are 16.4%, 52.5% and 31.1%, respectively. Cross-over between sister chromatids is observed in 7.3% of embryos. The 6:0 segregation is not observed and the meiotic segregation could not be determined in 43.0% of embryos (unpublished data). The incidence of embryos whose meiotic segregation mode cannot be determined is significantly higher ($P < 0.001$) in three-way translocation carriers (43.0%, unpublished data) than reciprocal translocation (16.8%) or Robertsonian translocation (14.4%) carriers (Ko et al., 2010, 2013). Various genomic, genetic and non-genetic factors can influence the meiotic segregation of translocation configurations (Jalbert et al., 1980). The number of chromosomes involved in a translocation is included in the factors. Also, the occurrence of double interstitial chiasma may affect the meiotic segregation of three-way translocation (Cifuentes et al., 1998). Therefore, in three-way translocation carriers, chaotic segregation may occur in at a higher incidence than two-way translocation carriers due to the involvement of three chromosomes and the occurrence of double interstitial chiasma although more extensive studies are needed to solve this problem.

Table 3. PGD outcomes of complex chromosome rearrangement carriers

References	Biopsy stage	% of normal or balanced embryos	% of clinical pregnancy	% of abnormal pregnancy*	% of ET cancel**	PGD techniques
Lim et al. (2008)	Cleavage stage	7.4 (4/54)	33.3	0	25.0	FISH
Escudero et al. (2008)	Cleavage stage	6.4 (9/140)	50.0		69.2	FISH
Scriven et al. (2014)	Cleavage stage	16.2 (6/37)	100		66.7	FISH
Ko et al. (Unpublished)	Cleavage stage	5.0 (22/444)	33.3	66.7	64.7	FISH

* Percent per embryos transfer.

** Embryo transfer was canceled because of the absence of normal/balanced embryos or insufficient quality of embryos for transfer.

Couples in whom one partner is a carrier of two different chromosome rearrangements has underwent PGD cycles. The incidence of embryos available to embryo replacement is low (4.9%, unpublished data) in those PGD cycles. Normal or balanced embryos for two different chromosome rearrangements are rare. Therefore, cancellation of embryo replacement due to the absence of normal or balanced embryos is high (about 77.8%, unpublished data). The two different chromosome rearrangements form two separate chromosomal configurations and undergo meiosis independently (Miller and Flatz, 1984; Zahed et al., 1998). In these PGD cycles, the meiotic segregation analysis is difficult as the number of analyzed embryos is very small and various chromosomes are involved in CCRs.

In some couples, both partners are carriers of chromosome rearrangements. Those couples can benefit from PGD. In PGD cycles of those couples, the incidence of normal or balanced is also very low (5.6%, unpublished data). The cancellation rate of embryo replacements is also very high. Similar to couples in whom one partner is a carrier of two different chromosome rearrangements, most of embryos that are normal or balanced in one chromosome rearrangement are unbalanced in the other chromosome rearrangement. The couples in whom both partners are carriers of chromosome rearrangements seem to have more reproductive risks than couples in whom one partner is a carrier of a chromosome rearrangement (Beyazyurek et al., 2010). Instead, the reproductive risks of those couples appear to be similar to those of couples in whom one partner is a carrier of two different chromosome rearrangements. The rate of normal or balanced embryos is similar between couples in whom both partners are carriers of chromosome rearrangements and couples in whom one partner is a carrier of two different chromosome rearrangements (5.6% vs. 4.9%, unpublished data). Strictly speaking, couples in whom both partners are carriers of chromosome rearrangements are not CCR carriers. However, the respective chromosome rearrangements of both partners seem to have a similar effect on pregnancy outcomes of PGD cycles to that of CCR carriers.

APPLICATION OF COMPREHENSIVE CHROMOSOME SCREENING (CCS) INTO PGD FOR BALANCED CHROMOSOME REARRANGEMENTS

With the development of PGD techniques, comprehensive chromosome screening using array CGH, SNP array or NGS is currently applied to PGD for balanced chromosome rearrangement carriers. In 2000s, FISH was the main technique of PGD for chromosome rearrangement. However, the techniques that can diagnose 24 chromosomes are currently applied to PGD for chromosome rearrangement carriers (Alfarawati et al., 2011; Fiorentino et al., 2011; Treff et al., 2011; Colls et al., 2012; van Uum et al., 2012; Huang et al., 2013; Tan et al., 2013). PGD for chromosome rearrangements is possible using these techniques and single blastomere biopsied from cleavage stage embryo, but recently 24-chromosome PGD combined with trophectoderm biopsy are expanding and gradually replacing the polar body or blastomere biopsy.

Compared to FISH-PGD using blastomere biopsied at day 3, array-based PGD combined with trophectoderm biopsy (day 5 or 6) results in higher clinical pregnancy rate and implantation rate (Tan et al. 2013). The number of embryos available for PGD is significantly higher in FISH-PGD than in array-based PGD. Only translocation-related chromosomes are diagnosed in FISH-PGD but 24 all chromosomes are diagnosed in array-based PGD. Therefore, the number of transferable embryos might decrease in array-based PGD. However, the average number of transferable embryos per embryo transfer is not significant different between FISH-PGD and array-based PGD (Tan et al., 2013). Instead, the rate of transferable embryos is higher in array-based PGD than in FISH-PGD. The higher rate of transferable embryos in array-based PGD might result from the small number of embryos available for PGD. And developmental potential of unbalanced embryos seems to be lower than that of normal or balanced embryos since the rate of transferable embryos is higher in array-based PGD than in FISH-PGD although the number of embryos available for PGD is larger in FISH-PGD than array-based PGD. In addition, miscarriage rate is higher in FISH-PGD than in array-based PGD although there is no significant difference (Tan et al., 2013). Abnormalities of chromosome rearrangement-unrelated chromosomes are observed in aborted tissues of FISH-PGD but not in aborted tissues of array-based PGD. Overall, array- or NGS-based PGD combined with trophectoderm biopsy can improve the clinical outcomes compared to FISH-PGD in PGD for chromosome rearrangement carriers. In the near future, array- or NGS-based PGD will be the major PGD technique for chromosome rearrangement carriers since its use is expanding.

CONCLUSION

Chromosome rearrangement carriers have achieved pregnancies by PGD after its introduction into human IVF-ET programs. In 2000s, FISH was widely used to PGD for chromosome rearrangement carriers and healthy babies were born. However, some of pregnancies achieved by PGD resulted in miscarriages that might be caused by abnormalities of chromosome rearrangement-unrelated chromosomes. PGD techniques for chromosome rearrangements have developed from diagnosis for only rearranged chromosomes to diagnosis for all chromosomes, from FISH to CCS using array CGH or NGS. With the development of

PGD techniques, miscarriages due to abnormalities of chromosome rearrangement-unrelated chromosomes will decrease and clinical outcomes of chromosome rearrangement carriers are expected to be improved. And although PGD techniques are developed, distinguishment between chromosomally normal embryos and balanced embryos is impossible with the PGD techniques we have today. Therefore, the techniques that can distinguish chromosomally normal embryos from balanced embryos have to be developed in the near future.

Lastly, though there are some exceptions, the possibility of obtaining normal or balanced embryos is higher in cycles that a large number of embryos are available for PGD than in cycles that a small number of embryos are available. Therefore, PGD might be performed by using as many embryos as possible. And also, the possibility of obtaining normal or balanced embryos is higher in cycles where a large number of embryos are obtained in one cycle compared to cycles where embryos are collected over several cycles. However, care should be taken to avoid excessive stimulation that may be harmful to patients and cause other side effects. With the development of PGD techniques, elaborate care for chromosome rearrangement carriers enable them to have healthy babies and will make possible the improvement of clinical outcomes of chromosome rearrangement carriers.

REFERENCES

- Alfarawati S, Fragouli E, Colls P, Wells D. First births after preimplantation genetic diagnosis of structural chromosome abnormalities using comparative genomic hybridization and microarray analysis. *Hum. Reprod.*, 2011, 26, 1560-1574.
- Batista DAS, Tuck-Muller CM, Martinez JE, Kearns WG, Pearson PL, Stetten G. A complex chromosomal rearrangement detected prenatally and studied by fluorescence in situ hybridization. *Hum. Genet.*, 1993, 92, 117-121.
- Batista DAS, Pai GS, Stetten G. Molecular analysis of a complex chromosomal rearrangement and a review of familial cases. *Am. J. Med. Genet.*, 1994, 53, 255-263.
- Bernicot I, Schneider A, Mace A, Hamamah S, Hedon B, Pellestor F, Anahory T. Analysis using fish of sperm and embryos from two carriers of rare rob(13;21) and rob(15;22) robertsonian translocation undergoing PGD. *Eur. J. Med. Genet.*, 2012, 55, 245-251.
- Beyazyurek C, Ekmekci CG, Sağlam Y, Cinar C, Kahraman S. Preimplantation genetic diagnosis (PGD) for extremes-successful birth after PGD for a consanguineous couple carrying an identical balanced reciprocal translocation. *Fertil. Steril.*, 2010, 93, 2413.e1-5.
- Bielanska M, Tan SL, Ao A. Chromosomal mosaicism throughout human preimplantation development in vitro: incidence, type, and relevance to embryo outcome. *Hum. Reprod.*, 2002, 17, 413-419.
- Blanco J, Egozcue J, Vidal F. Interchromosomal effects for chromosome 21 in carriers of structural chromosome reorganizations determined by fluorescence in situ hybridization on sperm nuclei. *Hum. Genet.* 2000, 106, 500-505.
- Capalbo A, Rienzi L, Cimadomo D, Maggiulli R, Elliott T, Wright G, Nagy ZP, Ubaldi FM. Correlation between standard blastocyst morphology, euploidy and implantation: an observational study in two centers involving 956 screened blastocysts. *Hum. Reprod.*, 2014, 29, 1173-1181.

- Capalbo A, Ubaldi FM, Cimadomo D, Maggiulli R, Patassini C, Dusi L, Sanges F, Buffo L, Venturella R, Rienzi L. Consistent and reproducible outcomes of blastocyst biopsy and aneuploidy screening across different biopsy practitioners: a multicentre study involving 2586 embryo biopsies. *Hum. Reprod.*, 2015, 31, 199-208.
- Chang EM, Han JE, Kwak IP, Lee WS, Yoon TK, Shim SH. Preimplantation genetic diagnosis for couples with a Robertsonian translocation: practical information for genetic counseling. *J. Assist. Reprod. Genet.*, 2012, 29, 67-75.
- Cimadomo D, Capalbo A, Ubaldi FM, Scarica C, Palagiano A, Canipari R, Rienzi L. The impact of biopsy on human embryo developmental potential during preimplantation genetic diagnosis. *Biomed. Res. Int.*, 2016, 2016, 7193075.
- Cifuentes P, Navarro J, Míguez L, Egozcue J, Benet J. Sperm segregation analysis of a complex chromosome rearrangement, 2;22;11, by whole chromosome painting. *Cytogenet. Cell Genet.* 1998, 82, 204-209.
- Cohen J, Wells D, Munne S. Removal of 2 cells from cleavage stage embryos is likely to reduce the efficacy of chromosomal tests that are used to enhance implantation rates. *Fertil. Steril.*, 2007, 87, 496-503.
- Colls P, Escudero T, Fischer J, Cekleniak NA, Ben-Ozer S, Meyer B, Damien M, Grifo JA, Hershlag A, Munne S. Validation of array comparative genome hybridization for diagnosis of translocations in preimplantation human embryos. *Reprod. Biomed. Online*, 2012, 24, 621-629.
- De Rycke M, Belva F, Goossens V, Moutou C, SenGupta SB, Traeger-Synodinos J, Coonen E. ESHRE PGD Consortium data collection XIII: cycles from January to December 2010 with pregnancy follow-up to October 2011. *Hum. Reprod.* 2015, 30, 1763-1789.
- De Vos A and Van Steirteghem A. Aspects of biopsy procedures prior to preimplantation genetic diagnosis. *Prenat. Diag.*, 2001, 21, 767-780.
- Dumoulin JC, Bras M, Coonen E, Dreesen J, Geraedts JP, Evers JL. Effect of $\text{Ca}^{2+}/\text{Mg}^{2+}$ -free medium on the biopsy procedure for preimplantation genetic diagnosis and further development of human embryos. *Hum. Reprod.*, 1998, 13, 2880-2883.
- Eldar-Geva T, Srebnik N, Altarescu G, Varshaver I, Brooks B, Levy-Lahad E, Bromiker R, Schimmel MS. Neonatal outcome after preimplantation genetic diagnosis. *Fertil. Steril.*, 2014, 102, 1016-1021.
- Escudero T, Estop A, Fischer J, Munne S. Preimplantation genetic diagnosis for complex chromosome rearrangements. *Am. J. Med. Genet. A*, 2008, 146A:1662-1669.
- Fischer J, Colls P, Escudero T, Munne S. Preimplantation genetic diagnosis (PGD) improves pregnancy outcome for translocation carriers with a history of recurrent losses. *Fertil. Steril.*, 2010, 94, 283-289.
- Fiorentino F, Spizzichino L, Bono S, Biricik A, Kokkali G, Rienzi L, Ubaldi FM, Iammarrone E, Gordon A, Pantos A. PGD for reciprocal and Robertsonian translocations using array comparative genomic hybridization. *Hum. Reprod.*, 2011, 26, 1925-1935.
- Fiorentino F, Biricik A, Bono S, Spizzichino L, Cotroneo E, Cottone G, Kokocinski F, Michel CE. Development and validation of a next-generation sequencing-based protocol for 24-chromosome aneuploidy screening of embryos. *Fertil. Steril.*, 2014a, 101, 1375-1382.
- Fiorentino F, Bono S, Biricik A, Nuccitelli A, Cotroneo E, Cottone G, Kokocinski F, Michel CE, Minasi MG, Greco G. Application of next-generation sequencing technology for comprehensive aneuploidy screening of blastocysts in clinical preimplantation genetic screening cycles. *Hum. Reprod.*, 2014b, 29, 2802-2813.

- Geber S, Bossi R, Lisboa CB, Valle M, Sampaio M. Laser confers less embryo exposure than acid tyrode for embryo biopsy in preimplantation genetic diagnosis cycles: a randomized study. *Reprod. Biol. Endocrinol.*, 2011, 28, 58.
- Giardino D, Corti C, Ballarati L, Finelli P, Valtorta C, Botta G, Giudici M, Grosso E, Larizza L. Prenatal diagnosis of a de novo complex chromosome rearrangement (CCR) mediated by six breakpoints, and a review of 20 prenatally ascertained CCRs. *Prenat. Diagn.*, 2006, 26, 565-570.
- Goldman AS and Hulten MA. Meiotic analysis by FISH of a human male 46,XY,t(15;20)(q11.2;q11.2) translocation heterozygote: quadrivalent configuration, orientation and first meiotic segregation. *Chromosoma*, 1993, 102, 102-111.
- Goossens V, De Rycke M, De Vos A, Staessen C, Michiels A, Verpoest W, Van Steirteghem A, Bertrand C, Liebaers I, Devroey P, Sermon K. Diagnostic efficiency, embryonic development and clinical outcome after the biopsy of one or two blastomeres for preimplantation genetic diagnosis. *Hum. Reprod.*, 2008, 23, 481-492.
- Gorski JL, Kistenmacher ML, Punnett HH, Zackai EH, Emanuel BS. Reproductive risks for carriers of complex chromosome rearrangements: analysis of 25 families. *Am. J. Med. Genet.*, 1988, 29, 247-261.
- Grace J, El-Toukhy T, Scriven P, Ogilvie C, Pickering S, Lashwood A, Flinter F, Khalaf Y, Braude P. Three hundred and thirty cycles of preimplantation genetic diagnosis for serious genetic disease: clinical considerations affecting outcome. *BJOG*, 2006, 113, 1393-1401.
- Gui B, Yao Z, Li Y, Liu D, Liu N, Xia Y, Huang Y, Mei L, Ma R, Lu S, Liang D, Wu L. Chromosomal analysis of blastocysts from balanced chromosomal rearrangement carriers. *Reprod.*, 2016, 151, 455-464.
- Hamerton JL, Canning N, Ray M, Smith S. A cytogenetic survey of 14,069 newborn infants. I. Incidence of chromosome abnormalities. *Clin. Genet.*, 1975, 8, 223-243.
- Handyside AH, Kontogianni EH, Hardy K, Winston RML. Pregnancies from biopsied human preimplantation embryos sexed by Y-specific DNA amplification. *Nature*, 1990, 344, 768-770.
- Huang CC, Chang LJ, Tsai YY, Hung CC, Fang MY, Su YN, Chen HF, Chen SU. A feasible strategy of preimplantation genetic diagnosis for carriers with chromosomal translocation: Using blastocyst biopsy and array comparative genomic hybridization. *J. Formos. Med. Assoc.*, 2013, 112, 537-544.
- Huang J, Yan L, Fan W, Zhao N, Zhang Y, Tang F, Xie XS, Qiao J. Validation of multiple annealing and looping-based amplification cycle sequencing for 24-chromosome aneuploidy screening of cleavage stage embryos. *Fertil. Steril.* 2014, 102, 1685-1691.
- Idowu D, Merrion K, Wemmer N, Mash JG, Pettersen B, Kijacic D, Lathi RB. Pregnancy outcomes following 24-chromosome preimplantation genetic diagnosis in couples with balanced reciprocal or Robertsonian translocations. *Fertil. Steril.* 2015, 103, 1037-1042.
- Jacobs PA, Melville M, Ratcliffe S, Keay AJ, Syme J. A cytogenetic survey of 11,680 newborn infants. *Ann. Hum. Genet.*, 1974, 37, 359-376.
- Jalbert P and Sele B. Factors predisposing to adjacent 2 and 3:1 disjunctions: study of 161 human reciprocal translocations. *Med. Genet.*, 1979, 16, 467-478.
- Jalbert P, Se'le B, Jalbert H. Reciprocal translocations: a way to predict the mode of imbalanced segregation by pachytene-diagram drawing. A study of 151 human translocations. *Hum. Genet.*, 1980, 55, 209-222.

- Jones AE, Wright G, Kort HI, Straub RJ, Nagy ZP. Comparison of laser-assisted hatching and acidified Tyrode's hatching by evaluation of blastocyst development rates in sibling embryos: a prospective randomized trial. *Fertil. Steril.*, 2006, 85, 487-491.
- Kausch K, Haaf T, Köhler J, Schmid M. Complex chromosomal rearrangement in a woman with multiple miscarriages. *Am. J. Med. Genet.*, 1988, 31, 415-420.
- Keymolen K, Staessen C, Verpoest W, Liebaers I, Bonduelle M. Preimplantation genetic diagnosis in female and male carriers of reciprocal translocations: clinical outcome until delivery of 312 cycles. *Eur. J. Hum. Genet.*, 2012, 20, 376-380.
- Ko DS, Cho JW, Park SY, Kim JY, Koong MK, Song IO, Kang IS, Lim CK. Clinical outcomes of preimplantation genetic diagnosis (PGD) and analysis of meiotic segregation modes in reciprocal translocation carriers. *Am. J. Med. Genet. A.* 2010, 152A, 1428-1433.
- Ko, DS, JW Cho, HS Lee, JY Kim, IS Kang, KM Yang, CK Lim. Preimplantation genetic diagnosis outcomes and meiotic segregation analysis of robertsonian translocation carriers. *Fertil. Steril.*, 2013, 99, 1369-1376.
- Liao Y, Wang L, Zhang D, Liu C. Identification of a balanced complex chromosomal rearrangement involving chromosomes 3, 18 and 21 with recurrent abortion: case report. *Mol. Cytogenet.*, 2014, 7, 39.
- Lim CK, Jun JH, Min DM, Lee HS, Kim JY, Koong MK, Kang IS. Efficacy and clinical outcome of preimplantation genetic diagnosis using FISH for couples of reciprocal and Robertsonian translocations: the Korean experience. *Prenat. Diagn.*, 2004, 24, 556-561.
- Lim CK, Cho JW, Song IO, Kang IS, Yoon YD, Jun JH. Estimation of chromosomal imbalances in preimplantation embryos from preimplantation genetic diagnosis cycles of reciprocal translocations with or without acrocentric chromosomes. *Fertil. Steril.*, 2008a, 90, 2144-2151.
- Lim CK, Cho JW, Kim JY, Kang IS, Shim SH, Jun JH. A healthy live birth after successful preimplantation genetic diagnosis for carriers of complex chromosome rearrangements. *Fertil. Steril.*, 2008b, 90, 1680-1684.
- Loup V, Bernicot I, Janssens P, Hedon B, Hamamah S, Pellestor F, Anahory T. Combined FISH and PRINS sperm analysis of complex chromosome rearrangement t(1;19;13): an approach facilitating PGD. *Mol. Hum. Reprod.*, 2010, 16, 111-116.
- Lukaszuk K, Puksza S, Ochman K, Cybulska C, Liss J, Pastuszek E, Zabielska J, Woclawek-Potocka I. Healthy Baby Born to a Robertsonian Translocation Carrier following Next-Generation Sequencing-Based Preimplantation Genetic Diagnosis: A Case Report. *AJP Rep.* 2015, 5, e172-5.
- Madan K, Nieuwint AWM, Van Bever Y. Recombination in a balanced complex translocation of a mother leading to a balanced reciprocal translocation in the child. Review of 60 cases of balanced complex translocations. *Hum. Genet.*, 1997, 99, 806-815.
- Martin RH and Hulten M. Chromosome complements in 695 sperm from three men heterozygous for reciprocal translocations, and a review of the literature. *Hereditas*, 1993, 118, 165-175.
- Mau-Holzmann UA. Somatic chromosomal abnormalities in infertile men and women. *Cytogenet. Genome Res.*, 2005, 111, 317-336.
- McArthur SJ, Leigh D, Marshall JT, DeBoer KA, Jansen RPS. Pregnancies and live births after trophoctodermiopsy and preimplantation genetic testing of human blastocysts. *Fertil. Steril.*, 2005, 84, 1628-1636.

- Miller K and Flatz SD. Segregation of two independent chromosome translocations in one family. *Hum. Genet.*, 1984, 68, 93-95.
- Morel F, Roux C, Bresson JL. FISH analysis of the chromosomal status of spermatozoa from three men with 45,XY,der(13;14)(q10;q10) karyotype. *Mol. Hum. Reprod.* 2001, 7, 483-488.
- Moutou C, Goossens V, Coonen E, De Rycke M, Kokkali G, Renwick P, SenGupta SB, Vesela K, Traeger-Synodinos J. ESHRE PGD Consortium data collection XII: cycles from January to December 2009 with pregnancy follow-up to October 2010. *Hum. Reprod.*, 2014, 29, 880-903.
- Munne S, Morrison L, Fung J, Marquez C, Weier U, Bahce M, Sable D, Grundfeld L, Schoolcraft B, Scott R, Cohen J. Spontaneous abortions are reduced after preconception diagnosis of translocations. *J. Assist. Reprod. Genet.* 1998, 15, 290-296.
- Munne S, Sandalinas M, Escudero T, Fung J, Gianaroli L, Cohen J. Outcome of preimplantation genetic diagnosis of translocations. *Fertil. Steril.*, 2000a, 73,1209-1218.
- Munne S, Escudero T, Sandalinas M, Sable D, Cohen J. Gamete segregation in female carriers of Robertsonian translocations. *Cytogenet. Cell Genet.* 2000b, 90, 303-308.
- Munne S. Preimplantation genetic diagnosis of numerical and structural chromosome abnormalities. *Reprod. Biomed. Online*, 2002, 4, 183-196.
- Nielsen J and Wohler M. Chromosome abnormalities found among 34,910 newborn children: results from a 13-year incidence study in Arhus, Denmark. *Hum. Genet.*, 1991, 87, 81-83.
- Nussbaum RL, McInnes RR, Willard HF. (2016) Principles of Clinical Cytogenetics and Genome Analysis. In: Nussbaum RL, McInnes RR, Willard HF (8th Ed.). *Thompson and Thompson genetics in medicine* (57-74), Philadelphia: Elsevier.
- Oliver-Bonet M, Navarro J, Codina-Pascual M, Carrera M, Egozcue J, Benet J. Meiotic segregation analysis in a t(4;8) carrier: Comparison of FISH methods on sperm chromosome metaphases and interphase sperm nuclei. *Eur. J. Hum. Genet.* 2001, 9, 395-403.
- Otani T, Roche M, Mizuike M, Colls P, Escudero T, Munne S. Preimplantation genetic diagnosis significantly improves the pregnancy outcome of translocation carriers with a history of recurrent miscarriage and unsuccessful pregnancies. *Reprod. Biomed. Online*, 2006, 13, 869-874.
- Pai GS, Thomas GH, Mahoney W, Migeon BR. Complex chromosome rearrangements. Report of a new case and literature review. *Clin. Genet.*, 1980, 18, 436-444.
- Pellestor F, Puechberty J, Weise A, Lefort G, Anahory T, Liehr T, Sarda P. Meiotic segregation of complex reciprocal translocations: direct analysis of the spermatozoa of a t(5;13;14) carrier. *Fertil. Steril.*, 2011b, 95, 2433. e17-22.
- Pey R, Vial C, Schatten G, Hafner M. Increase of intracellular Ca²⁺ and relocation of E-cadherin during experimental decompaction of mouse embryos. *Proc. Natl. Acad. Sci. US*, 1998, 95, 12977-12982.
- Pujol A, Benet J, Staessen C, Van Assche E, Campillo M, Egozcue J, Navarro J. The importance of aneuploidy screening in reciprocal translocation carriers. *Reproduction* 2006, 131, 1025-1035.
- Rius M, Obradors A, Daina G, Ramos L, Pujol A, Martínez-Passarell O, Marquès L, Oliver-Bonet M, Benet J, Navarro J. Detection of unbalanced chromosome segregations in preimplantation genetic diagnosis of translocations by short comparative genomic hybridization. *Fertil. Steril.*, 2011, 96, 134-142.

- Ruiz C, Grubs RE, Jewett T, Cox-Jones K, Abruzzese E, Pettenati MJ, Rao PN. Prenatally diagnosed de novo apparently balanced complex chromosome rearrangements: two new cases and review of the literature. *Am. J. Med. Genet.*, 1996, 64, 478-484.
- Saadallah N and Hulten M. A complex three breakpoint translocation involving chromosome 2,4 and 9 identified by meiotic investigations of a human male ascertained for subfertility. *Hum. Genet.*, 1985, 71, 312-320.
- Scriven PN, Handyside AH, Ogilvie CM. Chromosome translocations: segregation modes and strategies for preimplantation genetic diagnosis. *Prenat. Diagn.*, 1998, 18, 1437-1449.
- Scriven PN, O'Mahony F, Bickerstaff H, Yeong CT, Braude P, Mackie Ogilvie C. Clinical pregnancy following blastomere biopsy and PGD for a reciprocal translocation carrier: analysis of meiotic outcomes and embryo quality in two IVF cycles. *Prenat. Diagn.*, 2000, 20, 587-592.
- Scriven PN, Bint SM, Davies AF, Ogilvie CM. Meiotic outcomes of three-way translocations ascertained in cleavage-stage embryos: refinement of reproductive risks and implications for PGD. *Eur. J. Hum. Genet.*, 2014, 22, 748-753.
- Sefton M, Johnson MH, Clayton L, McConnell JM. Experimental manipulations of compaction and their effects on the phosphorylation of uvomorulin. *Mol. Reprod. Dev.*, 1996, 44, 77-87.
- Tan YQ, Tan K, Zhang SP, Gong F, Cheng DH, Xiong B, Lu CF, Tang XC, Luo KL, Lin G, Lu GX. Single-nucleotide polymorphism microarray-based preimplantation genetic diagnosis is likely to improve the clinical outcome for translocation carriers. *Hum. Reprod.*, 2013, 28, 2581-2592.
- Tan Y, Yin X, Zhang S, Jiang H, Tan K, Li J, Xiong B, Gong F, Zhang C, Pan X, Chen F, Chen S, Gong C, Lu C, Luo K, Gu Y, Zhang X, Wang W, Xu X, Vajta G, Bolund L, Yang H, Lu G, Du Y, Lin G. Clinical outcome of preimplantation genetic diagnosis and screening using next generation sequencing. *Gigascience*, 2014, 4, 30.
- Tobler KJ, Brezina PR, Benner AT, Du L, Xu X, Kearns WG. Two different microarray technologies for preimplantation genetic diagnosis and screening, due to reciprocal translocation imbalances, demonstrate equivalent euploidy and clinical pregnancy rates. *J. Assist. Reprod. Genet.*, 2014, 31, 843-850.
- Treff NR, Northrop LE, Kasabwala K, Su J, Levy B, Scott RT Jr. Single nucleotide polymorphism microarray-based concurrent screening of 24-chromosome aneuploidy and unbalanced translocations in preimplantation human embryos. *Fertil. Steril.* 2011, 95, 1606-12.e1-2.
- Van Assche E, Staessen C, Vegetti W, Bonduelle M, Vandervorst M, Van Steirteghem A, Liebaers I. Preimplantation genetic diagnosis and sperm analysis by fluorescence in-situ hybridization for the most common reciprocal translocation t(11;22). *Mol. Hum. Reprod.*, 1999, 5, 682-690.
- Van Dyke DL, Weiss L, Roberson JR, Babu VR. The frequency and mutation rate of balanced autosomal rearrangements in man estimated from prenatal genetic studies for advanced maternal age. *Am. J. Hum. Genet.*, 1983, 35, 301-308.
- van Uum CM, Stevens SJ, Dreesen JC, Drüsedau M, Smeets HJ, Hollanders-Crombach B, Die-Smulders CE, Geraedts JP, Engelen JJ, Coonen E. SNP array-based copy number and genotype analyses for preimplantation genetic diagnosis of human unbalanced translocations. *Eur. J. Hum. Genet.* 2012, 20, 938-944.

- Wells D and Delhanty JD. Comprehensive chromosomal analysis of human preimplantation embryos using whole genome amplification and single cell comparative genomic hybridization. *Mol. Hum. Reprod.*, 2000, 6, 1055-1062.
- Wells D, Kaur K, Grifo J, Glassner M, Taylor JC, Fragouli E, Munne S. Clinical utilisation of a rapid low-pass whole genome sequencing technique for the diagnosis of aneuploidy in human embryos prior to implantation. *J. Med. Genet.*, 2014, 51 553-562.
- Xiong B, Tan K, Tan YQ, Gong F, Zhang SP, Lu CF, Luo KL, Lu GX, Lin G. Using SNP array to identify aneuploidy and segmental imbalance in translocation carriers. *Genomics Data*, 2014, 2, 92-95.
- Ye Y, Qian Y, Xu C, Jin F. Meiotic segregation analysis of embryos from reciprocal translocation carriers in PGD cycles. *Reprod. Biomed. Online*. 2012, 24, 83-90.
- Zahed L, Der Kaloustian V, Batanian JR. Familial complex chromosome rearrangement giving rise to balanced and unbalanced recombination products. *Am. J. Med. Genet.*, 1998, 79, 30-34.

BIOGRAPHICAL SKETCH

Name: Lim, Chun Kyu

Affiliation: Laboratory of Reproductive Medicine, Cheil General Hospital and Women's Healthcare Center, Dankook University College of Medicine, South Korea

Position Title: Associate Professor

E-Mail: seungzzang@paran.com

Education:

Hanyang University, B.S., 1992, Biology

Hanyang University, M.S., 1997, Reproductive Endocrinology

Hanyang University, Ph.D., 2008, Reproductive Endocrinology

A. Positions

1995-2010 - Researcher, MIT, Lab. of Reproductive Medicine, Cheil General Hospital and Women's Healthcare Center

2010-2014 - Assistant Professor, Cheil General Hospital and Women's Healthcare Center, Kwandong University College of Medicine

2014-2015 - Associate Professor, Cheil General Hospital and Women's Healthcare Center, Catholic Kwandong University College of Medicine

2015-present - Associate Professor, Cheil General Hospital and Women's Healthcare Center, Dankook University College of Medicine

B. Selected Peer-Reviewed Publications

- [1] Lim CK, Jun JH, Min DM, Lee HS, Kim JY, Koong MK, Kang IS. (2004) Efficacy and clinical outcome of preimplantation genetic diagnosis using FISH for couples of reciprocal and Robertsonian translocations: the Korean experience. *Prenat. Diagn.* 24 (7), 556-561.
- [2] Lee HS¹, Choi HW, Lim CK, Koong MK, Kang IS, Yoo HW, Choi JH, Jun JH. (2006) Identification of a novel single nucleotide polymorphism of HADHA gene at a referred primer-binding site during pre-diagnostic tests for preimplantation genetic diagnosis. *J. Korean Med. Sci.* 21(5), 794-799.
- [3] Lee HS, Jun JH, Choi HW, Lim CK, Yoo HW, Koong MK, Kang IS. (2007) Preimplantation genetic diagnosis for ornithine transcarbamylase deficiency by simultaneous analysis of duplex-nested PCR and fluorescence in situ hybridization: a case report. *J. Korean Med. Sci.* 22 (3), 572-576.
- [4] Lim CK, Cho JW, Kim JY, Kang IS, Shim SH, Jun JH. (2008) A healthy live birth after successful preimplantation genetic diagnosis for carriers of complex chromosome rearrangements, *Fertil. Steril.* 90(5), 1680-1684.
- [5] Lim CK, Cho JW, Song IO, Kang IS, Yoon YD, Jun JH. (2008) Estimation of chromosomal imbalances in preimplantation embryos from preimplantation genetic diagnosis cycles of reciprocal translocations with or without acrocentric chromosomes. *Fertil. Steril.* 90(6), 2144-2151.
- [6] Lim CK, Kim SK, Ko DS, Cho JW, Jun JH, An SY, Han JH, Kim JH, Yoon YD. (2009) Differential cytotoxic effects of mono-(2-ethylhexyl) phthalate on blastomere-derived embryonic stem cells and differentiating neurons. *Toxicology* 264(3), 145-154.
- [7] Ko DS, Cho JW, Park SY, Kim JY, Koong MK, Song IO, Kang IS, Lim CK. (2010) Clinical outcomes of preimplantation genetic diagnosis (PGD) and analysis of meiotic segregation modes in reciprocal translocation carriers. *Am. J. Med. Genet. A.* 152A(6), 1428-1433.
- [8] Ko DS, Cho JW, Lee HS, Kim JY, Kang IS, Yang KM, Lim CK. (2013) Preimplantation genetic diagnosis outcomes and meiotic segregation analysis of robertsonian translocation carriers. *Fertil. Steril.* 99(5), 1369-1376.
- [9] Park YS, Kim MK, Lim CK, Lee SH, Park DW, Seo JT, Yang KM. (2014) Efficacy of cryopreservation of embryos generated by intracytoplasmic sperm injection with spermatozoa from frozen testicular tissue. *J. Assist. Reprod. Genet.* 31(10), 1331-1336.
- [10] Park YS, Lee SH, Lim CK, Cho JW, Yang KM, Seo JT. (2015) Effect of testicular spermatozoa on embryo quality and pregnancy in patients with non-obstructive azoospermia. *Syst. Biol. Reprod. Med.*, 61(5), 300-306.

Chapter 72

**A FAMILY-BASED ASSOCIATION STUDY BETWEEN
THE BDNF GENE AND ATTENTION DEFICIT
HYPERACTIVITY DISORDER IN MEXICAN CHILDREN
AND ADOLESCENTS**

***Alan López-López¹, Miriam Margot Rivera-Angles¹,
Carlos Alfonso Tovilla-Zárate^{2,*}, Roberto Molina-Solís¹,
Araceli Valencia-Hernández¹, Luis Gómez-Valencia¹,
Thelma Beatriz González-Castro³, Isela E Juárez-Rojop⁴,
María Lilia López-Narváez⁵, Ana Fresan⁶
and Yazmin Hernández-Díaz³***

¹Servicio de Citogenética, Hospital Regional de Alta Especialidad
“Dr. Rodolfo Nieto Padrón”. Villahermosa, Tabasco, México

²Universidad Juárez Autónoma de Tabasco, División Académica Multidisciplinaria de
Comalcalco, Comalcalco, Tabasco, México

³División Académica Multidisciplinaria de Jalpa de Méndez; Universidad Juárez
Autónoma de Tabasco; Jalpa de Méndez, Tabasco

⁴Universidad Juárez Autónoma de Tabasco, División Académica de Ciencias de la Salud,
Villahermosa, Tabasco, México

⁵Hospital General de Yajalón. Secretaría de Salud, Yajalón, Chiapas, México

⁶Subdirección de Investigaciones Clínicas, Instituto Nacional de Psiquiatría Ramón de la
Fuente Muñiz, México, D. F., México

* Corresponding Author's Email: alfonso_tovillaz@yahoo.com.mx (Dr. Carlos Alfonso Tovilla Zárate, Universidad Juárez Autónoma de Tabasco, Ranchería Sur, Cuarta Sección, C.P. 86650, Comalcalco, Tabasco, México; Tel: 52 9933581500 ext. 6901).

ABSTRACT

Aims: Attention deficit hyperactivity disorder (ADHD) is a multifactorial psychiatric and neurobehavioral disorder. The brain-derived neurotrophic factor gene (*BDNF*) has been proposed as a strong candidate for this pathology. The aim of this study was to determine a family-based association between three polymorphisms of the *BDNF* gene and the ADHD in a Tabascan-Mexican population.

Methods: We analyzed the rs6265, rs12273363 and rs11030119 polymorphism of the *BDNF* gene through a family-based association study. A total of 105 individuals grouped in family-trios (mother, father and ADHD patient) were studied. Allelic and haplotypic transmission were assessed through transmission disequilibrium test (TDT), using HaploView software.

Results: No statistically significant association was observed between the *BDNF* gene polymorphisms and the ADHD etiology in Tabascan-Mexican families: rs6265 ($\chi^2 = 1.33$; $p = 0.24$); rs12273363 ($\chi^2 = 1.33$; $p = 0.24$); rs11030119 ($\chi^2 = 0.66$; $p = 0.41$). Furthermore, no preference of transmission was observed for any of the haplotypes.

Conclusions: It was not possible to prove any association between the *BDNF* gene polymorphic variants and ADHD in a Mexican population. Future studies comprising larger samples are necessary to determine the potential role of the *BDNF* gene in ADHD.

Keywords: gene; Brain Derived Neurotrophic Factor (BDNF); Mexican population; Attention Deficit Hyperactivity Disorder (ADHD)

INTRODUCTION

Attention deficit hyperactivity disorder (ADHD) is one of the most common neuropsychiatric diseases in infancy and adolescence, with a higher frequency in men than women [1, 2]. Its world prevalence in the general population is high 3.4% (IC del 95%, 2.6-4.5) [3] and affects between 2 to 10% children in school age [4], of whom according to recent reports [5, 6], 80% continue showing ADHD symptoms through ought their lives. The frequency of its symptoms are catalogued as warning signs of this pathology as they highly affect the social, educational and working contexts of these individuals [7, 8]. There is genetic evidence that consistently supports the polygenic nature of ADHD with a heritability estimated between 75% and 91% [9, 10]. In this context, different alterations in neurotransmission pathways such as the dopaminergic [11], glutamatergic [12] and serotonergic [13] have been associated to the etiology of ADHD [14, 15]. Due to its contribution on the neural development, its role on the pharmacological action and its function on the dopaminergic pathway, literature proposes that the Brain Derived Neurotrophic Factor (*BDNF*) is a candidate gene that participates in the ADHD pathogenesis [16]. The *BDNF* gene is located on chromosome 11 at 11p14.1; at least 122 known polymorphisms have been studied for this gene (snpper.chip.org/bio/find-gene). One of the most studied polymorphisms is the Val66Met (G196A) [17], which has functional effects over the intracellular traffic and the pro-BDNF protein secretion when a Val66 is substituted by Met66 (18). However, the relation between the Val66Met of the *BDNF* and ADHD remains controversial because studies have shown positive and negative associations [19-31]. It should be noted that the majority of the studies about the *BDNF* as a candidate gene have been case-control studies [23, 29-33]; only a few

have evaluated the allele transmission in families [27, 29, 34], which we consider is the main reason for the inconclusive results obtained so far. Therefore, we decided to perform a family-based association study with the aim of comparing the allele and haplotypes transmission of the Val66Met *BDNF* polymorphism (rs6265) and the presence of ADHD. The rs12273363 and rs11030119 polymorphisms were also evaluated.

MATERIAL AND METHODS

Patients and Participants

The study comprised a total of 105 individuals of a Tabascan-Mexican population, grouped in 35 family-trios. Of the 35 probands, 32 were men and 3 were women (average age 7.7 years; age range 4-14 years).

Ethic Considerations

Patients were recruited from the outpatient clinic of the Children High Speciality Regional Hospital “Dr. Rodolfo Nieto Padrón” and the High Speciality Regional Hospital “Dr. Gustavo A. Rovirosa Pérez” in Villahermosa City of Tabasco State, Mexico. After receiving a verbal and written explanation of the study objectives, all participants voluntarily accepted to participate and their parents or legal guardians signed an informed consent to authorize the research. The study was approved by the research and bioethics committee of the Children High Speciality Regional Hospital “Dr. Rodolfo Nieto Padrón”.

Clinical Evaluation

ADHD diagnosis was engaged by a child psychiatrist using diagnostic criteria from the Diagnostic and Statistical Manual of Mental Disorders-Fifth Edition (DSM-V), [35]. All individuals were assessed based on their symptomatology and the information provided by the semi-structured interviews with parents and teachers; only patients that met the DSM-V criteria for ADHD were included in the study. Patients with another comorbid psychiatric condition, patients whose parents belong to different ethnicities and patients who refused the blood sample taking were excluded.

Genotyping Tests

The genomic DNA was obtained via the leukocytes extraction from peripheral blood using the Qiagen N. V. protocol of DNA extraction and purification (Puregene Blood Core Kit C, No. Cat. 15838).

Three *BDNF* gene polymorphisms were genotyped (rs6265 (Val66Met), rs12273363 and rs11030119) through the PCR amplifying technique using TacMan[®] 5' probe assays of Applied

Biosystems®. The primer sequences and their characteristics to amplify the three polymorphisms (rs6265, rs12273363 and rs11030119) are shown in Table 1. The fluorescence intensity was measured with the Applied Biosystems® Fast Real-Time PCR Systems 7900HT equipment and the genotypes were determined through allelic discrimination using the algorithm provided by the manufacturer along with the ABI PRISM® 7900HT SDS software version 2.4. A complete standardization was set following the standards of the Laboratory of Psychiatric and Neurodegenerative Diseases at the National Institute of Genomic Medicine (INMEGEN in Spanish) in Mexico City. The genotyping was performed blind to the clinical condition of the individuals.

Table 1. Main characteristics of the *BDNF* gene polymorphism SNPs rs6265, rs12273363 and rs11030119 (Data obtained from Applied Biosystems®)

SNP ID	Gene	Location	Sequence [VIC®/FAM™]	Polymorphism	[VIC®/FAM™]
rs6265	<i>BDNF</i> -AS	Chr.11:27679916	TCCTCATCCAACAGCTCTTCTATCA[G/A]GTGTTTCGAAAGTGTCAGCCAATGAT	G/A Transition and substitution	VIC: C FAM: T
rs11030119	<i>BDNF</i>	Chr.11:27728102	TTAAGTCACCACTCAGACTTTTCTC[A/G]TAGCAAAAGATCAGATCTCACAACC	A/G Transition and substitution	VIC: A FAM: G
rs12273363	<i>BDNF</i>	Chr.11:27744859	TGAGGCACAGCGATGCTGCAGAAGA[C/T]GGTGGATAGCTCTTAAGTTTCAGAC	C/T Transition and substitution	VIC: C FAM: T

Table 2. Allelic frequency and TDT analysis of the SNPs in the *BDNF* gene

Polymorphism	Alleles	HW	Lower frequency allele	Transmitted allele	T:NT	X ²	p value
Val66Met (rs6265)	G: A	1.0	0.12	C	8: 4	1.33	0.24
rs11030119	G: A	0.3	0.17	G	4: 2	0.66	0.41
rs12273363	T: C	1.0	0.12	T	8: 4	1.33	0.24

T – Transmitted allele; NT – Non-Transmitted allele; HW – Hardy – Weinberg equilibrium; p value – Probabililty value for the hypothesis tes (level of significance $p < 0.05$).

Statistical Analysis

The Hardy-Weinberg equilibrium was ascertained using the Pearson's Chi square (χ^2); all the genetic markers of parents and children genotyped were verified using the same test. The family-based association analysis was established using the Transmission Disequilibrium Test (TDT) [36], included in the HaploView Statistic Program version 4.2 (available on: www.broad.mit.edu/mpg/haploview) [37]; where the haplotypes were built for all the markers and the linkage disequilibrium values were examined with a Lewontin's D' minimum selected value of 0.08. The significance level was set as $p \leq 0.05$ with a 95% confidence interval (CI).

Table 3. Transmission haplotypes for the BDNF gene in ADHD families

Haplotypes block 1	Frequency	T:NT	χ^2	p value
CGT	0.700	17.0: 8.0	3.214	0.073
TGT	0.131	3.8: 8.5	1.826	0.176
CAC	0.112	3.0: 7.8	2.149	0.142
CAT	0.041	2.5: 2.7	0.009	0.922
CGC	0.011	0.3: 0.2	0.017	0.896

RESULTS

The TDT analysis results are shown in Table 2. No statistical significance was observed for the Val(G) allele transmission ($\chi^2 = 1.33$; $p = 0.24$). Similar results were seen for the other polymorphisms: rs11030119 ($\chi^2 = 0.66$; $p = 0.41$) and rs12273363 ($\chi^2 = 1.33$; $p = 0.24$). When the haplotypes analysis were performed, five common haplotypes were observed; however, no statistically significant differences were detected for any of the haplotype-blocks transmission (Table 3).

DISCUSSION

The objective of the present study was to evaluate the family-based association between the rs6265, rs12273363 and rs11030119 polymorphisms of the *BDNF* gene and attention deficit hyperactivity disorder in a Mexican population. Initially, the allelic transmission analysis was performed and the haplotypes analysis subsequently. In our population, no family-based association was observed, as none of the alleles (of the three polymorphisms studied) appeared to be over-transmitted from parents to probands. The result of no association based on families was also observed when haplotypes were analyzed. To our knowledge, this is the first study conducted in a Mexican population that analyses the genetic association between the *BDNF* gene and attention deficit hyperactivity disorder utilizing the transmission disequilibrium test (TDT).

Previous family-based association studies show that there is correlation between the *BDNF* gene and ADHD [23, 26, 30]; however, our results do not evidence a family transmission of the Val66Met (rs6265) polymorphism. Nevertheless, our results are similar to other studies [24, 25, 27-29, 31]; therefore, the rs6265 polymorphism role remains controversial. Furthermore, there are at least four meta-analysis that have evaluated the relation between the *BDNF* rs6265 polymorphism and ADHD and their evidence shows no association [16, 18, 19, 38], two of them specifically evaluated cases and controls [16, 18], while the other two included family-based studies as well as case-control studies [19, 38].

Some differences can explain the wide variation of the results. First, population ethnicities; so far, studies have been conducted on European and Asian populations (Caucasian and Mongoloid); as a consequence, there is high heterogeneity among the populations studied. As an example, the G/A alleles frequency in the Mexican population has been described as 79%

for the “G” allele and 20% for the “A” allele. In the present study, the prevalence observed for the “A” allele was lower than the previously reported (www.hapmap.org). When other populations have been analyzed, the frequency for the “A” allele can increased up to 61% as seen in Asian population; while the European population shows this allele as the least frequent in a range around 19%. Second, sample sizes of the studies that have used the TDT to analyze patients with ADHD have been very different; for instance, our analysis included 35 families, another study with significant values used a sample of 64 families [23], while reports with the largest samples have included 342 and 454 families [26, 29], respectively.

Finally, our “no association” observation was also seen in the haplotypes analysis. Likewise, Lee et. al., (2007) did not observe any correlation of the polymorphic variants rs2049046, rs6265 and rs11030104 in their sample (24); however, they found the A-G-G haplotype as the most frequent in relation with ADHD. On the other hand, Cho et. al., (2010) studied the rs6265 polymorphisms and other molecular markers (rs11030101 and rs16917204) and did not observe any possible risk related haplotype [31]. Overall, the results of the aforementioned studies (including ours), do not show a group of haplotypes involved in the ADHD pathogenesis.

Furthermore, there are just a few reports that have analyzed the rs12273363 and rs11030119 polymorphisms in other neuropsychiatric diseases [39-42]. To our knowledge, the present study is the first one to search for an association between the *BDNF* polymorphisms and ADHD (www.ncbi.nlm.nih.gov) in a Mexican population, though no association was found with any of the three SNPs proposed.

We acknowledge some limitations in our research. First, the small sample size made impossible to perform a sub-analysis by gender and also limits the statistical power of our analyses. Second, we did not evaluate other variables such as the severity of dominant-symptoms of the disorder. Therefore, the negative results found in our study, are not necessarily a refutation of the possible association between the *BDNF* gene and ADHD; the SNPs rs6265, rs12273363 and rs11030119 should still be important for future analyses, maybe for individual associations with ADHD symptoms.

There are also some strengths in our research: First, children with ADHD were evaluated and diagnosed by a child psychiatrist. Second, the association technique used in this research is more robust than the used in case-control studies; family-based association gives more information and one of the features of the transmission disequilibrium test is that it prevents possible spurious associations [43, 44]. Finally, only individuals from Tabasco State were included, which discards possible heterogeneity in the studied population.

CONCLUSION

Our findings suggest no association between the *BDNF* gene polymorphic variants and attention deficit hyperactivity disorder in a Mexican population. However, to determine the potential role of the *BDNF* gene and the development of ADHD we suggest that future studies comprise larger samples.

ACKNOWLEDGMENTS

The authors thank to the outpatient clinic of the Children High Speciality Regional Hospital “Dr. Rodolfo Nieto Padrón” and the High Speciality Regional Hospital “Dr. Gustavo A. Rovirosa Pérez” who collaborated in the recruitment of the patients.

FINANCIAL SUPPORT

None.

STATEMENT OF INTEREST

None.

ETHICAL STANDARDS

The authors assert that all procedures contributing to this work comply with the ethical standards of the relevant national and institutional committees on human experimentation and with the Helsinki Declaration of 1975 revised in 2008.

REFERENCES

- [1] Zulauf C. A., Sprich S. E., Safren S. A., Wilens T. E. The complicated relationship between attention deficit/hyperactivity disorder and substance use disorders. *Current psychiatry reports*. 2014;16(3):436.
- [2] Thapar A., Cooper M. Attention deficit hyperactivity disorder. *Lancet* (London, England). 2016;387(10024):1240-50.
- [3] Polanczyk G. V., Salum G. A., Sugaya L. S., Caye A., Rohde L. A. Annual research review: A meta-analysis of the worldwide prevalence of mental disorders in children and adolescents. *Journal of child psychology and psychiatry, and allied disciplines*. 2015;56(3):345-65.
- [4] Hawi Z., Cummins T. D., Tong J., Johnson B., Lau R., Samarrai W., et al. The molecular genetic architecture of attention deficit hyperactivity disorder. *Molecular psychiatry*. 2015;20(3):289-97.
- [5] Cheung C. H., Rijdsdijk F., McLoughlin G., Faraone S. V., Asherson P., Kuntsi J. Childhood predictors of adolescent and young adult outcome in ADHD. *Journal of psychiatric research*. 2015;62:92-100.
- [6] van Lieshout M., Luman M., Twisk J. W., van Ewijk H., Groenman A. P., Thissen A. J., et al. A 6-year follow-up of a large European cohort of children with attention-

- deficit/hyperactivity disorder-combined subtype: outcomes in late adolescence and young adulthood. *European child & adolescent psychiatry*. 2016.
- [7] Cunill R., Castells X. [Attention deficit hyperactivity disorder]. *Medicina clinica*. 2015;144(8):370-5.
 - [8] Thapar A., Cooper M., Eyre O., Langley K. What have we learnt about the causes of ADHD? *Journal of child psychology and psychiatry, and allied disciplines*. 2013;54(1):3-16.
 - [9] Zhang L., Chang S., Li Z., Zhang K., Du Y., Ott J., et al. ADHDgene: a genetic database for attention deficit hyperactivity disorder. *Nucleic acids research*. 2012;40(Database issue):D1003-9.
 - [10] Akutagava-Martins G. C., Rohde L. A., Hutz M. H. Genetics of attention-deficit/hyperactivity disorder: an update. *Expert review of neurotherapeutics*. 2016;16(2):145-56.
 - [11] Hasler R., Salzmann A., Bolzan T., Zimmermann J., Baud P., Giannakopoulos P., et al. DAT1 and DRD4 genes involved in key dimensions of adult ADHD. *Neurological sciences: official journal of the Italian Neurological Society and of the Italian Society of Clinical Neurophysiology*. 2015;36(6):861-9.
 - [12] Naaijen J., Bralten J., Faraone S., Glennon J., Franke B., Buitelaar J. P. 3.002 Glutamatergic and GABAergic gene-sets in ADHD: association to overlapping traits in ADHD and autism. *European Neuropsychopharmacology*. 2016;26:S47-S8.
 - [13] Thissen A. J., Bralten J., Rommelse N. N., Arias-Vasquez A., Greven C. U., Heslenfeld D., et al. The role of age in association analyses of ADHD and related neurocognitive functioning: A proof of concept for dopaminergic and serotonergic genes. *American journal of medical genetics Part B, Neuropsychiatric genetics: the official publication of the International Society of Psychiatric Genetics*. 2015.
 - [14] Liu D. Y., Shen X. M., Yuan F. F., Guo O. Y., Zhong Y., Chen J. G., et al. The Physiology of BDNF and Its Relationship with ADHD. *Molecular neurobiology*. 2015;52(3):1467-76.
 - [15] Li Z., Chang S. H., Zhang L. Y., Gao L., Wang J. Molecular genetic studies of ADHD and its candidate genes: a review. *Psychiatry research*. 2014;219(1):10-24.
 - [16] Lee Y. H., Song G. G. BDNF 196 G/A and COMT Val158Met Polymorphisms and Susceptibility to ADHD: A Meta-Analysis. *Journal of attention disorders*. 2015.
 - [17] Park S., Kim B. N., Kim J. W., Jung Y. K., Lee J., Shin M. S., et al. The role of the brain-derived neurotrophic factor genotype and parenting in early life in predicting externalizing and internalizing symptoms in children with attention-deficit hyperactivity disorder. *Behavioral and brain functions: BBF*. 2014;10:43.
 - [18] Sanchez-Mora C., Ribases M., Ramos-Quiroga J. A., Casas M., Bosch R., Boreatti-Hummer A., et al. Meta-analysis of brain-derived neurotrophic factor p.Val66Met in adult ADHD in four European populations. *American journal of medical genetics Part B, Neuropsychiatric genetics : the official publication of the International Society of Psychiatric Genetics*. 2010;153b(2):512-23.

- [19] Forero D. A., Arboleda G. H., Vasquez R., Arboleda H. Candidate genes involved in neural plasticity and the risk for attention-deficit hyperactivity disorder: a meta-analysis of 8 common variants. *Journal of psychiatry & neuroscience: JPN*. 2009;34(5):361-6.
- [20] Ribases M., Hervás A., Ramos-Quiroga J. A., Bosch R., Bielsa A., Gastaminza X., et al. Association study of 10 genes encoding neurotrophic factors and their receptors in adult and child attention-deficit/hyperactivity disorder. *Biological psychiatry*. 2008;63(10):935-45.
- [21] Brookes K., Xu X., Chen W., Zhou K., Neale B., Lowe N., et al. The analysis of 51 genes in DSM-IV combined type attention deficit hyperactivity disorder: association signals in DRD4, DAT1 and 16 other genes. *Molecular psychiatry*. 2006;11(10):934-53.
- [22] Oades R. D., Lasky-Su J., Christiansen H., Faraone S. V., Sonuga-Barke E. J., Banaschewski T., et al. The influence of serotonin- and other genes on impulsive behavioral aggression and cognitive impulsivity in children with attention-deficit/hyperactivity disorder (ADHD): Findings from a family-based association test (FBAT) analysis. *Behavioral and brain functions: BBF*. 2008;4:48.
- [23] Lanktree M., Squassina A., Krinsky M., Strauss J., Jain U., Macciardi F, et al. Association study of brain-derived neurotrophic factor (BDNF) and LIN-7 homolog (LIN-7) genes with adult attention-deficit/hyperactivity disorder. *American journal of medical genetics Part B, Neuropsychiatric genetics: the official publication of the International Society of Psychiatric Genetics*. 2008;147b(6):945-51.
- [24] Lee J., Laurin N., Crosbie J., Ickowicz A., Pathare T., Malone M., et al. Association study of the brain-derived neurotrophic factor (BDNF) gene in attention deficit hyperactivity disorder. *American journal of medical genetics Part B, Neuropsychiatric genetics: the official publication of the International Society of Psychiatric Genetics*. 2007;144b(8):976-81.
- [25] Xu X., Mill J., Zhou K., Brookes K., Chen C. K., Asherson P. Family-based association study between brain-derived neurotrophic factor gene polymorphisms and attention deficit hyperactivity disorder in UK and Taiwanese samples. *American journal of medical genetics Part B, Neuropsychiatric genetics: the official publication of the International Society of Psychiatric Genetics*. 2007;144b (1):83-6.
- [26] Kent L., Green E., Hawi Z., Kirley A., Dudbridge F., Lowe N., et al. Association of the paternally transmitted copy of common Valine allele of the Val66Met polymorphism of the brain-derived neurotrophic factor (BDNF) gene with susceptibility to ADHD. *Molecular psychiatry*. 2005;10(10):939-43.
- [27] Tzang R. F., Hsu C. D., Liou Y. J., Hong C. J., Tsai S. J. Family-based association of the brain-derived neurotrophic factor gene in attention-deficit hyperactivity disorder. *Psychiatric genetics*. 2013;23(4):177-8.
- [28] Schimmelmann B. G., Friedel S., Dempfle A., Warnke A., Lesch K. P., Walitza S., et al. No evidence for preferential transmission of common valine allele of the Val66Met polymorphism of the brain-derived neurotrophic factor gene (BDNF) in ADHD. *Journal of neural transmission (Vienna, Austria: 1996)*. 2007;114(4):523-6.

- [29] Thakur G. A., Sengupta S. M., Grizenko N., Choudhry Z., Joobor R. Family-based association study of ADHD and genes increasing the risk for smoking behaviours. *Archives of disease in childhood*. 2012;97(12):1027-33.
- [30] Lanktree M., Muglia P., Squassina A., Krinsky M., Jain U., Macciardi F., et al., editors. A potential role for brain derived neurotrophic factor (BDNF) in adult ADHD. *American Journal Of Medical Genetics Part B-Neuropsychiatric Genetics*; 2004: Wiley-Liss Div John Wiley & Sons Inc, 111 River St, Hoboken, Nj 07030 Usa.
- [31] Cho S. C., Kim H. W., Kim B. N., Kim J. W., Shin M. S., Chung S., et al. Gender-specific association of the brain-derived neurotrophic factor gene with attention-deficit/hyperactivity disorder. *Psychiatry investigation*. 2010;7(4):285-90.
- [32] Aureli A., Del Beato T., Sebastiani P., Marimpietri A., Melillo C. V., Sechi E., et al. Attention-deficit hyperactivity disorder and intellectual disability: a study of association with brain-derived neurotrophic factor gene polymorphisms. *International journal of immunopathology and pharmacology*. 2010;23(3):873-80.
- [33] Friedel S., Horro F. F., Wermter A. K., Geller F., Dempfle A., Reichwald K., et al. Mutation screen of the brain derived neurotrophic factor gene (BDNF): identification of several genetic variants and association studies in patients with obesity, eating disorders, and attention-deficit/hyperactivity disorder. *American journal of medical genetics Part B, Neuropsychiatric genetics: the official publication of the International Society of Psychiatric Genetics*. 2005;132b(1):96-9.
- [34] Mick E., Wozniak J., Wilens T. E., Biederman J., Faraone S. V. Family-based association study of the BDNF, COMT and serotonin transporter genes and DSM-IV bipolar-I disorder in children. *BMC psychiatry*. 2009;9:2.
- [35] Association A. P., Association A. P. *Diagnostic and statistical manual of mental disorders* (DSM). Washington, DC: American psychiatric association. 1994:143-7.
- [36] Spielman R. S., McGinnis R. E., Ewens W. J. Transmission test for linkage disequilibrium: the insulin gene region and insulin-dependent diabetes mellitus (IDDM). *American journal of human genetics*. 1993;52(3):506-16.
- [37] Barrett J. C., Fry B., Maller J., Daly M. J. Haploview: analysis and visualization of LD and haplotype maps. *Bioinformatics*. 2005;21(2):263-5.
- [38] Gizer I. R., Ficks C., Waldman I. D. Candidate gene studies of ADHD: a meta-analytic review. *Human genetics*. 2009;126(1):51-90.
- [39] Ropret S., Zupanc T., Komel R., Videtic Paska A. Single nucleotide polymorphisms in the BDNF gene and suicide in the Slovenian sample. *Neuroscience letters*. 2015;602:12-6.
- [40] Gao Y., Galante M., El-Mallakh J., Nurnberger J. I., Jr., Delamere N. A., Lei Z., et al. BDNF expression in lymphoblastoid cell lines carrying BDNF SNPs associated with bipolar disorder. *Psychiatric genetics*. 2012;22(5):253-5.
- [41] Mercader J. M., Ribases M., Gratacos M., Gonzalez J. R., Bayes M., de Cid R., et al. Altered brain-derived neurotrophic factor blood levels and gene variability are associated with anorexia and bulimia. *Genes, brain, and behavior*. 2007;6(8):706-16.

- [42] Driscoll I., Martin B., An Y., Maudsley S., Ferrucci L., Mattson M. P., et al. Plasma BDNF is associated with age-related white matter atrophy but not with cognitive function in older, non-demented adults. *PloS one*. 2012;7(4):e35217.
- [43] Balding D. J., Bishop M., Cannings C. *Handbook of statistical genetics*: John Wiley & Sons; 2008.
- [44] Kapelski P., Skibinska M., Maciukiewicz M., Pawlak J., Permoda-Osip A., Twarowska-Hauser J. Family-based association study of interleukin 6 (IL6) and its receptor (IL6R) functional polymorphisms in schizophrenia in the Polish population. *Journal of neuroimmunology*. 2015;285:62-7.

Chapter 73

PETROLEUM MICROBIOLOGY AND GENETICS

Joana Montezano Marques, Carla Thais Moreira Paixão,

Gabriel Villar Monte Palma Pantoja,

Artur Luiz da Costa da Silva and Diego Assis das Graças*

Federal University of Pará, Biological Sciences Institute, Center of Genomics and Systems Biology, Laboratory of Genomics and Bioinformatics, Belém, Pará, Brazil

ABSTRACT

Petroleum can be characterized as an oily, flammable substance, less dense than water, with a distinctive scent and color ranging from black to dark brown. The petroleum industry includes some global processes that can be highlighted by the environmental risks they present. At offshore platforms, the danger of a spill is aggravated by the density of the oil, which floats and is carried quickly through sea currents. Thus, in addition to the marine fauna, coastal fauna and flora (e.g., mangroves and estuaries) can also be affected by a leakage. Petroleum is predominantly composed of hydrocarbons, these can be degraded by several species of bacteria, which are of great interest for the bioremediation of contaminated environments. Among these species we can mention *Pseudomonas*, *Sphingomonas*, *Mycobacterium*, *Microbacterium* and *Gordonia*. The success of a bioremediation process depends on numerous factors such as microbial biomass, population diversity, enzymatic activity, pH, temperature, and carbon source. Moreover, the strains have their genetic potential for bioremediation investigated through methods of analysis concerning genes related to the degradation of aliphatic and aromatic hydrocarbons mostly by oxygenases, such as the polycyclic aromatic hydrocarbon ring-hydroxylating dioxygenases (PAH-RHD α), and *alkB* (for n-alkane degradation) genes. The study of autochthonous microbial communities is of crucial importance for the understanding of the genetic and biotechnological potential of bioremediation in environments close to the areas of extraction and susceptible of contamination; it is also of great interest for industry and biotechnological development. This chapter covers the main aspects of petroleum, regarding its exploitation aspects, the process of geological formation, and environmental impacts. It will include topics in microbiology and genetics; for instance, metabolic pathways of hydrocarbon biodegradation, biofilm dynamics

* Corresponding Author's Email: diego.a87@gmail.com.

associated to the oil industry, metagenomics of communities in marine environments, corrosion influenced by microorganisms (CIM), microbial control methods related to the petroleum industry, and bioremediation will be discussed.

1. INTRODUCTION

The importance of oil to the daily lives of the population and global industry is huge, but also carries with it a great risk to the environment. Risks are associated with accidental spills and leaks during the exploration, production, refinement and especially transportation and storage of this product and its derivatives. Large environmental disasters, such as the Exxon Valdez ship and the Deepwater Horizon, wreck the local ecosystem and therefore draw attention to community studies in contaminated areas or possible oil contamination.

Petroleum consists predominantly of hydrocarbons, which can account for 97% of its composition and, to a lesser extent, organic, sulfur, nitrogenous, oxygenated and organometallic derivatives. Due to the predominance of hydrocarbons in petroleum, these are the most used compounds as indicators of this type of pollution, and can be divided into four classes: saturated, aromatic, asphaltenes (phenols, fatty acids, ketones, esters and porphyrins), and the resins (pyridines, quinolines, carbazols, sulfoxides and amides).

The hydrocarbons released into the environment are one of the worst forms of pollution that exists for water, soil, flora and fauna. This is because they are generally very hydrophobic compounds capable of penetrating cell membranes causing mutations, and in humans can cause cancer. These hydrocarbons also accumulate in the trophic chain, making their concentration even greater for humans. Microbiological and genetic aspects will be addressed in the following topics in order to provide a background of the importance of microbes to the petroleum industry and to the environment.

1.1. Formation of Petroleum in Relation to Geological Processes

Petroleum (from Greek: *petra*: “rock” + *oleum*: “oil”) is a mixture of oily substances, flammable, generally less dense than water, with a distinctive scent and coloring that can range from colorless or light brown to black, through to green and brown (brown), naturally occurring, found in geological formations beneath the Earth's surface. A complex mixture of hydrocarbons and varying amounts of non-hydrocarbons, among these molecules we can point out from n-alkanes to aromatic compounds containing sulfur (sulfides) and nitrogen (pyridines and indoles). In addition, the petroleum still has some organometallic constituents such as vanadium and nickel [1]. When it occurs in the liquid state in subsurface or surface reservoirs, it is termed oil (or crude oil, to differentiate from the refined oil. It is known as condensate the hydrocarbon mixture which is in the gaseous state in subsurface and becomes liquid at the surface. The term natural gas refers to the fraction of oil that occurs in the gaseous state or in solution in the oil in subsurface reservoirs [2].

1.1.1. The Origins of Oil

The first theories that tried to explain the occurrence of the petroleum postulated an inorganic origin, from the reactions that would occur in the mantle. Even today there are authors

who advocate an inorganic origin for petroleum, either from the polymerization of methane from the mantle and migrated through faults, or from reactions equivalent to those employed in the Fischer-Tropsch synthesis, and which would find favorable conditions for its occurrence in the subduction zones [3]. Several facts, however, favor an organic origin for most of the hydrocarbons found near the Earth's surface, in particular for those with two or more carbon atoms. Oil is formed in the earth's crust as a result of the partial anaerobic decomposition of organic matter (animal, plant, and plankton) deposited in the bottom of seas and lakes. The product generated from this partial anaerobic decomposition is subsequently transformed under high pressure and at temperatures up to 150°C. Transformation reactions occur at catalytic sites present in the adjacencies of rock surfaces in the presence of water, sulfuric acid, sulfur, and other inorganic compounds [4].

1.1.2. Determinant Factors of the Petroleum's Occurrence in Sedimentary Basins

The most accepted hypothesis considers that, with the increase of the temperature, the molecules of kerogen found in the Earth's underground layers would begin to be broken, generating liquid and gaseous organic compounds, in a process called catagenesis. In order to have an accumulation of oil, it would be necessary to migrate the oil and/or gas through the layers of adjacent and porous rocks, until a sealing rock and a structure That accumulates oil and/or gas in a porous rock called reservoir rock. It is accepted by most geologists and geochemists that it is formed from organic substances from the earth's surface (organic debris), but this is not the only theory about its formation [4].

The formation of an accumulation of oil in a sedimentary basin requires the association of a number of factors:

- The existence of rocks rich in organic matter, called generating rocks;
- The generating rocks must be submitted to the appropriate conditions (time and temperature) for the generation of the oil;
- The existence of rocks with porosity and permeability required for the accumulation and production of oil, called reservoir rocks;
- The presence of favorable conditions for the migration of oil from the generating rock to the reservoir rock;
- The existence of an impermeable rock that holds the oil, called a sealant rock or cape;
- A geometric arrangement of the reservoir and sealant rocks that favors the accumulation of a significant volume of oil.

A commercial accumulation of oil is the result of an adequate association of these factors in time and space. The absence of only one of these factors makes the formation of an oil deposit impossible.

1.1.2.1. Generating Rock

A generating rock must have organic matter in adequate quality and quantity and submitted to the stage of thermal evolution necessary for degradation of the kerogen. It is generally accepted that a generating rock should contain a minimum of 0.5 to 1.0% of total organic carbon (TOC) content. The volumetric aspects of the generating rock (thickness and lateral extension)

should also not be ignored, since a rock with adequate quantity and quality of organic matter can be, for example, too thin to generate commercial quantities of petroleum [5, 6]. The term organic matter refers to the material present in sedimentary rocks, which is derived from the organic part of living things. The quantity and quality of the organic matter present in the sedimentary rocks reflects a series of factors, such as the nature of the biomass, the balance between production and preservation of organic matter, and the physical and chemical conditions of the depositional paleoenvironment [5].

1.1.2.2. Reservoir Rocks

The crude oil formed and the other substances do not remain in the rock on which they are generated, the matrix rock, but they move to the sedimentary basins or reservoir rocks where they accumulate occupying the pores. Layers or porous sheets of sand, sandstone or limestone forms the reservoir rocks. It is the permeability of reservoir rocks that will allow the production of oil. Because it has a lower density than the rocks that make up the subsoil, oil tends to migrate to the surface. When petroleum encounters an impermeable structure on the way to the surface that confines it and prevents its migration, slow formation (a few thousand years) of an oil reservoir will occur [6–8]. In this place are the natural gas, in the highest part, and oil and water in the lowest [7].

1.1.2.3. Conversion of Kerogen into Oil

Three phases are recognized in the evolution of organic matter as a function of temperature increase: diagenesis, catagenesis and metagenesis [8, 9]. The diagenesis occurs after the deposition of the organic matter, under small depths and low temperatures, resulting in the transformation of the original organic matter into kerogen. During diagenesis, methane is the only hydrocarbon generated in significant amounts. During catagenesis, the kerogen is subjected to even higher temperatures (in the range of 50 to 150°C), which results in the successive formation of oil, condensate, and moist gas [9]. The end of the catagenesis is reached at the stage where the kerogen has completed the loss of its aliphatic chains. In metagenesis, achieved under very high temperature (above 150–200°C), the organic matter is represented basically by dry gas (methane) and a carbonaceous residue.

In order to characterize the evolution of the kerogen transformation process in petroleum, two parameters are used: the genetic potential (or potential), defined as the amount of oil (oil and gas) that a kerogen is able to generate, and the rate of transformation, defined as the ratio between the amount of oil generated and the original genetic potential [7, 9]. The original generator potential refers to the kerogen that has not yet been subjected to the catagenesis, that is, whose transformation rate is zero. From the beginning of the catagenesis, the conversion of the kerogen into petroleum causes a progressive increase of the transformation rate associated with the reduction of the generating potential, which is called residual [8].

1.1.3. Petroleum Migration and Accumulation: Primary and Secondary Migration

The process of expulsion of oil from the generating rocks, an essential factor for the formation of commercial accumulations, is called primary migration. Various theories and hypotheses have been proposed in order to explain the mechanisms and factors controlling the expulsion of oil from its generating rock [6, 8, 9]. During primary migration, gas, and oil travel together as a single liquid phase due to the high pressures in the generating rock as these pressures usually become higher than the bubble point pressure. After some time, expulsion

from the matrix rock may occur in pulses due to the closure of pores and fractures from the first expulsion. Gradually, petroleum creates enough pressure to reopen these fractures causing a second expulsion, and the pulses continue to happen as long as the petroleum can rebuild enough pressure to reopen its path out of the rock. After migration, the pressure decreases and the pores close. Lastly, petroleum migrates out of the generating rock and, pressures decay [8, 9].

Among the suggested mechanisms, the migration of oil in solution in water and by molecular diffusion is highlighted. With the advances within this field, it was shown that these mechanisms, although working, do not have the necessary efficiency for the expulsion of significant volumes of petroleum.

The dislocation of the oil between the generating rock and the trap is called secondary migration. It consists of a continuous flow, driven by the fluid potential gradient. This potential can be subdivided into three components: I. pressure unbalance caused by compaction, II. buoyancy, consisting of the vertical force resulting from the difference in density between oil and forming water; And, III the capillary pressure resulting from the interfacial tension between the oil and water phases and the rocks [9]. Finally, it is crucial to mention necessities for favorable petroleum accumulations as these accumulations will supply the world with its needs of petroleum.

1.2. Petroleum Composition

Oil contains hundreds of different compounds. In elementary terms, petroleum consists essentially of carbon (80 to 90% by weight), hydrogen (10 to 15%), sulfur (up to 5%), oxygen (up to 4%), nitrogen and other elements (e.g., nickel, vanadium, etc.). The composition of petroleum is generally described in terms of the proportion of saturated hydrocarbons, aromatic hydrocarbons, and non-hydrocarbons [2].

Saturated hydrocarbons, C and H compounds attached by single bonds, include normal alkanes (normal paraffins or n-alkanes), isoalkanes (isoparaffins or branched alkanes) and cycloalkanes (cyclic alkanes or naphthenes). N-alkanes of less than 5 carbon atoms (methane, ethane, propane, and butane) occur as gas at normal pressure and temperature, while those with 5 to 15 carbon atoms are liquid and those with more than 15 carbon atoms vary from viscous liquids to solids. Most of the normal alkanes present in petroleum have up to 40 carbon atoms. Isoalkanes are present primarily with compounds of up to 10 carbon atoms, although occurring up to 25 atoms. Cycloalkanes may have up to 6 carbon rings, each having 5 or 6 carbon atoms, and cycloalkanes occur mainly in the liquid state [2].

Aromatic hydrocarbons are compounds that have the aromatic ring (benzene) and always occur in the liquid state. They may contain more than one aromatic ring, such as naphthalenes (2 rings) and phenanthrene (3 rings). Toluene, with only one benzene nucleus, is the most common aromatic compound in petroleum, followed by xylene and benzene. Finally, non-hydrocarbons are compounds that contain other elements, other than carbon and hydrogen, called heteroatoms. As the elements nitrogen, sulfur and oxygen are the most common heteroatoms, these compounds are generally known as NSO. It is also common the occurrence of metals (especially nickel and vanadium) associated with organic matter in compounds called organometallic. Resins and asphaltenes are high molecular weight NSO compounds, poorly soluble in organic solvents. Its basic structure consists of 'layers' of condensed polyaromatic

compounds, stacked in the form of aggregates. The proportion of resins and mainly asphaltenes in petroleum is directly proportional to their viscosity [10].

Petroleum hydrocarbons comprise n-alkanes, isoalkanes, cycloalkanes, and aromatics. Among these, the predominant are the n-alkanes and the alkanes with branched chains. These compounds contain amounts of carbon varying from 1 to 78 atoms in some types of oil [11]. The most important branched group is formed by isoprenoids containing 13 carbon atoms, pristane and phytane having 19 and 20 carbon atoms, respectively [12]. The cyclo alkanes, also called the paraffin or naphthenes cycle, are also important constituents, but found in lesser amounts in petroleum. Aromatic naphthenes have saturated and aromatic cyclic structures at the same time. Refined products like gasoline, diesel, lubricating oils, kerosene, fuel oil contain the same compounds as petroleum, but with a range of different boiling points. In addition, in a refining process, such as cracking, there are olefins (alkenes and cycloalkenes), which exist in high concentration in gasoline [13].

There are basically two types of oil classifications. Those proposed by engineers are based on the composition and physical-chemical properties of the oil (density, viscosity, etc.) and are geared to the production and refining areas. On the other hand, the classifications proposed by geologists emphasize the composition, being directed to the origin and evolution of the oil. Among the classifications of geological character, one of the most used is the one proposed by Tissot & Welte (1978) that divides the oils into six types: paraffinic, paraffinic-naphthenic, naphthenic, aromatic intermediate, aromatic-asphaltic and aromatic-naphthenic. The composition of each type reflects the origin, the degree of thermal evolution and the alteration processes to which the oil was subjected.

Another form of occurrence of hydrocarbons are gas hydrates, which consist of ice crystals with gas molecules (ethane, propane and, mainly, methane). Gas hydrates occur under very specific conditions of pressure and temperature, being more common in shallow deposits in the polar regions or in deep waters at various points on the planet. It is commonly refined into various types of fuels. Components of petroleum are separated using a technique called fractional distillation i.e., separation of a liquid mixture into fractions differing in boiling point by means of distillation, typically using a fractionating column [14].

1.3. General Concepts Concerning Oil Industry: Exploitation, Recovery and Refining Process

The petroleum results from anaerobic organic matter degradation that takes around thousands of years to finally form an oil quarry [15]. Although this long-term formation process, petroleum quarries last no longer than 32 years since its discovery until oil well abandonment. The well abandonment occurs in the last phase of a reservoir operational life caused by low costs-benefits in oil/gas production. A premature well closing can also happen and its commonly associated with microbiological presence/activity in the oil fields which leads to adverse risks to environmental and human security [16, 17].

The operational life of an oil well begins with the discovery step, which generally involves prospecting with seismic methods. The prospecting process can suggest the most suitable site for a well drilling, signaling a potential oil quarry location. The seismic reflection method is one of the main tools for oil prospection used by petroleum industries. This method is based on seismic waves emitted by a generator source (any source of mechanical vibrations) that can be

detected by a receptor. The mechanical vibrations can be generated by explosives, air guns (offshore prospecting), weight-dropping (onshore prospecting) among others. The waves that were reflected from the interface between different rocks, with distinct petrophysics composition, are captured by the receptor (recording equipment). The seismic interpretation is based in the relationship between rocks density and seismic waves rate. Since the rocks present different densities, the waves will pass through dense rocks faster when compared with porous rocks. The different reflection coefficient from the surface rocks can provide geologists an advice of a sedimentary basin with a petroleum reservoir rock [18].

The oil well drilling is the most expensive step in the petroleum exploitation process with great economic risks. The well drilling is performed in a 3-8-stage process using a wheel probe. The numbers of stages involved are related to rocks characteristics and well deep. During a stage, the wheel probe drills the rocks surface using drill rotation and the own weight of the drilling column. The rocks fragments are removed by the drilling fluid or mud, a complex mixture of solids, liquid and gas, which is pump injected inside the drilling column and returns to the surface using the annulus of the well (the space between well walls and the drilling column). After that, the drill column is replaced by a steel coating column and the annulus space is concrete fulfilled. Thereafter, a new stage begins with the introduction of a thinner drilling column in the well. The oil well drilling also contents some additional steps as coring, logging and completion. The coring process aims to obtain a real sample from the subsurface rock (core sample) as a representative sample of the formation. This step can predict well productivity. In the following, logging process characterizes an oil well as cost-effective or not. This process consists on the assessment of rock properties using a logging probe that can indicate if there are oil/gas in quantity to warrant the completion process. This last one is an operational set that aims to prepare the well to start oil/gas production with or without fluids injection in a safe and cost-effective manner [19, 20].

It is also important to highlight that during drilling process potential and dangerous accidents can happens such as oil productive well explosion after a blowout process. The blowout is characterized by the uncontrolled flow of the formation fluid in the surface way and it starts as a kick. The kick happens when the well pressure (fluid formation) is higher than drilling mud pressure. Consequently, the formation fluids came from the drilling column pushing the drilling mud to the surface and can result in a fire accident [19].

In the following step, the oil recovery consists basically in three stages. The first stage is called primary recovery or production. In this stage, hydrocarbons are displaced from the reservoir into the wellbore up the surface using natural reservoir energy, such as gasdrive, waterdrive or gravity drainage. In the primary production, the reservoir natural pressure is significantly higher than the bottom hole pressure, which can drive the oil toward the well and up to the surface. Although, as long as the reservoir pressure declines, so does the oil well production. This natural pressure could be increased using artificial lift system such as rod pumps (horsehead pump) among others. The primary recovery stage ends when the natural pressure is too low that oil production is not economical or when water/gas production is too high in the production stream. The primary production provides only a small percentage (around 15%) of the initial hydrocarbons present in the reservoir [21, 22].

The next stage in oil recovery is called secondary recovery or production. This secondary stage aims to maintain reservoir pressure by the injection of an external fluid into the reservoir to displace the oil towards the wellbore. The secondary recovery usually employs water or gas to enhance reservoir pressure. These technics are named as waterflooding and gas injection,

respectively. In this stage, the injection fluid is injected into the reservoir by the injection wells and the oil is collected in the end of this process in the production well. The injection and production wells have fluid communication under the rock surface. In this process, the gas is usually injected into the gas cap and water is injected into the production zone to displace hydrocarbons from the well. The secondary recovery finishes when oil production is not economical anymore and the injected fluid is produced in a considered amount. The secondary production along with primary production can recover up to 60% of the initial oil in place [18, 22].

The third and last stage in oil recovery was traditionally named as tertiary recovery. The tertiary recovery is characterized as a set of methods used to enhance oil recovery that originally followed primary and second production process. Nowadays, these enhanced methods can be applied at any time during the hydrocarbons production process and the term tertiary recovery became outdated. Therefore, the term enhanced oil recovery (EOR), also known as improved oil recovered, is more commonly used to specify the use of sophisticated technics that aims not only to maintain reservoir pressure, but also to promote hydrocarbons displacement with fluid flow inside the oil well. For this purpose, EOR technics alter oil original properties and they are based in three operational major types: i. the chemical flooding (alkaline or micellar-polymer approaches); miscible displacement (carbon dioxide and hydrocarbon injection); and thermal recovery (steamflood or in situ combustion). Some reservoirs conditions (temperature, pressure, depth, permeability, rock porosity among others), residual oil and water saturations inside the reservoir, and petroleum properties such as oil API gravity and viscosity are main factors that can influence the optimal application of each type of EOR [18, 23, 24].

The EOR technics can also include the use of microorganisms to improve oil recovery in an oil well. In this case, this technic is known as microbial enhanced oil recovery (MEOR). The MEOR technics involve basically target microorganisms and/or their metabolic products (gas, acid, solvents, polymers, biosurfactants and bioemulsifiers) injection into the reservoir, or nutrient injection to promote this target microbiota development inside the well. A variety of mechanisms can be considered when MEOR is applied to enhance oil recovery. These mechanisms used by microorganisms to improve oil production are summarized in Table X. Despite of MEOR mechanism involved, these mechanisms are based on oil viscosity decreasing, enhance of fluid mobility and oil displacement from the reservoirs rocks, and enhance of oil solubility or miscibility in water. The MEOR consists of an interesting green and economical alternative for EOR technics that includes complex technics with high energy uptake (thermal recovery) and financial costs (chemicals improvement) that can lead to environmental impacts [25, 26].

After the oil extraction, the fluids go to a tank for oil/water separation and then, the oil is transported from onshore and offshore platforms basically by pipelines and oils tankers to petroleum refiners. The crude oil is a complex mixture of hydrocarbons which requires refining to be converted in petroleum derivatives such as liquified petroleum gas (LPG, propane and butane) gasoline, diesel, kerosene (jet fuel), naphtha, heating oil, base oil (lubricants) and asphalts (bitumen). The oil by-products are generated from three major refining processes: separation, conversion and treating [27–29].

Table 1. The microbial enhanced oil recovery (MEOR) main mechanisms used by microorganisms to improve oil production

Bioproduct	Example	Activity
Gas	Carbon dioxide and methane	Preferential solubility in oil
Acids	Acetic, butyric and lactic acids	Enhance porosity of reservoirs rocks
Solvents	Ethanol, butanol and acetones	Affect oil/rock interactions
Biosurfactants	Glycolipids and lipopeptides	Decreases surface tension between oil/rock and oil/water
Biopolymers	Polysaccharides and proteins	Controls oil physical mobility/Bioemulsifiers
Microbial Biomass	Anaerobic and aerobic microorganisms	Hydrocarbons degradation/Biofilm formation

The separation step consists in hydrocarbons segregation in a distillation column. Inside the column, the oil is heated with high temperatures (around 350-400°C) leading hydrocarbons to vaporize according to their molecular weight. During this process, small chains hydrocarbons vaporize first while heaviest molecules remain at the bottom of the distillation column without vaporizing. In the following, the molecules condense into liquids according to the temperature inside the column and petroleum fractions or cuts are collected. In the end of the separation process, the distillation column presents highly viscous hydrocarbons like asphalt (bitumen) and gases at the column bottom and top, respectively. The conversion step aims to “crack” heavy molecules, remaining from the previous step, in to lighters products to reach the market demand. This step also applies high temperatures together with a catalyst to speed up heavy products conversion into gas, gasoline and diesel. The final refining step is known as treating and it consists on remove or reduces significantly corrosive and potential polluting molecules from refined products, especially the sulfur. As an example, the diesel desulfurization or sulfur removal occurs at high temperatures and pression with hydrogen addition. The hydrogen combined with the sulfur forms hydrogen sulfide (H₂S) which is treated and removed. Another refine products as kerosene, butane and propane are treated with caustic soda (sodium hydroxide) to remove thiols (mercaptans) in a sweetening process. The automotive fuels steel needs a treatment to produce high-octane products. This treatment, known as catalytic reforming, involves chemical reactions at high temperature and pressure using platinum as a catalyst. The catalytic reforming promotes naphthenic hydrocarbons conversion into higher-octane rate hydrocarbons, as the aromatics ones. Another process that increases octane rating is the alkylation, one of the main chemical reactions used in the petrochemical industry [27–29].

1.4. Petroleum and the Environment: Pollution and Environmental Impacts

Oil is a naturally occurring substance, its presence in the environment is not necessarily the result of human intervention, such as accidents and extraction, refining and combustion. Phenomena such as exudates are examples of areas where oil affects the environment without the man's involvement. Regardless of the source, the effects of oil are similar when released into the environment. In the last century, there has been a large increase in global pollution. Industrial development, population growth, urbanization, and unconcern with the ecological consequences of the dumping of chemical compounds in nature have contributed to the current level of pollution and, consequently, environmental issues. A large number of industries, especially the oil and gas industries, have contributed significantly to the environmental pollution reaching the present dangerous proportions. In addition, there has been a drastic increase in the diversity of industrially produced organic compounds that have been inconsistently discarded in the environment. Due to this fact, there are now innumerable chemical contaminants that are toxic to biological systems from both natural (biogenic and geochemical) sources and anthropogenic sources [30].

1.4.1. Acidification of the Oceans

Ocean acidification is the name given to the chemical imbalance by decrease in pH in the oceans caused by the increase in atmospheric carbon dioxide (CO₂), meaning an increase in acidity. Since the beginning of the 19th century, when carbon emissions began to rise rapidly, the pH of the ocean surface decreased by about 0.1 on the logarithmic pH scale. Although this difference seems small in reason of the type of scale used, it represents an increase of about 10% in the concentration of H⁺ hydrogen ions, the direct ones responsible for acidification. The rise of CO₂ levels in the atmosphere has its origins in human activities, especially in the burning of fossil fuels (oil, natural gas, coal and, others), but also in deforestation, in some industrial processes and in other smaller sources. Ocean acidification inhibits all marine life - has a greater impact on smaller organisms and then affects larger organisms [31].

Contamination may result from the deliberate dumping and discarding of hazardous materials, from accidental spills, and from the migration of hazardous substances from spills occurring elsewhere. Soil pollution can lead to the contamination of underground aquifers as the contaminant begins to percolate over time through rainwater paths in the soil leading to the water table. Contamination of groundwater can affect at least two million people in the world who depend directly on aquifers for drinking water [30].

1.4.2. Oil Spills

Soil and water contamination represents a serious problem to human health and to the affected area's ecology [30]. Since the most toxic compounds are the most soluble and volatile components, the chemical impact is greatest in the first few days after the spill in most cases. The chemical effect or impact is associated with the toxicity of the compounds present in the oil. Saturated hydrocarbons have anesthetic and necrotic effects on living beings. Alkanes, popularly known as paraffins, and which account for much of the crude oil, may also cause anesthetic and narcotizing effects [32]. Low molecular weight hydrocarbons (C₁₂ to C₂₄) have an intense acute toxic effect, mainly due to their high solubility and consequent bioavailability [33].

The contact of organisms with toxic fractions of the oil can lead to death by intoxication, especially associated with the fractions of aromatic compounds. Among the most toxic components are benzene, toluene, and xylene. Aromatic hydrocarbons having multiple benzene rings are known as Polycyclic Aromatic Hydrocarbons (PAHs). It is known that these compounds are more resistant to the microbiological biodegradation and quite persistent in the environment, being strongly adsorbed in the sediments. Some common examples of PAHs and derivatives in petroleum are naphthalene, anthracene, phenanthrene and benzopyrene and their various isomers. PAHs are especially toxic, potentially carcinogenic to humans, and the tendency of these compounds to be incorporated into fatty tissues and to cause damage to organs such as the liver and kidneys of humans has been proven [34]. Its mutagenic activity is strongly related to its molecular structure. The molecular form of PAH isomers, therefore, is directly related to biological activity and, consequently, to its toxicity [35].

The environmental impacts caused by the petroleum industry come from the exploration, transportation, refining and use of its products [32–35]. Crude oil and refined fuel spills from tanker ship accidents have damaged natural ecosystems in Alaska, the Gulf of Mexico, the Galapagos Islands, France, and many other places. These activities are involved in 90% of the accidents resulting from the exploration and use of oil, and only the remaining 10% are attributed to oil spill catastrophes in the high seas resulting in contamination of coastal regions [34,35]. As an example of an environmental catastrophe, we can mention the case of the Exxon Valdez oil tanker, which in 1989 spilled about 50,000 m³ to 150,000 m³ of oil in the Alaska Sea [36]. As a result of the spill, thousands of animals died in the following months, among them: sea birds, otters, eagles, killer whales, and billions of salmon eggs. In France, March 1978, the oil tanker Amoco Cadiz ran aground on Portsall Rocks, 5 km (3.1 mi) from the coast of Brittany, and ultimately split in three and sank, the consequences from this accident resulted in one of the largest oil spills of its kind in history to the present date.

The consequences of industrial pollution for human health and the environment have led to public demands for the cleaning and restoration of the environment, as well as the creation of governmental regulations and legal actions. The result has been the search for remediation technologies that can be actively applied to cleaning soil and aquifers [31].

1.5. The Microbiome of Oil Reservoirs

Oil reservoirs are characterized by high pressure, high osmosis (salinity or hydrophobicity) and high temperature, and these factors were considered impossible for a long time, being oil considered sterile [37]. Besides the harsh conditions, oil reservoirs contain inorganic ions such as iron, sulfate, nitrate and organic compounds from petroleum (see section 1) and derived from microbial activity [38].

The first microbiological study describing bacteria in oil-reservoirs was published by Bastin in 1926, and he raised a question: the bacteria found in oil-producing wells could be considered indigenous or not? [39]. Indeed, is very difficult to access the indigenous community without to contaminate with exogenous microbes, but it is probably dominated by anaerobic ones, once the deep subsurface has no oxygen [40].

Technical advances of the molecular biology as well as microbiology shone light on the microbial diversity of extreme environments and it made possible to detect and sometimes isolate microbes from oil-producing reservoirs that were able to grow at those challenging

conditions [1, 37, 40–43]. Several studies have been done to assess not only the microbial diversity of oil reservoirs, but (i) how these communities react to physical-chemical changes, (ii) hydrocarbon degradation mechanisms, (iii) microbial steel corrosion in oil-producing wells [1, 40, 43–45].

Many factors can affect microbial composition, such as temperature and pH and biological interactions, kind of reservoir (onshore, offshore, freshwater, seawater, surface soil) [40]. Temperatures above 82°C are considered the limit for growth on oil, because biological molecules denature at this temperature. However, some hyperthermophilic exogenous microbes growing at 103°C were isolated from oil reservoir contaminated with seawater [40]. The pH of an oil reservoir ranges from 3 to 7 and pressure can be up to 500 atm. These conditions can limit the microbiological activity.

The absence of oxygen at deep subsurface and availability of sulphate, carbonate, CO₂ and H₂, led anaerobes to realize sulphate reduction, methanogenesis, acetogenesis and fermentation process. In fact, oil reservoir microbiome is dominated by heterotrophics, thermophiles and lithotrophics bacteria, and methanogenic archaea.

Sulfate reducing microbes oxidize organic compounds reducing sulfate (SO₄²⁻) into sulfide (H₂S). The main taxa found in oil reservoir are *Desulfovibrio*, *Desulfobacter*, *Desulfotomaculum*, *Desulfobacterium*, *Desulfomicrobium*, *Thermodesulfobacterium*, *Thermodesulforhabdus* bacterial genera and *Archaeoglobus* archaeal genus. Mesophilic sulfate reducing bacteria (SRB) have optimum growth temperature around 37°C, while thermophilic SRBs grow better at 60°C, but tolerates up to 80°C. Both mesophilic and thermophilic can oxidize a wide range of substrate such as acetate, lactate, pyruvate or even C17 organic compounds [40].

Methanogenic archaea are the only microorganisms able to produce methane from simple carbon compounds (C₁-C₄). In oil reservoirs, methanogens reduce CO₂ to CH₄ using H₂ as electron donor, the so-called hydrogenotrophic pathway, but sometimes acetoclastic (using acetate instead of CO₂) route can be used [46]. Methanogens can tolerate salinity of up to 30%, and temperature up to 80°C [40].

Fermentative bacteria can also be found in oil-producing wells, among them many bacterial genera able to use carbohydrate to produce a wide range of acids, alcohol and gases under anaerobic conditions. Many halotolerant mesophilic and thermophilic fermentative bacteria have been reported, such as *Haloanaerobium*, *Spirocheta*, *Geotoga*, *Anaerobaculum* are examples.

Mesophilic and thermophilic iron reducing bacteria can be detected in oil field fluids due to the presence of metals in the apparatus of oil extraction [47]. *Shewanella putrefaciens* (formerly *Alteromonas putrefaciens*) is a bacterium that can reduce sulphur, sulphite and thiosulphate into sulphide using H₂ and Fe [47].

1.6. Petroleum Degrading Genes and Metabolic Pathways

As seen on the previous topic, oil reservoirs are characterized by harsh conditions for life and the microbial activity is low. However, after petroleum is extracted from wells, it is contaminated with exogenous microbes (mainly from seawater or marine sediments) whose begins to degrade hydrocarbons. Microbial degradation of petroleum hydrocarbons depends on

the nature of hydrocarbon and physical-chemical conditions such as availability of oxygen, nutrients, temperature, salinity and others [48].

Biochemical mechanisms for hydrocarbon degradation is well described on the literature [48–54]. Under aerobic conditions, oxygen is introduced on hydrocarbons by mono- and dioxygenases and is faster than the anaerobic degradation. Mono and dioxygenases add oxygen atoms to hydrocarbons producing fatty acids that go to β -oxidation and generate acetyl-CoA for the tricarboxylic acid (TCA) cycle. At this process, NADH and FADH are the electron donors [52]. After TCA cycle, carbon intermediary compounds are assimilated for biosynthesis and cell growth [48].

Anaerobic degradation needs to insert fumarate into hydrocarbon molecules and uses inorganic ions such as sulfate, nitrate and iron as terminal electron acceptors [52], resulting in fatty acids that go to β -oxidation producing acetyl-CoA or benzoyl-CoA [48]. Some aromatic compounds, such as benzoate, can be degrade by adding coenzyme A producing acetate and propionyl-CoA (addressed to TCA cycle) [48].

Methane is a hydrocarbon that can be degraded aero and anaerobically. Under aerobic conditions is oxidized to CO_2 by methylophilic bacteria (able to use methane and other C1 compounds). Methanotrophic bacteria use enzymes called methane monooxygenases to oxidize methane, which is the sole source of carbon and energy of this bacterial group. In marine sediments, methane can be oxidized without oxygen by a consortium of sulfate-reducing bacteria (SRBs) and archaea from ANME (anaerobic methanotroph) group. ANME archaea oxidize methane by using it as electron donor, which go to SRBs to reduce SO_4^{2-} to H_2S [55].

As discussed on the beginning of the chapter, petroleum is a heterogeneous compound, which requires several classes of genes and metabolic pathways to be degraded. Environmental conditions such as pH, oxygen availability, nutrient (mainly N and P), temperature and salinity can affect petroleum degradability and enzyme activities [56, 57]. The main microbial enzymes responsible for petroleum degradation are summarized on the table below [48, 58].

Table 2. Main microbial enzymes classes related to petroleum compounds degradation

Enzyme class	Substrate range	Hydrocarbon	Example of bacterial genus
Soluble methane Monooxygenases (sMMO)	$\text{C}_1\text{-C}_8$	Alkanes, alkenes and cycloalkanes	<i>Methylococcus</i>
Particulate Methane Monooxygenases (pMMO)	$\text{C}_1\text{-C}_5$	Alkanes and cycloalkanes	<i>Methylococcus</i>
Alkane hydroxylases (mostly AlkB)	$\text{C}_5\text{-C}_{16}$	Alkanes, fatty acids, alkyl benzenes, cycloalkanes	<i>Pseudomonas</i>
Bacterial P450 oxygenase system	$\text{C}_5\text{-C}_{16}$	Alkanes and cycloalkanes	<i>Acinetobacter</i>
Dioxygenases	$\text{C}_{10}\text{-C}_{30}$	Alkanes and polycyclic aromatic	<i>Acinetobacter</i>

The AlkB protein is the most studied alkane hydroxylase, whose function is to catalyze terminal hydroxylation of alkanes, generating n-alkanol and it requires iron and oxygen [59]. Other genes (*alkFGHJKL* cluster) are necessary to complete alkane degradation to fatty acids and gene order and regulation is quite diverse in natural environments [59–61]. The presence of transposase genes into the cluster is reported in some studies and indicates the spreading of gene cluster in nature and shows opportunities for evolutionary events [61].

1.7. Biofilms in the Petroleum Industry

The metabolic pathways in petroleum context go far beyond the degrading pathways showed in last topic. The genetic interactions concerning life around petroleum metabolism is indeed considered noble from the biotechnology point of view towards the development in bioremediation strategies. However, there is second class of high-level of importance microbial community: Biofilms [62, 63].

Biofilms are complex matrixes of microorganisms that adhere to each other over a surface, usually recruiting extracellular polymeric substances (EPS), so that layers within the biofilm with different patterns of gene expression are formed. In industry, the development of biofilms control strategies, because of the enhanced quorum-sensing activity, and the complexity of gene expression of over the film layers. Antibiotic resistance and toxic compound production, are some of the risks that biofilms can perform hindering efficiency of bioprocesses [64].

The petroleum industry encompasses the biofilm termination in two main categories: (i) those who promote novelties in bioremediation, usually related to bioprospecting of bacteria with the potential to be immobilized in supports for oil degradation, and (ii) the biofilms that hinder petroleum extraction, and cause structural damages to oil rigs, and by-products of petroleum [38].

Sulfate reducing bacteria (SRB), described initially in the last topic, are considered a major challenge in the industry of petroleum. Currently there are over 220 described species (most of the *Desulfovibrio* genus), mostly gram-negative as well as absolute anaerobes, that have been isolated from petroleum oilfields. Thus, they use sulfate (or other sulfur compounds, e.g., sulfites, thiosulfate, tetrathionate) in the respiration producing sulfide [62]. To that end, sulfide ions generated from reduction of sulfate reacts with water, producing hydrogen sulfide by SRB causes corrosive damages over the inner surface of oil ducts, generating one of the major issues off shore petroleum scene [38].

Communities of SRB have been described as one of the hardest to study. Since the environmental conditions of oil reservoirs usually are extremely challenging to be reproduced in the laboratory, microorganisms are often absolute anaerobes, the temperature and pressure in the reservoir are extreme, salinity and pH are. Considering that, novelties in molecular biology making use of 16S rRNA, sulfate reductase (*dsrAB*) genes, and metagenomic approaches have been fundamental to the understanding of the community diversity, composition, and genetic profile [16]. I.e., New techniques have demonstrated how different species of genus as *Desulfobacterium*, *Desulfotomaculum*, *Desulfobacterium*, *Desulfomicrobium*, and *Desulfobulbus* correlate with temperature, salinity, depth, and the concentration of other chemical compounds in the oil mixture, as acetate, propionate and sulfur levels [65].

1.8. Metagenomics for Marine Microbial Communities

The concept of metagenomics was defined by Jo Handelsman [66] as “the genomic analysis of a population of microorganisms” and encompasses molecular biology methods to directly access the genetic material of microbes from environmental samples [67]. The main advantage of using a metagenomic approach is to overcome the culture limitation, since only the minor (0,1-1%) part of the microbes are cultivable using the methods we know.

In the beginning of the 2000's, two main approaches for metagenomic analysis were preferred, and involved the total DNA extraction from samples, followed by (i) PCR, cloning and sequencing – the so-called amplicon approach, mostly the 16S rRNA gene - or (ii) fragmentation, cloning and sequencing – the shotgun approach. Using those methods, molecular microbiologists could (and still can) get a better understanding about microbial diversity and their interaction with the environment, host, human, plant and others.

In 2004, 454 Life Sciences (being acquired by Roche Diagnostics in 2007) has developed the first massively parallel DNA sequencing technology that were able to sequence thousands of DNA fragments in a single run, and it has eliminated the necessity of cloning. Several platforms for DNA sequencing were released after the Roche 454 being known as next generation sequencing (NGS) technologies and today we can achieve tens of billions of base pairs in single run that takes a few days long. The main characteristics of the most popular NGS platforms are shown on Table 3.

Combining the metagenomic approach with NGS, the microbial communities can be deeply analyzed in any environment. Marine environment covers approximately 70% of the Earth surface and hosts a huge amount of microbes ($> 3 \times 10^{30}$) and these microbes consists of bacteria, archaea, protists and fungi. Marine microbial communities play a key role on the global biogeochemical cycles of important elements, such as carbon, oxygen, nitrogen, phosphorus and sulphur [69]. Many studies have been made to determine which taxons are present in marine environment (mostly using 16S rRNA as phylogenetic marker), how they are distributed in space and time and the reaction to physical and chemical changes.

The marine microbial communities are represented by (i) phototrophic (in the upper layers), (ii) chemotrophic primary producers, (iii) denitrifiers, (iv) nitrifiers, (v) nitrogen fixers, (vi) archaea, (vii) heterotrophs. Photo and chemotrophic microbes obtain biomass from carbon dioxide (CO_2), but phototrophics obtain energy from light and chemotrophics from chemical reactions. Besides them, heterotrophic microbes can also be found, but they obtain biomass from carbon reduced and recycling nutrients. Denitrifiers, nitrifiers and nitrogen fixers are involved in the nitrogen biogeochemical cycle, where atmospheric nitrogen (N_2) is fixed by bacteria and is transformed to ammonia, nitrite and nitrate; ammonia can be oxidized into nitrite and nitrate by nitrifiers, and nitrate can be reduced to N_2 by denitrifiers. Viruses, protists, grazers and other organisms less abundant also compose the marine biomass, but are not the focus of this chapter.

It is known that microbial communities changes over time, space and environmental conditions. Fuhrman and colleagues have considered microbial changes occurring in four timescales – (i) hours, (ii) daily to weekly, (iii) monthly to seasonal and (iv) interannual. In hours, rare taxa can become abundant on the sea surface due to day-light cycle, to light, to rapidly cell death and maybe another reasons not yet documented but hypothesized [70]. Daily to weekly timescale can reveal how microbial communities reacts to environmental conditions (and changes), such as virus infection, weather, interaction with other microbes or larger

organisms, or even phytoplankton blooms. Solar angle, seasonal changes and nutrient availability influence microbial changes from monthly to seasonal periods. Interannual changes are caused by anthropogenic actions such as overfishing, global warming and pollution, as well as environmentally natural changes (e.g., Pacific Decadal Oscillation).

Table 3. NGS characteristics. Adapted from Braga and colleagues [68]

NGS Platform/Company	Read Length	Throughput
3730xl DNA Analyzer/Life Technologies (http://www.lifetechnologies.com/)	Up to 900 bp	690 – 2100 Kbases
454 GS FLX System/Roche (http://www.454.com/)	Up to 1000 bp	700 Mb
Illumina HiSeq 2500/Illumina (http://www.illumina.com/)	2 x 150 bp	150 – 180 Gb (Dual Flow Cell)
ABI SOLiD 5500xl/Life Technologies (http://www.lifetechnologies.com/)	75 bp (Fragments)	Up to 250 Gb
	75x35 bp (Paired-end)	
	60x60 bp (Mated-paired)	
Ion Torrent Personal Genome Machine/Life Technologies (http://www.lifetechnologies.com/)	Up to 400 bp	Up to 100 Mb (314 chip v2)
		Up to 1 Gb (316 chip v2)
		Up to 2 Gb (318 chip v2)
PacBio RS System/Pacific Biosciences (http://www.pacificbiosciences.com/)	Over 40,000 bp	Up to 1 Gb

Metagenomic has aided to get knowledge about the genetic diversity of marine microbes. A remarkable work made by Venter and colleagues (2004) estimated 1800 genomic species on the Sargasso sea, including 148 previously unknown bacterial phylotypes and 1.2 million of previously unknown genes [71]. Thirteen years late, many discoveries were done and we are able to couple metagenomics with other techniques like FACS (Fluorescent-Activated Cell Sorting), MDA (Multiple Displacement Amplification) to analyze single cell genomics (SCG) of marine microbes as reported in some studies [72, 73].

The use of SGC techniques is shining light on the called “microbial dark matter” (MDM), which is the major portion of the microbial diversity we are not able to cultivate and consequently do not know the function in the environment. Rinkle and colleagues (2013), on a notable article, have reported a novel amino acid use for the opal stop codon, an archaeal-type

purine synthesis in Bacteria and complete sigma factors (proteins that helps transcription) in Archaea similar to those in Bacteria [73].

Evolution is responsible for the great microbial diversity and the biotechnological potential of their enzymes are still under explored. The biotechnological potential of marine microbial enzymes is focus of important research areas. Proteases, lipases, polysaccharide-degrading and extremozymes (derived from microbes living on extreme environments) have been isolated from marine bacteria and archaea [74]. Biosurfactants and enzymes able to degrade hydrocarbons from petroleum are considered one of most searched (macro) molecules on the nature for its potential use in bioremediation process [44]. There are more than 79 bacterial genera able to use hydrocarbons as sole source of energy and carbon, many of them uncultured, and metagenomics can be a powerful tool if coupled to other techniques to get knowledge about these enzymes and bioproducts.

1.9. Microbiologically-Influenced Corrosion (MIC)

Corrosion can naturally occur in natural and artificial environments. Materials made by pure or mixed compositions of metals chemically decay in a process called oxidization. Corrosion is a electrochemical process which involves a anodic (metal 'A' ionization) reaction and a cathodic (metal 'B' reduction) reaction. Those reactions could be influenced by microbial action, primarily when microbial colonies arrange themselves in biofilms. It is well known that biofilm formation is one of the most problematic interferences caused in technological systems. The accumulation of growing biofilms in a given system that causes corrosion and further material decay is calling biofouling [75].

The microbial activity over the metal and semi-metal species is most catalyzed by many metabolism products such as enzymes, organic and inorganic acids, volatile compounds and exopolymers, which all potentially have influence on anodic/cathodic reactions over the exposed material. Those metabolites alter the electrochemical balance in the interface biofilm/metal leading to a gradual and constant corrosion. The corrosion lead by microbial action is known as biocorrosion or Microbiologically-Influenced Corrosion (MIC) [76]. The MIC is commonly due to bacterial accumulation, but can occur by fungal and algae as well [77].

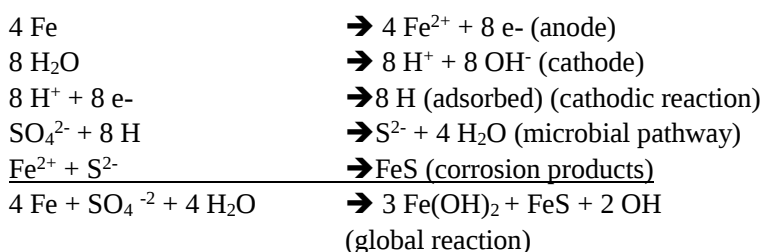
The major organisms frequently involved in MIC processes are the sulfate reducing bacteria, the sulfur oxidizing bacteria, iron reducing/oxidizing bacteria, manganese oxidizing bacteria, organic acid and exopolimers synthesizing bacteria. Since 1930, the sulfate reducing bacteria is studied due to investigation about the corrosion of many metals and semi-metals. Also, the corrosion action of this class of bacteria is spread through land and water environments under aerobic and anaerobic conditions. Many models have been proposed in order to elucidate the mechanisms under such different corrosion processes [76].

The biofilm formation induces an oxygen gradient that can explain the oxygen concentration in the aerobic zones of a microbiologically-induced corrosion. When the MIC is occurring under anaerobic process at neutral pH, it is common to find sulfur reducing bacteria in the biofilms [78]. According to Aguiar [79] suggests that sulfur reducing bacteria are using three major pathways. The first pathway is a naturally occurring due to metabolic needs, whereas those bacteria use their surface hydrogenases to consume the cathodic H₂ film as a consequence of the electrons derivate from the metal ionization from the anodic end. In normal

conditions, the hydrogen formed in the cathodic end can be absorbed to the material resulting in polarization which slows the corrosion process. However, in the presence of sulfate reducing bacteria, the cathodic end is depolarized which increases the hydrogen consume locally, the reaction is shifted to deionized even more water molecules, resulting in increasing Fe^{2+} in solution, and so the metal decay reaction speeds.

The second pathway is also a direct way to corrosion. It occurs due to the metabolic products expelled by those bacteria, such as H_2S (hydrogen sulfate) which reacts with Fe^{2+} forming FeS (iron sulfate) which increases the H^+ concentration and so stimulating the electron consume in the cathodic end. In the case of biofilms, the reposition of hydrogen ions in the metallic surface can be seen as a cathodic repolarization of this surface which protects it from the corrosion. The third pathway is an indirect method occurring by the generation of oxygenized zones over the metal surface where specific metabolites are deposited [79]. Some studies argue that none of the pathways cited above are predominant over one another due to the quantity of factors involved in the process as a whole [80, 81].

The following electrochemical reactions describing the CIM process induced by sulfate reducing bacteria were suggested by van Wolzogen Kuhr and van der Vlugt in 1930 [82], as well as the global reaction describing the final process of depolarization of the cathodic end [77].



The corrosion of steel by sulfate reducing bacteria has a very characteristic pattern called pitting with pits, or small corrosion holes where the iron sulfate is deposited over time [83]. The characteristics of a corrosion pit are dictated by the physical and chemical corrosion products involved in the process after the oxygen has accessed this environment [81]. The general characteristics of corrosion process can be more related to the metabolic state of the bacterial cells over the total bacterial biomass adhered to the biofilms (9 - Beech et al., 1994). As described by Pedersen (1988) [78], the microbial species and the carbon source absorbed by these species are also very important parameters for regulation of the CIM process. As well, the exopolymers have their hole in the corrosion process since they allow the bacterial cells to group together irreversibly to the metal surface and provide protection to the bacterial cells against foreign hazardous molecules and other possible competitors [84].

Besides the fact that sulfate reducing bacteria are involved in corrosion process in petroleum extraction fields by the formation of biofilms and synthesis of corrosive metabolites, other group of bacteria needs to be acknowledged for contributing direct or indirectly for the corrosion of metal bonds [78]. Moreover, it is important to notice that in nature, many corrosion processes reflect directly the variety of microorganism involved in not in isolated processes, but in collaboration, with one serving on another in a particular environment [84].

The impacts caused by biofilms formation are estimated to be 1% of GNP (gross national product) of industrialized nations, where it is classified as an oversized process, biofilms and interruption of production expenses, and population health risk [75].

1.10. Microbial Control Methods Related to Oil Industry

The treatments usually used to control microbial growth, specially of the SRB group, and to mitigate damage processes associated to SRB presence in oil wells are mainly based on the inhibition of biologic activity (IBA) or on the specific prevention of anaerobic activity (PAA) [85]. In this last one, one strategy used that aims to prevent anaerobic activity through sulfate elimination from the aqueous phase is known as nanofiltration. This technic is based on water filtration using porous membranes (pore diameter of 1 mm, approximated). Comparatively, this membrane pores are slightly larger than reverse osmosis membranes one, and smaller than ultrafiltration membranes pores. When nanofiltration is applied, sulfate concentrations in sea water can be reduced from 2,700 ppm to less than 40 ppm. However, this technic is very expensive to be applied for large amounts of water as oil industry demands [86].

Considering IBA, the introduction of chemical biocides in injection water is commonly used in operational wells during waterflooding to prevent microbial growth. The biocides are divided as oxidative (chlorines and chloramines) or non-oxidative agents (amines, anthraquinones and aldehydes). The biocides that are more commonly used by petroleum industry are the anthraquinones, glutaraldehyde and the THPS (tetrakis hydroxymethyl phosphonium sulfate). The oxidative ones depend on factors such as light, pH and temperature to be effective. In the other hand, the non-oxidative are more stable in environmental conditions than the oxidative ones. Because of it, the non-oxidative can be used in a variety of environments [87].

The oil industry uses biocides to treat the petroleum recovery system, the pipelines, the tanks, the refrigerator tower, among others. The biocides effectiveness is the point here, since biocides treatments can cause bacterial resistance and it can also lead to environmental impacts. Because of this, the biocide type chosen for a specific treatment depends on the laws that dispose on environmental protection and on bacterial populations that are present in the field. Therefore, new strategies have been already considered to better control SRB presence, and their known consequences as biofilm formation, biocorrosion of equipment's and reservoirs acidulation [88, 89].

A variety of methods were proposed to inhibit SRB populations in the oil wells. Some of these methods, were based on microbial competition stimulation through favorable thermodynamically compounds introduction in the reservoir environment. These compounds, such as oxygen (O_2) and nitrate (NO_3), are electron acceptors more favorable than sulfate (SO_4) for heterotrophic metabolism in aerobic and anaerobic environments, respectively. Thus, when nitrate is introduced in an environment, the bacterial group known as nitrate reducing bacteria (NRB) are benefited. The SRB and the NRB competes directly for carbon organic compounds and, as sulfate reduction has a reduced energy gain when compared to nitrate reduction reaction, the NRB won this competition [90]. In summary, nitrate introduction is known as a competitive exclusion technology and it have been extensively studied [91–94].

Jurelevicius et al. (2008) observed nitrate effects in the SRB populations in a two-months experiment that consisted on the continuous introduction of nitrate in a water/oil storage tank.

Consequently, in the same study, the hydrogen sulfide production was controlled by nitrate injection. In another work, Marques et al. (2012) also observed nitrate injection effects in total bacterial community present in the produced water, in a three-months bioreactor experiment. In this study, SRB populations were reduced by nitrate injection but, in contrast, the metal corrosion enhanced with nitrate introduction during the experiment. As another disadvantage, Engelbrektson et al. [95] also highlights that nitrate needs to be injected in high concentrations in an intermittent way to directly inhibit SRB activity which commit logistical and economically the entire process.

As an alternative to nitrate introduction, Engelbrektson et al. [95] suggest the use of another electron acceptor, thermodynamically more favorable than sulfate (SO_4), the perchlorate. In this study, the authors observed that the perchlorate introduction led to a reduction in hydrogen sulfide production and the stimulation of sulfur oxidizing bacteria (SOB) populations. As some advantages, the authors point to the direct SRB-inhibition by perchlorate, the stability of this compound in the sulfide presence, and the absence of residual sulfide in the end of the treatment.

Another interesting alternative for SRB control, with less environmental impact than biocides, includes microbial-stimulating of antimicrobial substance producers (AMS) strains. This method could be used apart or as a complementation to the synthetic biocides. Some bacterial strains well-known as SAM producers are affiliated to Actinomycetes (Gram-positives filamentous bacteria) and spore-forming bacteria genera (*Bacillus*, among others). Korenblum et al. [96] observed that the strain H2O-1 affiliated to *Bacillus* genera, could produce SAM, classified as a surfactin-like, against a *Desulfovibrio alaskensis* strain. After that, da Rosa et al. [97] also demonstrated the inhibition of the same SRB by *Streptomyces lunalinharesii* 235 strain, another AMS producer. In another work, da Rosa et al. [97] demonstrated that *Streptomyces lunalinharesii* 235 strain could prevent the formation of SRB biofilm.

The natural products, or plant products, or bioproducts such as essential oils and the plants extracts can present antimicrobial activity against microorganisms. The essential oils (OE) are characterized as a mix of lipophilic volatile compounds produced by plants. The plants extracts are characterized as an herb/alcohol liquid solution, popularly known as popular medications. The natural products present low costs for industries, less environmental impacts, and they do not represent security and health risks for oil and gas employees. Thus, the bioproducts that presents antimicrobial activity against SRB represents another interesting alternative for the synthetic biocides [98].

Korenblum et al. [99] demonstrated that the OE of the lemon grass and its active compound, the cytral, showed antimicrobial activity against a *Desulfovibrio alaskensis* strain. In the same work, the OE and the cytral also presented an anticorrosive effect. Recently, de Souza et al. [100] demonstrated the antimicrobial activity of a variety of OE against a *Desulfovibrio alaskensis* strain and a bacterial consortium present in the produced water. The plant extracts SRB antagonistic activity was also demonstrated by Bhola et al. [101]. All these studies contribute to ensure the potential of the bioproducts for biotechnology application in the petroleum industry.

1.11. Bioremediation of Organic and Hydrocarbon Pollutants

One of the major oil spill is the Deepwater Horizon oil rig, in the Gulf of Mexico in 2010 [102–104]. At the time, the site became surrounded by oil, and the environmental impacts reached distances up to 2000 km of shoreline besides, the spill devastated many natural habitats. bioremediation was one of the key instruments to mitigate the environmental impacts and hazard in those depredated ecosystems [105, 106]. Bioemediation is a biobased, eco-friendly non-invasive strategy to remove ou degrade pollutants by using mostly microorganisms. Currently it's been considered it is also cheaper than conventional chemical methods, as chlorine usage and manipulation of highly oxidative agents [107].

One of the main strategies in the bioremeditation of alkanes is the aerobic pathway [108]. Alkanes, likewise plant paraffins and fatty acids have a ubiquitous profile in nature so that many microorganisms' metabolism use n-alkanes as carbon source. In this context, high Biological oxygen demand (BOD) is fundamental to the performance of two types of enzyme pathways: the monooxygenases, which take a straight forward strategy, seing that monooxygenase enzymes act by a incorportating oxygen atoms at terminal carbons of the linear alkanes chains, generating then primary alcohols. The second type of enzymes are the dioxygenases, are capable of inserting two oxygen atoms, forming now hydroperoxidades, instead of a primary alcohol. Insteringly, both enzyme pathways generate primary fatty acids as final products, opening up a diversified metabolic window (mostly go into the B-oxidation pathway) to microorganisms use it, e.g., cell wall composition, storage into lipid storage [108].

Bioremediation applications associated with hydrocarbons (e.g., petroleum, kerosene and crude oil) have showed robustness in terms of environmental diversity, as glaciers, mangroves, soil, and water [109, 110]. Crude oil removal by is one of the main activities in soil decontamination, with huge potential of scalability and development of novelties for *in situ* bioremediation [107]. Modern technical approaches have involved the development of mixed biofilms strains of *Pseudomonas monteilii* and *Gordonia* sp immobilized in polyurethane foam (PUF) as support for the biofilm and have been demonstrated a robust economic viability as a bioremediation tool [111, 112].

The biodegradation of petroleum hydrocarbons, however, faces major challenges usually concerning physiological requirements of microorganisms in *in situ* remediation [107, 113], where nitrogen and phosphate shortage have been reported as main contributors to biodegradation insucess in nutrient deploited environments [114]. In reference to that issue, permeable reactive barriers (PRB) have been used to minimize harzard and environmental impacts caused by diesel fuel spill in Casey Station, Antarctica. Techniques to mitigate the issue of nitrogen dependency, and cutting-edge approaches have involved the use of ammonium exchanged zeolite (sieves for ion exchange improvement) with PRB in the enhancement of biodegraation to support biofilm growth and sustainability of n-alkane degrading bacteria, as *Actinomycetales* and *Burkholderiales* [115].

1.12. Reservoirs Souring

The sulfur is one of the most important chemical elements from the earth and it is present in rocks as pyrite (FeS_2) and plaster (CaSO_4). The major reservoir of inorganic sulfur in our planet is in the ocean sediments where this element is found as sulfate ion (SO_4^{2-}). In these

sediments, the microorganisms play an important role to sulfur-cycling in anoxic areas and the sulfate reduction can be divided in assimilatory and dissimilatory reduction processes. In the first case, some microorganisms reduce sulfate in small amounts to incorporate sulfur in organic compounds to generate cysteine, methionine among others. In the last case, sulfate is reduced in large amounts, sulfide generated is cell-excreted, air oxidized and used by other organisms or complexed with metals forming metals sulfides. The dissimilatory sulfate reduction aims to generate energy for a bacterial group known as sulfide reducing bacteria (SRB) in their anaerobic respiration metabolism. In this process, a variety of electrons donors can be used for SRB, but the main ones are hydrogen, lactate and pyruvate [116, 117].

The SRB are directly involved in carbon-cycling in the anoxic sediments as they degrade carbon organic compounds using sulfate as the final electron acceptor producing hydrogen sulfide, in a reaction that requires eight electrons and a variety of intermediary stages. The dissimilatory process occurs in the cytoplasmic membrane, in the electron transport chain. In summary, the sulfate ion is activated in APS (adenosine 5'-phosphosulfate) by the enzyme ATP-sulfurylase. In the following, APS is reduced to sulfite by the enzyme APS reductase. Finally, the sulfite reductase enzyme catalyzes the sulfite reduction to sulfide. The sulfate is an electron acceptor less favorable than oxygen (O_2) and nitrate (NO_3) and, in response to the reduced energy gain of this reaction, it requires a low redox potential to SRB metabolism [117]. Besides, in sulfate absence, some SRB can use nitrate as the final acceptor or even some organic compounds as pyruvate [118].

The SRB diversity is related to their habitats which demonstrate the biological complexity of this bacterial group. In numbers, more than 220 bacterial species were affiliated to 59 SRB genera from Bacteria and Archaea domains [119]. The SRB genera described was *Desulfovibrio* and *Desulfotomaculum*, being this first one the most studied [117, 120]. The SRB are widespread, occurring in natural and artificial environments where sulfate is present. The SRB have already been isolated from marine and fresh water sediments, hydrothermal vents, volcano mud, microbial aggregates formed in hypersaline conditions, plants rhizosphere, aquifers among others. Furthermore, the SRB were also isolated from oil fields [116].

The microbial development in oil fields is limited by electron acceptors concentrations, such as oxygen, nitrate, ferrous metal or sulfate. The presence of sulfate ions inside the oil well creates an appropriate environment for SRB activity, which leads to hydrogen sulfide production, in a process known as reservoir souring [121].

The souring process is usually observed in offshore platforms where the sea water is used in the waterflooding process to enhance reservoir pressure and displace the oil from reservoirs rocks during the secondary recovery. The sea water contents high concentrations of salts, specially chlorates, sulfates and significant densities of viable SRB populations [122]. The sea water is collected from places nearby the platforms and submitted to several treatments, before it can be injected in the well. First, the water is treated to remove the oxygen to avoid pipeline corrosion process. After that, the water is filtered and a variety of chemicals are introduced in this fluid as antifoam agents, crusting inhibitors and sulfide and bisulfide ions, these last ones added to water to interact with dissolved oxygen forming sulfate [123, 124].

The sea water after treatment is known as injection water. The injection water is injected in the well and mixes with formation water, that consists on a residual fluid generated during the quarry development process. The result of the injection and the formation water combination is the production water [125]. This last one is collected with the oil in the production well in the end of the waterflooding. The production water is characterized as a

water/oil emulsion, which is rich in electron donors, carbon sources (acetate, propionate among others), nitrogen sources (such as ammonia) with higher salinity than sea water (chlorates and bicarbonates high concentrations). The produced water is re-injected in the well several times during the waterflooding process before it can be discarded [126, 127].

The produced water contains high concentrations of sulfate ion and an even higher density of viable SRB populations than that one present in sea water. The re-injection process with produced water contributes to reservoir souring since this fluid brings the preferential electron acceptor in SRB anaerobic respiration (sulfate), the SRB population in high densities and organic compounds (hydrocarbons present in water and/or in the well) used in SRB heterotrophic metabolism. Moreover, the waterflooding cools the reservoir reaching lower temperatures and providing the perfect environment for SRB development as they are mesophilic bacteria. Consequently, the SRB present in produced water can generate hydrogen sulfide in extremely higher concentrations [119, 128, 129].

The hydrogen sulfide is responsible for several economic losses for the petroleum industry. The hydrogen sulfide leads to oil, gas and produced water contamination, reducing the petroleum quality. To promote produced water re-injection, the iron sulfide generated from SRB metabolism needs to be removed from the aqueous phase and it implicates in more financial costs. The waterflood oil yield is also affected by the iron sulfide as it can lead to reservoir-plugging, reducing the rock permeability. The sulfide gas is toxic, inflammable and represents a potential risk for employee's security and health as well. Consequently, an oil well acidulation involves additional costs to control employees' sulfide-exposition. Besides, extra costs concerning souring-control and prevention also count. Such mechanisms involve: i. Chemical treatments to remove sulfide; ii. Transport and storage requirement to treat contaminated materials; iii. Biocides introduction; among others. All these mechanisms cause major logistical implications for the oil business [89, 130].

According to Davidova et al. [131], the oil reservoirs are not the only and nor even the more affected ones by the sulfide production in the petroleum fields. The sulfide production and SRB activity also causes clear damages as the corrosion of equipment's surface as water/oil storage tanks or ponds, pipelines among others. The corrosion process leads to a variety of economic problems for the petroleum industries. The National Association of Corrosion Engineers (NACE - guideline MR0175/ISO-15156) establish that sulfide gas partial pressure cannot exceed the oil facilities resistance level. In case of it, the oil industry is responsible to take some security measures adopting sulfide corrosion protection services. These services can result in millions of dollars of additional costs for each well, and therefore for the entire oil drilling, extraction and recovery [124, 89, 130].

CONCLUSION

Petroleum exploration is one of the most important activities in the world and has several issues to be comprehended. Microbes have an essential role in many steps, since formation, until in eventual process of environmental cleaning. Understanding microbial genetics and how they interact with oil and environment is crucial to improve exploration and for the so-called one health care.

REFERENCES

- [1] Van Hamme JD, Singh A, Ward OP. Recent Advances in Petroleum Microbiology. *Microbiol Mol Biol Rev* 2003;67:503–49. doi:10.1128/MMBR.67.4.503.
- [2] Saraceno A da SP. Estudo Do Fenômeno De Parafinação A Partir De Um Óleo Cru. [Study of the Phenomenon of Paraffinization from a Raw Oil]. *COPPE* 2007;1:79.
- [3] Glasby GP. Abiogenic origin of hydrocarbons: An historical overview. *Resour Geol* 2006;56:83–96. doi:10.1111/j.1751-3928.2006.tb00271.x.
- [4] Speers G, Whithehead E. *Organic Geochemistry*. Berlin, Heidelberg: Springer Berlin Heidelberg; 1969. doi:10.1007/978-3-642-87734-6.
- [5] England WA, Mackenzie AS, Mann DM, Quigley TM. The movement and entrapment of petroleum fluids in the subsurface. *J Geol Soc London* 1987;144:327–47. doi:10.1144/gsjgs.144.2.0327.
- [6] Scott HP, Hemley RJ, Mao HK, Herschbach DR, Fried LE, Howard WM, et al. Generation of methane in the Earth's mantle: In situ high pressure-temperature measurements of carbonate reduction. *Proc Natl Acad Sci U S A* 2004;101:14023–6. doi:10.1073/pnas.0405930101.
- [7] Leinz V, Amaral SE. *Geologia Geral*. 1969.
- [8] Tissot BP, Welte DH. *Petroleum Formation and Occurrence*. Berlin, Heidelberg: Springer Berlin Heidelberg; 1978. doi:10.1007/978-3-642-96446-6.
- [9] Libes SM. *Introduction to marine biogeochemistry*. vol. 42. 2009. doi:10.1016/0304-4203(93)90016-H.
- [10] Posthuma J. The composition of petroleum. *Rapp Procès-Verbaux Des Réunions, Cons ...* 1977;171:7–16.
- [11] Ludwig FJ. Analysis of Microcrystalline and Paraffin Waxes by Means of Infrared Spectra in the Molten State. *Anal Chem* 1965;37:1737–41. doi:10.1021/ac60232a026.
- [12] Volkman JK, Holdsworth DG, Neill GP, Bavor Jr. HJ. Identification of natural, anthropogenic and petroleum hydrocarbons in aquatic sediments. *Sci Total Environ* 1992;112:203–19. doi:http://dx.doi.org/10.1016/0048-9697(92)90188-X.
- [13] Council NR, Sciences DEP, Commission on Physical Sciences MA, Update SCPME. *Oil in the Sea: Inputs, Fates, and Effects*. National Academies Press; 1985.
- [14] Burcik EJ. Properties of petroleum reservoir fluids. *J Environ Public Health* 1979;2014:656824. doi:10.1155/2014/656824.
- [15] Callaghan AV, Davidova IA, Savage-Ashlock K, Parisi VA, Gieg LM, Suflita JM, et al. Diversity of BenzyL- and Alkylsuccinate Synthase Genes in Hydrocarbon- Impacted Environments and Enrichment Cultures. *Environ Sci Technol* 2010;7287–94.
- [16] Thrasher DR, Vance I. Reservoir Souring: Mechanisms and Prevention. *Pet Microbiol* 2016;123–42. doi:10.1128/9781555817589.ch7.
- [17] Borges SMS. *Recuperação Avançada de Petróleo (EOR) com a Utilização da Glicerina Bruta (GB) co-produto da produção de Biodiesel* 2009:92. [Advanced Oil Recovery

(EOR) with the Use of Crude Glycerin (GB) as a co-product of Biodiesel production 2009:92].

- [18] Thomas JE, Triggia AA, Correia CA, Filho CV, Xavier JAD, Machando JCV, et al. *Fundamentos de Engenharia do Petróleo*. 2001. doi:f977.
- [19] Cardoso LC. *Petróleo: Do Poço ao Posto*. [From the Well to the Post]. 2nd ed. São Paulo: *Qualitymark*; 2008.
- [20] Rosa A, Carvalho R, Xavier J. *Engenharia de reservatórios de petróleo*. [Engineering of oil reservoirs]. Rio de Janeiro: 2006.
- [21] Beerstecher E. *Petroleum microbiology: an introduction to microbiological petroleum engineering*. Elsevier Press; 1954.
- [22] Corrêa OLS. *Petróleo: noções sobre exploração, perfuração, produção e microbiologia*. [Oil: notions on exploration, drilling, production and microbiology]. *Interciência*; 2003.
- [23] Lakatos I, Szabo JL. Global oil demand and role of chemical EOR methods in the 21st century. *Int J Oil, Gas Coal Technol* 2008;1:46. doi:10.1504/IJOGCT.2008.016731.
- [24] Carrero E, Queipo NV., Pintos S, Zepa LE. Global sensitivity analysis of Alkali-Surfactant-Polymer enhanced oil recovery processes. *J Pet Sci Eng* 2007;58:30–42. doi:10.1016/j.petrol.2006.11.007.
- [25] Mulligan CN, Sharma SK, Mudhoo A. *Biosurfactants: Research Trends and Applications*. Taylor & Francis; 2014.
- [26] Santos DKF, Rufino RD, Luna JM, Santos VA, Sarubbo LA. Biosurfactants: Multifunctional biomolecules of the 21st century. *Int J Mol Sci* 2016;17:1–31. doi:10.3390/ijms17030401.
- [27] Uller VC, Szklo AS. *Fundamentos Do Refino De Petróleo: Tecnologia E Economia*. [Fundamentals Of Oil Refining: Technology And Economics]. *Interciencia*; n.d.
- [28] Mariano JB. *Impactos ambientais do refino de petróleo*. [Environmental impacts of oil refining]. *Interciência*; 2005.
- [29] Leffler WL. *Petroleum Refining for the Non-Technical Person*. PennWell Corporation; 1985.
- [30] Atlas RM, Philp J, Washington AP. *Bioremediation Applied Microbial Solutions for Real-World Environmental Cleanup* Library of Congress Cataloging-in-Publication Data. vol. 5. 2005. doi:10.2136/sssaj2005.0002BR.
- [31] Pohlman JW, Bauer JE, Waite WF, Osburn CL, Chapman NR. Methane hydrate-bearing seeps as a source of aged dissolved organic carbon to the oceans. *Nat Geosci* 2011;4:37–41. doi:10.1038/ngeo1016.
- [32] Santelices B, Cancino J, Montalva B, Pinto R, Gonzales E. Estudios ecologicos en la zona costera afectada por contaminacion del “Northern Breeze” - II - Comunidades de playas de rocas. [Ecological studies in the coastal area affected by pollution of the "Northern Breeze" - II - Communities of rock beaches]. *Medio Ambient* 1977;2:65–83.
- [33] IMO/FAO/UNESCO/WMO/WHO/IAEA/UN/UNEP Joint Group of Experts on the Scientific Aspects of Marine Pollution (GESAMP). *Impact of oil and related chemicals and wastes on the marine environment*. London: 1993.

- [34] Cole GM. *Assessment and Remediation of Petroleum Contaminated Sites*. Taylor & Francis; 1994.
- [35] Donnelly JR, Betowski LD. Polycyclic aromatic hydrocarbons determination. In: Meyers R., Dittrick DK, editors. *Wiley Encycl. Environ. Pollut. Cleanup*, New York: John Wiley & Sons; 1999, p. 3831–7.
- [36] Margesin R, Schinner F. Biodegradation and bioremediation of hydrocarbons in extreme environments. *Appl Microbiol Biotechnol* 2001;56:650–63. doi:10.1007/s002530100701.
- [37] Head IM, Jones DM, Larter SR. Biological activity in the deep subsurface and the origin of heavy oil. *Nature* 2003;426:344–52. doi:10.1038/nature02134.
- [38] Varjani SJ, Gnansounou E. Microbial dynamics in petroleum oilfields and their relationship with physiological properties of petroleum oil reservoirs. *Bioresour Technol* 2017. doi:10.1016/j.biortech.2017.08.028.
- [39] Bastin ES. The Problem of the Natural Reduction of Sulphates. *Bull Am Assoc Pet Geol* 1926;10:1270–1299. doi:10.1306/3D932761-16B1-11D7-8645000102C1865D.
- [40] Magot M, Ollivier B, Patel BKC. Microbiology of petroleum reservoirs. Antonie van Leeuwenhoek, *Int J Gen Mol Microbiol* 2000;77:103–16. doi:10.1023/A:1002434330514.
- [41] Sugai Y, Purwasena IA, Sasaki K, Fujiwara K, Hattori Y, Okatsu K. Experimental studies on indigenous hydrocarbon-degrading and hydrogen-producing bacteria in an oilfield for microbial restoration of natural gas deposits with CO₂ sequestration. *J Nat Gas Sci Eng* 2012;5:31–41. doi:10.1016/j.jngse.2012.01.011.
- [42] Head IM, Gray ND, Larter SR. Life in the slow lane; biogeochemistry of biodegraded petroleum containing reservoirs and implications for energy recovery and carbon management. *Front Microbiol* 2014;5:1–23. doi:10.3389/fmicb.2014.00566.
- [43] Nie Y, Zhao JY, Tang YQ, Guo P, Yang Y, Wu XL, et al. Species divergence vs. functional convergence characterizes crude oil microbial community assembly. *Front Microbiol* 2016;7:1–11. doi:10.3389/fmicb.2016.01254.
- [44] Head IM, Jones DM, Röling WFM. Marine microorganisms make a meal of oil. *Nat Rev Microbiol* 2006;4:173–82. doi:10.1038/nrmicro1348.
- [45] Hazen TC, Dubinsky EA, DeSantis TZ, Andersen GL, Piceno YM, Singh N., et al. Deep-Sea Oil Plume Enriches Indigenous Oil-Degrading Bacteria. *Science* (80-) 2010;330:204–8. doi:10.1126/science.1195979.
- [46] Garcia JL, Patel BKC, Ollivier B. Taxonomic phylogenetic and ecological diversity of methanogenic Archaea. *Anaerobe* 2000;6:205–26.
- [47] Nazina TN, Ivanova AE, Golubeva OV, Ibatullin RR, Belyaev SS, Ivanov MV. Occurrence of Sulfate-Reducing and Iron-Reducing Bacteria in Stratal Waters of the Romashkinskoe Oil-Field. *Microbiology* 1995;64:203–8.
- [48] Das N, Chandran P, Das N, Chandran P. Microbial Degradation of Petroleum Hydrocarbon Contaminants: An Overview. *Biotechnol Res Int* 2011;2011:1–13. doi:10.4061/2011/941810.

- [49] Rojo F. Degradation of alkanes by bacteria: Minireview. *Environ Microbiol* 2009;11:2477–90. doi:10.1111/j.1462-2920.2009.01948.x.
- [50] Cerniglia CE. Biodegradation of polycyclic aromatic hydrocarbons. *Curr Opin Biotechnol* 1993;4:331–8. doi:10.1007/BF00129093.
- [51] Wilson SC, Jones KC. Bioremediation of soil contaminated with polynuclear aromatic hydrocarbons (PAHs): a review. *Environ Pollut* 1993;81:229–49. doi:10.1016/0269-7491(93)90206-4.
- [52] Peixoto RS, Vermelho a. B, Rosado a. S. Petroleum-Degrading Enzymes: Bioremediation and New Prospects. *Enzyme Res* 2011;2011:1–7. doi:10.4061/2011/475193.
- [53] Maletić S, Dalmacija B, Rončević S. Petroleum hydrocarbon biodegradability in soil. *Implic Bioremediation* 2013;22. doi:10.5772/50108.
- [54] Yakimov MM, Timmis KN, Golyshin PN. Obligate oil-degrading marine bacteria. *Curr Opin Biotechnol* 2007;18:257–66. doi:10.1016/j.copbio.2007.04.006.
- [55] Madigan M. Brock *Biology of Microorganisms*, 14th edn. 2012. doi:10.1016/B978-1-4832-3136-5.50010-3.
- [56] Das N, Chandran P. Microbial degradation of petroleum hydrocarbon contaminants: an overview. *Biotechnol Res Int* 2011;2011:941810. doi:10.4061/2011/941810.
- [57] Leahy JG, Colwell RR. Microbial degradation of hydrocarbons in the environment. *Microbiol Rev* 1990;54:305–15.
- [58] Cerniglia CE. Biodegradation of polycyclic aromatic hydrocarbons. *Biodegradation* 1992;3:351–68. doi:10.1007/BF00129093.
- [59] Smits THM, Balada SB, Witholt B, Van Beilen JB. Functional analysis of alkane hydroxylases from gram-negative and gram-positive bacteria. *J Bacteriol* 2002;184:1733–42. doi:10.1128/JB.184.6.1733-1742.2002.
- [60] Van Beilen JB, Li Z, Duetz WA, Smits THM, Witholt B. Diversity of Alkane Hydroxylase Systems in the Environment. *Oil Gas Sci Technol* 2003;58:427–40. doi:10.2516/ogst:2003026.
- [61] van Beilen JB, Panke S, Lucchini S, Franchini AG, Röthlisberger M, Witholt B. Analysis of *Pseudomonas putida* alkane-degradation gene clusters and flanking insertion sequences: Evolution and regulation of the alk genes. *Microbiology* 2001;147:1621–30. doi:10.1099/00221287-147-6-1621.
- [62] Chesnokova MG, Shalaj VV, Kraus YA, Cherkashina NV, Mironov AY. Analysis of Corrosion Defects on Oil Pipeline Surface Using Scanning Electron Microscopy and Soil Thionic and Sulfate-reducing Bacteria Quantification. *Procedia Eng* 2016;152:247–50. doi:10.1016/j.proeng.2016.07.698.
- [63] Li XX, Liu JF, Yao F, Wu WL, Yang SZ, Mbadinga SM, et al. Dominance of *Desulfotignum* in sulfate-reducing community in high sulfate production-water of high temperature and corrosive petroleum reservoirs. *Int Biodeterior Biodegradation* 2016;114:45–56. doi:10.1016/j.ibiod.2016.05.018.
- [64] Lillicrap A, Macken A, Wennberg AC, Grung M, Rundberget JT, Fredriksen L, et al. Environmental fate and effects of novel quorum sensing inhibitors that can control

- biofilm formation. *Chemosphere* 2016;164:52–8. doi:10.1016/j.chemosphere.2016.08.083.
- [65] Guan J, Xia LP, Wang LY, Liu JF, Gu JD, Mu BZ. Diversity and distribution of sulfate-reducing bacteria in four petroleum reservoirs detected by using 16S rRNA and *dsrAB* genes. *Int Biodeterior Biodegradation* 2013;76:58–66. doi:10.1016/j.ibiod.2012.06.021.
- [66] Handelsman J, Rondon MR, Brady SF, Clardy J, Goodman RM. Molecular biological access to the chemistry of unknown soil microbes: a new frontier for natural products. *Chem Biol* 1998;5:R245–9. doi:10.1016/S1074-5521(98)90108-9.
- [67] Handelsman J. Metagenomics: application of genomics to uncultured microorganisms. *Microbiol Mol Biol Rev* 2004;68:669–85.
- [68] Braga M de B, Veras AA de O, Gomes de Sá PHC, Graças DA das, Baraúna RA, Alves JTC, et al. Omics technologies applied to prokaryotes. In: Urbano K V, editor. *Adv. Genet. Rousources*. 16th ed., New York, NY: Nova Science Publisher; 2016, p. 201–25.
- [69] Falkowski PG, Fenchel T, Delong EF. The microbial engines that drive Earth's biogeochemical cycles. *Science* 2008;320:1034–9. doi:10.1126/science.1153213.
- [70] Fuhrman JA, Cram JA, Needham DM. Marine microbial community dynamics and their ecological interpretation. *Nat Rev Microbiol* 2015;13:133–46. doi:10.1038/nrmicro3417.
- [71] Venter JC, Remington K, Heidelberg JF, Halpern AL, Rusch D, Eisen J a, et al. Environmental genome shotgun sequencing of the Sargasso Sea. *Science* 2004;304:66–74. doi:10.1126/science.1093857.
- [72] Stepanauskas R. Single cell genomics: An individual look at microbes. *Curr Opin Microbiol* 2012;15:613–20. doi:10.1016/j.mib.2012.09.001.
- [73] Rinke C, Schwientek P, Sczyrba A, Ivanova NN, Anderson IJ, Cheng JF, et al. Insights into the phylogeny and coding potential of microbial dark matter. *Nature* 2013;499:431–7. doi:10.1038/nature12352.
- [74] Zhang C, Kim SK. Research and Application of Marine Microbial Enzymes: Status and Prospects. *Mar Drugs* 2010;8:1920–34. doi:10.3390/md8061920.
- [75] Schneider RP. *Microbiologia Ambiental*. [Environmental Microbiology]. 2nd ed. Jaguariúna, SP: *Embrapa Meio Ambiente*; 2008.
- [76] Beech IB, Gaylarde CC. Recent advances in the study of biocorrosion: an overview. *Microbiol Rev* 1999;30:177–90. doi:10.1590/S0001-37141999000300001.
- [77] Gentil V. *Corrosão induzida por microrganismos*. [Corrosion induced by microorganisms]. 4th ed. Rio de Janeiro: LTC; 2005.
- [78] Cavalcanti GH. Efeitos de biocidas sobre biofilmes bacterianos envolvidos em processos de biocorrosão em plataformas da petrobras. [Effects of biocides on bacterial biofilms involved in biocorrosion processes in petrobras platforms]. *Fiocruz*, 2001.
- [79] Aguiar LEV. Estudos eletroquímicos para desenvolvimento de um método de identificação e semiquantificação de bactérias redutoras de sulfato envolvidas em processo de corrosão. [Electrochemical studies for the development of a method of identification and semi-quantification of sulfate reducing bacteria involved in the corrosion process]. *Fiocruz*, 1991.

- [80] Lee W, Lewandowski Z, Nielsen PH, Hamilton WA. Biofouling: *The Journal of Bioadhesion and Biofilm Research* Role of sulfate - reducing bacteria in corrosion of mild steel: A review 2009;37–41. doi:10.1080/08927019509378271.
- [81] Hamilton W a. Bioenergetics of sulphate-reducing bacteria in relation to their environmental impact. *Biodegradation* 1998;9:201–12. doi:10.1023/a:1008362304234.
- [82] Lee W, Lewandowski Z, Nielsen PH, Hamilton WA. Role of sulfate-reducing bacteria in corrosion of mild steel: A review. *Biofouling* 1995;8:165–94. doi:10.1080/08927019509378271.
- [83] Hamilton WA. Sulphate-Reducing Bacteria and Anaerobic Corrosion. *Annu Rev Microbiol* 1985;39:195–217. doi:10.1146/annurev.mi.39.100185.001211.
- [84] Beech IB, Sunny Cheung CW, Patrick Chan CS, Hill MA, Franco R, Lino AR. Study of parameters implicated in the biodeterioration of mild steel in the presence of different species of sulphate-reducing bacteria. *Int Biodeterior Biodegrad* 1994;34:289–303. doi:10.1016/0964-8305(94)90089-2.
- [85] García De Lomas J, Corzo A, Gonzalez JM, Andrades JA, Iglesias E, Montero MJ. Nitrate promotes biological oxidation of sulfide in wastewaters: Experiment at plant-scale. *Biotechnol Bioeng* 2006;93:801–11. doi:10.1002/bit.20768.
- [86] Wei C, He W, Wei L, Li C, Ma J. The analysis of a microbial community in the UV/O₃-Anaerobic/Aerobic Integrated Process for Petrochemical Nanofiltration Concentrate (NFC) Treatment by 454-Pyrosequencing. *PLoS One* 2015;10:1–14. doi:10.1371/journal.pone.0139991.
- [87] Videla HA, Herrera LK. Microbiologically influenced corrosion: looking to the future. *Int Microbiol* 2005;8:169–80. doi:im2305026 [pii].
- [88] Albuquerque; AC, Neves CAB. Biocorrosão – Da Integridade Do Biofilme À Integridade Do Material. [Biocorrosion - From Biofilm Integrity to Material Integrity]. *Corrosão Proteção Mater* 2014;33:18–23.
- [1] Hubert C, Nemati M, Jenneman G, Voordouw G. Corrosion risk associated with microbial souring control using nitrate or nitrite. *Appl Microbiol Biotechnol* 2005;68:272–82. doi:10.1007/s00253-005-1897-2.
- [2] Larsen J, Rod MH, Zwolle S. Prevention of reservoir souring in the Halfdan field by nitrate injection. *Corros* 2004 2004:18.
- [3] Jurelevicius D, von der Weid I, Korenblum E, Valoni E, Penna M, Seldin L. Effect of nitrate injection on the bacterial community in a water-oil tank system analyzed by PCR-DGGE. *J Ind Microbiol Biotechnol* 2008;35:251–5. doi:10.1007/s10295-007-0297-2.
- [4] Gittel A, Sorensen KB, Skovhus TL, Ingvorsen K, Schramm A. Prokaryotic community structure and sulfate reducer activity in water from high-temperature oil reservoirs with and without nitrate treatment. *Appl Environ Microbiol* 2009;75:7086–96. doi:10.1128/AEM.01123-09.
- [5] Kumaraswamy R, Ebert S, Gray MR, Fedorak PM, Foght JM. Molecular- and cultivation-based analyses of microbial communities in oil field water and in microcosms amended with nitrate to control H₂S production. *Appl Microbiol Biotechnol* 2011;89:2027–38. doi:10.1007/s00253-010-2974-8.

- [6] Marques JM, de Almeida FP, Lins U, Seldin L, Korenblum E. Nitrate treatment effects on bacterial community biofilm formed on carbon steel in produced water stirred tank bioreactor. *World J Microbiol Biotechnol* 2012;28:2355–63. doi:10.1007/s11274-012-1043-0.
- [7] Engelbrektson A, Hubbard CG, Tom LM, Boussina A, Jin YT, Wong H, et al. Inhibition of microbial sulfate reduction in a flow-through column system by (per)chlorate treatment. *Front Microbiol* 2014;5:1–11. doi:10.3389/fmicb.2014.00315.
- [8] Korenblum E, Regina de Vasconcelos Goulart F, de Almeida Rodrigues I, Abreu F, Lins U, Alves P, et al. Antimicrobial action and anti-corrosion effect against sulfate reducing bacteria by lemongrass (*Cymbopogon citratus*) essential oil and its major component, the citral. *AMB Express* 2013;3:44. doi:10.1186/2191-0855-3-44.
- [9] Rosa JP da, Tibúrcio SRG, Marques JM, Seldin L, Coelho RRR. *Streptomyces lunalinharesii* 235 prevents the formation of a sulfate-reducing bacterial biofilm. *Brazilian J Microbiol* 2016;47:603–9. doi:10.1016/j.bjm.2016.04.013.
- [10] Street CN, Gibbs A. Eradication of the corrosion-causing bacterial strains *Desulfovibrio vulgaris* and *Desulfovibrio desulfuricans* in planktonic and biofilm form using photodisinfection. *Corros Sci* 2010;52:1447–52. doi:10.1016/j.corsci.2009.12.022.
- [11] Korenblum E, de Araujo L, Guimarães C, de Souza LM, Sasaki G, Abreu F, et al. Purification and characterization of a surfactin-like molecule produced by *Bacillus* sp. H2O-1 and its antagonistic effect against sulfate reducing bacteria. *BMC Microbiol* 2012;12:252. doi:10.1186/1471-2180-12-252.
- [12] De Souza PM, Goulart FRDV, Marques JM, Bizzo HR, Blank AF, Groposo C, et al. Growth inhibition of sulfate-reducing bacteria in produced water from the petroleum industry using essential oils. *Molecules* 2017;22. doi:10.3390/molecules22040648.
- [13] Bhola SM, Alabbas FM, Bhola R, Spear JR, Mishra B, Olson DL, et al. Neem extract as an inhibitor for biocorrosion influenced by sulfate reducing bacteria: A preliminary investigation. *Eng Fail Anal* 2014;36:92–103. doi:10.1016/j.engfailanal.2013.09.015.
- [14] Beyer J, Trannum HC, Bakke T, Hodson P V., Collier TK. Environmental effects of the Deepwater Horizon oil spill: A review. *Mar Pollut Bull* 2016;110:28–51. doi:10.1016/j.marpolbul.2016.06.027.
- [15] Follett L, Genschel U, Hofmann H. A graphical exploration of the Deepwater Horizon oil spill. *Comput Stat* 2014;29:121–32. doi:10.1007/s00180-013-0432-7.
- [16] Etnoyer PJ, Wickes LN, Silva M, Dubick JD, Balthis L, Salgado E, et al. Decline in condition of gorgonian octocorals on mesophotic reefs in the northern Gulf of Mexico: before and after the Deepwater Horizon oil spill. *Coral Reefs* 2016;35:77–90. doi:10.1007/s00338-015-1363-2.
- [17] Silva M, Etnoyer PJ, MacDonald IR. Coral injuries observed at Mesophotic Reefs after the Deepwater Horizon oil discharge. *Deep Res Part II Top Stud Oceanogr* 2016;129:96–107. doi:10.1016/j.dsr2.2015.05.013.

- [18] Prouty NG, Fisher CR, Demopoulos AWJ, Druffel ERM. Growth rates and ages of deep-sea corals impacted by the Deepwater Horizon oil spill. *Deep Res Part II Top Stud Oceanogr* 2016;129:196–212. doi:10.1016/j.dsr2.2014.10.021.
- [19] Dzionek A, Wojcieszyska D, Guzik U. Natural carriers in bioremediation: A review. *Electron J Biotechnol* 2016;23:28–36. doi:10.1016/j.ejbt.2016.07.003.
- [20] Sanchez-Salas JL, Raynal Gutierrez ME, Bandala ER. Aerobic Treatment of Petroleum Industry Effluents. *Curr. Dev. Biotechnol. Bioeng. Biol. Treat. Ind. Effluents*, 2016, p. 79–102. doi:10.1016/B978-0-444-63665-2.00003-5.
- [21] MR, Mehraşbi; B, Haghighi; M, Shariat; S, Naseri; KN. Biodegradation of Petroleum Hydrocarbons in Soil. *Iran J Publ Heal* 2003;32:28–32.
- [22] Rodgers-Vieira EA, Zhang Z, Adrion AC, Gold A, Aitken MD. Identification of anthraquinone-degrading bacteria in soil contaminated with polycyclic aromatic hydrocarbons. *Appl Environ Microbiol* 2015;81:3775–81. doi:10.1128/AEM.00033-15.
- [23] Alessandrello MJ, Juárez Tomás MS, Raimondo EE, Vullo DL, Ferrero MA. Petroleum oil removal by immobilized bacterial cells on polyurethane foam under different temperature conditions. *Mar Pollut Bull* 2017;122:156–60. doi:10.1016/j.marpolbul.2017.06.040.
- [24] Hodges A, Fica Z, Wanlass J, VanDarlin J, Sims R. Nutrient and suspended solids removal from petrochemical wastewater via microalgal biofilm cultivation. *Chemosphere* 2017;174:46–8. doi:10.1016/j.chemosphere.2017.01.107.
- [25] Kegler P, Baum G, Indriana LF, Wild C, Kunzmann A. Physiological response of the hard coral pocillopora verrucosa from Lombok, Indonesia, to two common pollutants in combination with high temperature. *PLoS One* 2015;10:e0142744. doi:10.1371/journal.pone.0142744.
- [26] Freidman BL, Speirs LBM, Churchill J, Gras SL, Tucci J, Snape I, et al. Biofilm communities and biodegradation within permeable reactive barriers at fuel spill sites in Antarctica. *Int Biodeterior Biodegradation* 2017;125:45–53. doi:10.1016/j.ibiod.2017.08.008.
- [27] Freidman BL, Gras SL, Snape I, Stevens GW, Mumford KA. The performance of ammonium exchanged zeolite for the biodegradation of petroleum hydrocarbons migrating in soil water. *J Hazard Mater* 2016;313:272–82. doi:10.1016/j.jhazmat.2016.03.095.
- [28] Muyzer G, Stams AJM. The ecology and biotechnology of sulphate-reducing bacteria. *Nat Rev - Microbiol* 2008;6:441–54. doi:10.1038/nrmicro1892.
- [29] Madigan, MT, Martinko, JM, Bender, KS, Buckley, DH. & Stahl DA. *Brock Microbiology*. 14th ed. Pearson/Benjamin Cummings; 2016.
- [30] Hao T wei, Xiang P yu, Mackey HR, Chi K, Lu H, Chui H kwong, et al. A review of biological sulfate conversions in wastewater treatment. *Water Res* 2014;65:1–21. doi:10.1016/j.watres.2014.06.043.
- [31] Barton LL, Fardeau ML, Fauque GD. Hydrogen sulfide: a toxic gas produced by dissimilatory sulfate and sulfur reduction and consumed by microbial oxidation. *Met Ions Life Sci* 2014;14:237–77. doi:10.1007/978-94-017-9269-1_10.

- [32] Postgate JR. *The sulphate-reducing bacteria*. 2nd ed. Cambridge: Cambridge University Press; 1984.
- [33] Hubert C, Voordouw G, Nemati M, Jenneman GE. Is souring and corrosion by sulfate-reducing bacteria in oil fields reduced more efficiently by nitrate or nitrite? *NACE Int Corros Conf Expo* 2004;4762.
- [34] Eckford RE, Fedorak PM. Chemical and microbiological changes in laboratory incubations of nitrate amendment “sour” produced waters from three western Canadian oil fields. *J Ind Microbiol Biotechnol* 2002;29:243–54. doi:10.1038/sj/jim/7000304.
- [35] Mcinerney, MJ, Sublette KL. Oil field microbiology. In: Hurst, CJ, Crawford, RL, Knudsen, GR, Mcinerney, MJ. & Stetzenbach LD, editor. *Man. Environ. Microbiol.*, Washington, D.C.: ASM Press; 2002.
- [36] Vance, I. & Thrasher DR. Reservoir Souring: Mechanisms and prevention. In: Ollivier, B. & Magot M, editor. *Pet. Microbiol*. 1st ed., Washington: ASM Press; 2005, p. 123–42.
- [37] Thomas JE. Fundamentos de engenharia de petróleo. [Fundamentals of petroleum engineering]. 2nd ed. Rio de Janeiro: *Interciência*; 2004.
- [38] Planckaert M. Oil reservoirs and oil production. *Pet Microbiol* 2005;3–19. doi:10.1128/9781555817589.ch1.
- [39] Stewart, M, Arnold K. *Produced Water Treatment System. Prod. water Treat. F. Man*. 1st ed., Gulf Professional Publishing; 2011, p. 1–134.
- [40] Sunde, E. & Torsvik T. Microbial control of hydrogen sulfide production in oil reservoirs. In: Ollivier, B. & Magot M, editor. *Pet. Microbiol*. 1st ed., Washington, D.C.: ASM Press; 2005, p. 201–14.
- [41] Enning D, Garrelfs J. Corrosion of iron by sulfate-reducing bacteria: New views of an old problem. *Appl Environ Microbiol* 2014;80:1226–36. doi:10.1128/AEM.02848-13.
- [42] Ollivier, B. & Cayol JL. Fermentative, iron-reducing, and nitrate-reducing microorganisms. In: Ollivier, B. & Magot M, editor. *Pet. Microbiol*. 1st ed., Washington, D.C.: ASM Press; 2005, p. 71–88.
- [43] Davidova I, Hicks M, Fedorak P, Suflita J. The influence of nitrate on microbial processes in oil industry production waters. *J Ind Microbiol Biotechnol* 2001;27:80–6. doi:10.1038/sj.jim.7000166.

Chapter 74

ESTROGEN ACTIVATED ERs ARE THE CHIEF GENETIC SAFEGUARDS OF SOMATIC AND REPRODUCTIVE HEALTH IN MAMMALIANS

Zsuzsanna Suba*

National Institute of Oncology,
Surgical and Molecular Tumor Pathology Center,
Budapest, Hungary

ABSTRACT

In healthy individuals, the incessant activity of regulatory genetic mechanisms ensures the metabolic and proliferative equilibrium of cellular activity. In case of accidental defects occurring in any part of the system, a coordinated counteraction of numerous mediators may successfully help in the restoration of physiologic processes by means of either overexpression or hyperactivity. Even serious defects of genome stabilizer mechanisms may be kept in balance for a long duration, showing the clinical signs of good health. By contrast, due to the exhaustion of the compensatory processes, DNA defects may develop and lead to the clinical manifestations of diseases. Estrogen activated estrogen receptors (ERs) are the primary initiators and organizers of the up-regulatory circle of genome stabilization in correlation and crosstalk with aromatase enzyme and genome safeguarding proteins, such as BRCA_s. The promoter regions of *ESR1*, *BRCA1*, and *CYP19* aromatase genes exhibit strong triangular partnership for the harmonized regulation of the synthesis of ERs, BRCA proteins and aromatase enzyme, which can be reconstructed from the meticulous details of earlier scientific results. Considering the extreme capacities of ER-signaling for self-restoration, it is obvious that antiestrogen treatment, either ER binding by a false ligand or inhibition of estrogen synthesis, may provoke extreme compensatory actions in genetically proficient cases. Analyses of the results of genetic studies on tumor cells have shown that upregulation of ER-signaling induced by natural estrogen or antiestrogen is a beneficial defensive process even in tumor cells, promoting their domestication and elimination. A schematic representation of the main stream of

* Corresponding Author's Email: subazdr@gmail.com (Professor Dr. Zsuzsanna Suba; Address: H-1122, Ráth György str. 7-9; Tel: 00 36 1 224 86 00; Fax: 00 36 1 224 86 20).

upregulative genome stabilizing circle visualizes the possibilities for extreme counteractions against the toxic effects of antiestrogens. The presented complex genome stabilizer mechanisms reveal that cancer cells may preserve their residual capacity for the upregulation of ER-signaling, even if the efforts are not satisfactory.

Keywords: activating mutations, antiestrogen resistance, apoptotic tumor cell death, artificial estrogens, breast cancer, cancer prevention, cancer therapy, DNA-repair, estrogen therapy, long non-coding RNAs

INTRODUCTION

A direct correlation between serum estrogen concentrations and breast cancer development was erroneously supposed and both epidemiologic and experimental studies tried to justify this preconception [1, 2]. Nevertheless, the controversial data of serum estrogen levels observed in breast cancer cases could not support the principle of estrogen induced carcinogenesis in either premenopausal or postmenopausal patients [3]. Moreover, estrogen withdrawal as an anticancer therapy could not equivocally inhibit the progression or recurrence of mammary tumors as it was expected [4].

These results seemed to show that highly complex regulatory failures may act in the background of tumor development instead of elevated estrogen concentrations [4, 5]. Despite all contradictory experiences, hypoestrogenism as an anticancer measure gained great popularity in medical practice and this misbelief led to a long-lasting erroneous pathway of breast cancer therapy throughout the 20th century, until now.

In the early 70s, tamoxifen, a synthetic ER-blocker was introduced for the treatment of ER-positive breast cancers [6]. Tamoxifen treatment resulted in low anticancer effectiveness similarly to earlier endocrine therapies, causing frequent and unexpected toxic effects, such as thromboembolic complications, stroke and malignancies particularly in the endometrium [7]. Later on, endocrine manipulations were introduced for the inhibition of estrogen biosynthesis by means of aromatase inhibitors [8]. Aromatase inhibition exhibited low effectiveness against breast cancer and frequently induced high toxicity in the targeted estrogen deficient postmenopausal breast cancer cases [9].

These failed efforts of medical practice and the pharmaceutical industry explain the confusion later experienced in association with the effects of synthetic estrogens and antiestrogens. Both types of compounds may transiently promote compensatory upregulation of ER signaling and tumor regression in short-term experiments, but unfortunately, long term treatment was seen to evoke toxic symptoms and tumor growth [10]. By contrast, natural estrogens show no toxicity and their advantageous anticancer effects are strengthened by higher doses [11]. All synthetic manipulators of ER-signaling are dangerous poisons that block ER-activation, while mimicking estrogenic activity by means of transient extreme compensatory increase of estrogen signaling. [10, 12].

In genetically proficient patients, upregulations of ER activity and estrogen synthesis caused by antiestrogens may evoke temporary strong anticancer defense mechanisms, while simultaneously promoting spontaneous tumor cell death. By contrast, in the majority of patients with mild to severe genetic disorders, antiestrogen administration induces a weak counteractive

increase of both estrogen synthesis and ER expression, which is not sufficient for restoring anticancer defense mechanisms and tumor growth inhibition [10].

Based on the lessons learned from the four decades of antiestrogen therapy, the chaotic manifestations of artificial ER-inhibition together with the encountered activating mutations aiming to restore ER-signaling have helped to disclose the exquisite significance of estrogens in human health [13].

FUNDAMENTAL ROLES OF ER SIGNALING IN THE PHYSIOLOGY OF MAMMALIANS

Over the past decades extensive knowledge has been gained on the complex regulatory functions of estrogen activated ERs in regard to the safeguarding of somatic and reproductive health in mammals. ERs activated by estrogens may act as a hub where all physiologic molecular pathways converge, allowing harmonization of their transcriptional activity in tune with all cellular functions [14].

Estrogens are synthesized by aromatase enzyme via conversion androgens to estrogens. Estradiol (E2) is the most potent and abundant estrogen in the circulation, while estrone (E1) and estriol (E3) have weaker estrogenic activities. Two estrogen receptor isoforms, ER-alpha and ER-beta are members of the nuclear receptor superfamily and they exhibit strong crosstalk and interplay. ER-beta is mainly responsible for cellular enlargement, while the role of ER-alpha is crucial during the course of cell proliferation [15]. Both ER isoforms are mandatory regulators of cellular glucose uptake since cell growth and mitotic activity require an appropriate supply of fuel for increased metabolic processes [16, 17]. Defective estrogen signaling induced by either estrogen deficiency or ER resistance leads to cellular insulin resistance [3, 18].

ER-alpha and ER-beta proteins are expressed via transcriptional activities on *ESR1* and *ESR2* genes. Estrogen bound ERs may directly operate as ligand activated transcription factor proteins on the various promoter regions of target genes. ERs can also regulate gene expression through indirect binding to deoxyribonucleic acid (DNA) via interaction with transcription factor proteins. Moreover, cell membrane associated ERs may also confer non-genomic signaling cascades to estrogen dependent target genes. Ligand independent activation of ERs can also be induced by mitogen-activated protein kinase (MAPK) or protein kinase B (Akt) pathways. Finally, genomic and non-genomic pathways of estrogen receptor signaling converge on the target genes [14].

The transcriptional activity of ERs partially results in the expression of protein coding ribonucleic acids (RNAs), whilst the vast majority of RNA transcripts are non-coding RNAs (ncRNAs) [19]. Protein coding RNA transcripts of ERs may define the synthesis of enzymes, receptors and further regulatory proteins. By contrast, ER-induced long non-coding RNA (lncRNA) transcripts are capable of promoting epigenetic gene modifications via their specific chromatin remodeling activities resulting in necessary mutations on targeted genes [20]. lncRNAs are in close interrelationship with genome stabilizer proteins, such as p53, suggesting a pivotal role of these transcripts in the promotion of genome protecting mutations [19].

Estrogens are outstanding hormones exhibiting a strong, unique upregulative feedback mechanism with their own receptors. Both low and high estrogen levels drive the increased

expression and transcriptional activity of ERs so as to restore or augment cellular ER signaling. In turn, both low and high ER expressions require upregulated estrogen synthesis for the improvement or augmentation of crucial estrogen signaling [11].

Upregulation of estrogen signaling displays a unique dichotomy through genomic stabilization, ensuring the survival and safe proliferative activity of healthy cells, while inducing spontaneous death of malignant tumor cells [13]. Both protein coding and chromatin modifier RNA transcripts of ERs may have crucial roles in the genome stabilizer machinery.

ER-SIGNALING IS IN STRONG MUTUAL CORRELATION WITH DNA STABILIZER GENES AND THEIR PROTEIN PRODUCTS

Estrogen activated ERs induce transcriptional activities on numerous estrogen regulated genes, which are responsible for the metabolic and proliferative functions of all cell types. The complexity of regulatory processes allows for estrogen-liganded ERs to choose the temporarily most suitable coregulators, promoter regions and transcriptional pathways in harmony with the optimal safeguarding of DNA replication [11].

CORRELATIONS BETWEEN ER-SIGNALING AND THE BRCA SYSTEM FOR SAFE DNA REPLICATION

BRCA1 and *BRCA2* genes have been implicated in a number of pivotal cellular functions and may be regarded as safeguards of the genome. Their ubiquitously expressed *BRCA1* and *BRCA2* protein products control DNA-replication including transcriptional processes and recombination as well as repair of DNA damages [21].

BRCA-gene mutation is associated with defects in ER-receptor expression as well as DNA safeguarding [11, 22]. The failure of genome stabilizing *BRCA*-proteins is a promoter of cancer development with specificity for highly estrogen dependent female breast and ovary [23, 24]. In *BRCA* mutation carriers, the exhibited compensatory upregulation of non-liganded ER signaling [25] may strongly reduce the risk of cancer development [11].

THE CIRCULAR FUNCTIONAL MAINSTREAM OF THREE CRUCIAL GENETIC UNITS INVOLVED IN MAINTAINING GENOMIC STABILITY

In healthy individuals, the incessant activity of regulatory genetic mechanisms ensures the metabolic and proliferative equilibrium of cellular health. In case of accidental defects occurring in any part of the system, a coordinated counteraction of numerous mediators may successfully help in the restoration of physiologic processes by means of either overexpression or hyperactivity. Even serious defects of genome stabilizer mechanisms may be kept in balance for a long duration, showing the clinical signs of good health. By contrast, due to the exhaustion

of the compensatory processes, DNA defects may develop and lead to the clinical manifestations of diseases [11].

Estrogen activated ERs are the primary initiators and organizers of the up-regulatory circle of genome stabilization in correlation and crosstalk with aromatase enzymes and genome safeguarding proteins, such as BRCA1. The promoter regions of *ESR1*, *BRCA1*, and *CYP19* aromatase genes exhibit strong triangular partnership for the harmonized regulation of the synthesis of ERs, BRCA proteins and aromatase enzyme, which can be reconstructed from the meticulous details of earlier scientific results [11, 13].

Considering the extreme capacities of ER-signaling for self-restoration, it is obvious that antiestrogen treatment, either ER binding by a false ligand or inhibition of estrogen synthesis, may provoke extreme compensatory actions in genetically proficient cases [10]. The upregulative genome stabilizing circle may be induced by antiestrogens; either ER blockers or aromatase inhibitors. In our model, the well-known *BRCA1* gene and its protein product represent all other genome stabilizing systems, which presumably exhibit similar mutual partnership with the expressions and transcriptional activities of ER-alpha and aromatase enzyme [26].

I. Transcriptional Activities of ESR1 Promoter Regions

The increased expression and activation of the *ESR1* gene are the main drivers of genome stabilizer mechanisms in both estrogen deficient and ER-resistant states.

Activated ER-alpha assembles the most effective coactivators and occupies the available *ESR1* promoter regions. The increased transcriptional activity of ER-alpha induces high expressions of protein coding ER-alpha-mRNAs and leads to a self-generating overexpression of ER-alpha synthesis [11].

At the same time, a number of activated ER-alphas occupy the promoter regions of long non-coding RNAs (lncRNAs), e.g., *HOTAIR*, which are ER responsive [27]. Highly expressed lncRNAs are capable of provoking epigenetic changes on targeted *ESR1* promoter regions resulting in activating mutations. Newly formed activating mutations of *ESR1* genes may lead to the overexpression and increased estrogen binding capacity of ERs in an estrogen deficient milieu [28]. Abundant lncRNA transcripts of ERs are capable of inducing beneficial activating mutations on *BRCA1* promoters for the upregulation of DNA-stabilization [29, 30].

During physiologic cell proliferation the increasing expression of ER-alpha strongly upregulates aromatase synthesis and leads to increased estrogen production [11]. There are no literary data supporting the capacity of activated ER-alpha to occupy the *CYP19A* promoter region and induce increased aromatase enzyme expression. It is likely that under physiologic conditions, ERs choose the safe circular pathway that induces increased aromatase expression through the activation of *BRCA1* protein for the achievement of appropriate estrogen synthesis [13].

On the contrary, it has been shown that in breast cancer cells lines, estradiol treatment can induce rapid increases of aromatase expression and activity by a nongenomic activation of ER-alpha via crosstalk with growth factor mediated pathways [31, 32]. The quick upregulation of ER-alpha displays the existence of short nongenomic autocrine loops between ER-alpha and

aromatase enzyme synthesis providing rapid increase in estrogen formation in the emergency situations of malignant tumors [13].

II. Transcriptional Activities of *BRCA1* Promoter Regions

Activated ERs have the capacity to occupy the *BRCA1* promoter regions owing to the fact that *BRCA* genes are ER-alpha responsive [23]. Increased expressions of protein-coding *BRCA1* mRNAs and elevated *BRCA1* protein synthesis are driven by ER overexpression and high transcriptional activity, ensuring the adequate safeguarding of DNA-replication [11]. In turn, *BRCA* protein abundance promotes the expression of *ESR1*, the gene of ER-alpha [22].

Newly formed, abundant *BRCA1* proteins are able to occupy the lncRNA promoters and also to increase the expression of lncRNA. The abundance of lncRNAs may provoke beneficial epigenetic changes on *ESR1* promoters causing activating mutations that result in increased ligand-activated and ligand-independent activations of ER-alpha [25, 33, 34]. Further, lncRNA transcripts of *BRCA1* may stimulate chromatin modifications on *CYP19* aromatase promoter genes and may induce highly increased aromatase synthesis and estrogen production [35, 36]. Moreover, *BRCA1* protein stimulated expressions of certain lncRNA transcripts confer activating mutations on *BRCA1* genes, ensuring greater effectiveness of newly formed *BRCA1* proteins on DNA safeguarding.

Multiple transcriptional activities of the genome stabilizer *BRCA1* protein seem to be responsible for the key balance between the expressions and activities of ER-alpha and the aromatase enzyme.

III. Transcriptional Activities of *CYP19* Aromatase Promoter Regions

Increased *BRCA1* protein expression provides the opportunity of the considerable occupancy of *CYP19* aromatase promoter genes, which are responsive to *BRCA1* protein. Increased expression of the protein coding A450 mRNA results in upregulated synthesis of the aromatase enzyme capable of converting androgens to estrogens. Increased estrogen concentrations bind and activate abundant amounts of ER-alpha, further stimulating the upregulative circle of genome stabilization [11].

lncRNA promoter activation and increased lncRNA transcription by the *BRCA1* protein may lead to activating mutation of the *CYP19* aromatase gene in order to increase aromatase synthesis and estrogen production. In *BRCA1* mutation carriers, *BRCA1* protein activity confers the selection of the appropriate *CYP19* aromatase promoter region for the compensatory intensifying of estrogen synthesis [36].

The compelling symmetry and safety of the regulatory processes are suggestive of the fact that in turn, lncRNA transcripts of the A450 aromatase protein exert a backward upregulative feedback mechanism on the *BRCA1* promoter, aiming to activate the *BRCA1* protein synthesis [13]. There are however, no available literary data that would justify the existence of this retrograde regulatory pathway.

UPREGULATIVE AND DOWN-REGULATIVE CROSSTALK BETWEEN ER-ALPHA AND BRCA1 IN PHYSIOLOGIC AND MALIGNANT CELL PROLIFERATIONS

During rapid physiologic cell proliferation, as in case of pregnancy, the increased levels of estrogens are capable of upregulating the expressions of protein coding ER-alpha mRNA transcripts and ER-alpha synthesis [37]. Meanwhile, the high ER-alpha levels upregulate the expressions of protein coding BRCA1 mRNA transcripts as well as BRCA1-protein synthesis, with the outcome of an increase in DNA stabilization [38, 39]. In turn, high BRCA1-protein levels cause further upregulation in ER-alpha expression [40] and estrogen synthesis by way of the increasing expressions of both protein coding A450 mRNA transcripts and aromatase enzyme synthesis [36]. This upregulative circle ensures strong DNA-protection for the estrogen activated ER-determined rapid proliferation and differentiation of maternal and fetal cells, whilst inducing apoptotic death in spontaneously initiated malignant cells [11].

In contrast, malignant cell proliferation manifests a self-repressing mutual down-regulation of the low and/or defective expressions of ER-alpha and BRCA1-protein [11]. Mutagenic alteration or decreased expression of ER-alpha suppresses the expression of BRCA1 mRNA transcripts and BRCA1-protein synthesis; inhibiting appropriate DNA-safeguarding [39]. In turn, decreased or defective synthesis of the BRCA1-protein leads to down-regulation of both ER-alpha mRNA expression and ER-alpha synthesis and suppresses ER-alpha signaling [41]. The down-regulative circle results in unrestrained proliferation of poorly differentiated tumor cells, which is attributable to the defect of ER signaling and the feeble control of DNA replication [11].

Moreover, the two key proteins, ER-alpha and BRCA1, are also capable of direct binding, thus mutually regulating their activities as transcriptional factors. Certain binding sites drive the upregulation of each other's transcriptional activity, while others may silence the transcriptional processes [40].

In sum, it could be established that in patients with cancer, the upregulation of ER-signaling by estradiol treatment may be the main target for the restoration of the physiologic circle of cell proliferation and controlled DNA replication [12].

TUMOR CELLS EXHIBIT REMNANTS OF THE MACHINERY OF GENOME STABILIZATION PROVIDING CIRCUMSTANCES FOR SELF-DIRECTED DOMESTICATION AND ELIMINATION

Estrogen induced upregulation of ER signaling is a beneficial process even in tumor cells, promoting their domestication and spontaneous death, whilst in case of antiestrogen administration the activation of ERs is a compensatory action in order to strengthen the remnants of genome stabilizer processes [10]. Tumor cells are capable of self-sacrifice by means of increasing the expression and activity of both ERs and the aromatase enzyme.

Intratumoral production of aromatase and estrogens was demonstrated in both cancer cells and stromal cells as well as in adipocytes adjacent to breast tumors [42, 43]. Increased aromatase activity and high in situ estrogen concentration were demonstrated at the invasive

front of cancers, where tumor-stromal interactions may define the spread or regression of cancer cells [43]. Close correlations were observed between the intense aromatase synthesis of invasive breast cancer cells and neighboring adipocytes supporting the fact that high estrogen concentration is a defensive action against the invasion of tumors even when it fails to perform a satisfactory blockade [44, 45].

Estrogen treatment has been observed to increase ER expression in tumor cells. A study was conducted involving two estrogen receptor-positive breast cancer cell lines, which were treated by four types of estrogens: estrone, estradiol, estriol, and estetrol. It was found that all four provoked significant increases in ER-expressions as compared with the untreated controls [46].

In an ER-alpha positive breast cancer cell line, estradiol treatment increased both the expression and transcriptional activity of membrane associated ER-alpha via the phosphatidylinositol 3-kinase (PI 3-K)/Akt system. At the same time, the expression and activity of nuclear ER-alpha was also enhanced by the estradiol activation of Akt [47].

Insertion of exogenous ERs into an ER-negative breast cancer cell line was shown to activate a number of estrogen-regulated genes, and treatment with estrogen led to a decrease in the invasive and metastatic potential of tumor cells [48]. This finding was the first experimental testimony of estrogen induced tumor regression.

In tumor cells, estrogen activated ERs are capable of upregulating aromatase expression. When ER-negative SK-BR-3 cells were transfected with ER alpha, aromatase activity was elevated by estradiol treatment in a dose-dependent manner. In breast cancer cells, estradiol may upregulate the expression of aromatase enzyme by a nongenomic activation of ER-alpha via growth factor-mediated pathways [31]. Estradiol may also upregulate aromatase activity in breast cancer cells by means of enhanced tyrosine phosphorylation of the enzyme mediated by c-Src kinase [32].

In situ (intracrine) formation of active sex steroids at the sites of their actions from biologically inactive precursors in the circulation have been demonstrated to play very important roles in neoplasms. In breast cancer cases, a direct correlation was observed between the aromatase activity of tumors and survival after relapse [49]. Extreme aromatase positivity of tumors showed a significant correlation with tumor grade suggesting that the higher the histologic grade of tumors, the stronger is the need for defensive estrogen synthesis. ER-alpha positive metastatic breast cancer cells demonstrated increased intracrine aromatase synthesis even after entering the blood and lymph circulation [50]. In advanced tumors, aromatase synthesis and increased estrogen concentrations provide a feeble, final possibility for the restoration of the genome stability even before the death of patient.

In breast cancer cell lines, long-term estradiol withdrawal resulted in estrogen hypersensitivity mediated by highly increased ER-alpha expression. The hypersensitive tumor cells were able to react to estrogen levels at concentrations 2-3 logs lower than required to stimulate wild type cells [51]. Acquired estrogen hypersensitivity of ER-positive tumor cells is a justification that tumors may have partially preserved or even overexpressed DNA stabilizer capacities in strongly estrogen deficient milieu [11].

Spontaneous *ESR1* gene amplification is frequent in breast malignancies. Ninety-nine percent of tumors with *ESR1*-gene amplification detected at 6q25, showed estrogen receptor protein overexpression, compared with 66% of cancers without *ESR1* amplification [52]. These observations are the manifestations of the aptitude of tumor cells to upregulate ER synthesis by means of activating mutation in the *ESR1* gene.

Overexpression of mutation activated ERs was observed in association with the proliferative activity of advanced tumors, which was mistakenly regarded as a causal factor of resistance to tamoxifen treatment [53]. In reality, the exhaustive tamoxifen treatment performed in these cases blocked all the newly formed abundant amount of ERs despite the activating mutation of the *ESR1* gene [10].

ER-alpha is capable of increasing BRCA1 protein expression even in tumor cells, since the *BRCA1* promoter is ER responsive. In MCF-7 tumor cells, a delayed upregulative expression of BRCA1 protein was achieved by estrogen treatment. The peak value was reached after 24 hours, considering that new protein synthesis takes a long time [29]. In turn, it was found that the BRCA1 protein promotes *ESR1*-gene activation and transcriptionally upregulates ER-alpha mRNA and ER-alpha expressions in breast cancer cell lines [22].

In tumor cells, ERs activated by estradiol treatment mediate the transcriptional regulation and increased expression of lncRNAs [27]. It was demonstrated that HOTAIR, a lncRNA is pervasively overexpressed in most human cancers compared with noncancerous adjacent tissues [54]. HOTAIR expression was found to vary widely in primary breast cancers [55]. Patients with high HOTAIR expression were found to have lower risks of relapse and mortality than those showing low HOTAIR expression [55]. DNA methylation was observed in association with unfavorable characteristics in tumors and HOTAIR expression was strongly increased with the methylation level of DNA.

These results suggest that in malignancies, the estradiol induced expression of lncRNAs is associated with epigenetic changes in both *ESR1* and *BRCA1* genes, with the aim to promote ER-alpha upregulation, DNA restoration and tumor remission.

CLINICAL DATA ON TUMOR RESPONSES TO ANTIESTROGEN TREATMENT

ER blockade induced by tamoxifen therapy is a potent activator of estrogen synthesis and the consequential hyperestrogenism [10]. It was demonstrated in premenopausal breast cancer patients that tamoxifen, as the single therapeutic agent, increased the circulating levels of estradiol, estrone, and progesterone 2-3 fold [56]. The found local estradiol concentrations in the breast tissue of tamoxifen-treated premenopausal women were 8.2 times higher than the levels observed in case of healthy, normally cycling women [57]. These experiences support the concept that in tamoxifen-treated women, highly elevated serum and mammary estrogen concentrations are the possible inducers of transient tumor response [10].

In the later stage of tamoxifen treatment, ovarian cysts and amenorrhoea were detected in almost half of the tamoxifen-treated premenopausal breast cancer patients, whilst serum E₂ levels were observed to be highly elevated [58]. In these symptomatically estrogen resistant patients, even the extremely elevated estrogen levels were not enough to protect the ER signaling against the advanced ER blockade of tamoxifen. Development of ovarian cysts was not induced by the erroneously presumed toxic effects of the elevated estrogen levels; but rather the advanced ER blocking effect of tamoxifen led to the pathology of ovaries [10].

Exhaustive tamoxifen treatment has been found to cause severe, frequently life-threatening toxic effects in women, including thromboembolic complications, stroke, myocardial infarct and malignancies at various sites, with particular regard to endometrial cancer [59]. These toxic

effects are mistakenly attributed to unfavorable estrogenic actions, whilst being the typical complications of subtotal artificial ER blockade [10].

Tamoxifen-induced tumor responses cannot simply be defined by elevated estrogen synthesis, considering that maintenance of ER transcriptional activity and concomitant breast cancer regression require appropriate amounts of available ERs. The high estrogen receptor content of breast cancers has been shown to demonstrate significant correlation with prolonged survival in case of tamoxifen-treated pre- and postmenopausal patients [60].

In aromatase inhibitor treated postmenopausal breast cancer patients, plasma estradiol concentrations were found to be much higher than expected, despite the noted extensive interindividual variations in increased estradiol levels [61]. These observations justify the diverse capacities for compensatory aromatase synthesis and explain the great differences between tumor responses manifested among those patients who were treated with aromatase inhibitors [10].

Among women with breast cancer, a significant direct correlation was observed between the aromatase activity of tumors and patient's survival time after relapse [49]. In young premenopausal breast cancer cases, clinical control and examination of removed tumor samples revealed that the absence of CYP19-aromatase activity carried a significantly high risk for local tumor recurrence and poor prognosis [62]. The higher the compensatory aromatase synthesis and estrogen production in breast cancers treated with aromatase inhibitor, the better is the tumor response and prognosis of the disease.

In conclusion, the tumor responses among antiestrogen treated patients are with certainty defined by the genetic capacity of the individual cases for strong upregulation of ER expression and estrogen synthesis against highly toxic medicaments.

EXPERIMENTAL DATA ON TUMOR RESPONSES TO ANTIESTROGEN TREATMENT

In antiestrogen treated tumors, the compensatory increase in ER expression and aromatase enzyme via activating mutations assists the domestication and self-directed elimination of tumors as well as the survival of patients [13].

In tumor cells, long-term aromatase inhibitor-induced estradiol deprivation causes a four to tenfold up-regulation of the amount of ER α and a strong increase in the basal transcription level of several estradiol stimulated genes. On the other hand, the rapid, ligand independent activation of ERs further enhances adaptive estrogen hypersensitivity by means of an increased utilization of plasma membrane mediated pathways, including the activation of MAP kinase as well as PI-3-kinase and mTOR [63].

ER-positive breast cancer cells may adapt to an estrogen deficient environment via the increased expression of noncoding RNAs, which may perform epigenetic activating changes on the *ESR1* gene locus, with an abundance of ER overexpression as a result [28].

In antiestrogen resistant MCF-7/Ral tumors, long-term tamoxifen or raloxifene treatment continuously stimulated ER expression, while the growth of tumors was strongly progressive. Following a 5-week estradiol treatment, estradiol statistically significantly reduced the size of tumors earlier stimulated by raloxifene or tamoxifen pretreatment [64]. These findings indicate that the exhaustive administration of ER-blockers increases ER expression in tumor cells, while

ER blockade by a false ligand decreases the accessible receptor mass and promotes the growth of tumors. By contrast, estradiol treatment of antiestrogen resistant tumors may induce tumor regression by the extreme upregulations of both ER expression and transcriptional activity [10].

In ER-positive breast cancers, the increased level of *ESR1* mRNA was the strongest linear predictor of benefit from tamoxifen, whilst a low level of *ESR1* mRNA expression was a determinant of tamoxifen resistance [65]. Activating *ESR1* mutations in tumor cells were shown to result in constitutive activity and continued responsiveness to antiestrogen therapies in vitro [66]. Moreover, in breast cancer patients, adjuvant tamoxifen monotherapy resulted a significantly longer survival time for women with cancers exhibiting strong *ESR1* gene amplification [52].

On the contrary, activating *ESR1* mutations seemed to be a key event in the aggressive behavior of antiestrogen treated tumors [67]. In reality, in case of progressive tumors showing activating *ESR1* mutation even the defensive increase in ER expression is not enough to protect against the advanced ER-blockade of exhaustive antiestrogen therapy.

In breast cancer cell lines, long-term tamoxifen exposure strongly increased aromatase promoter activity and aromatase expression via G-protein-coupled receptor GPR30/GPER [68]. Abundance of aromatase synthesis and high estrogen concentrations may be persistent even in the exhaustive phase of tamoxifen treatment, although they are not satisfactory for the reactivation of near completely blocked ER signaling.

In conclusion, the activating mutations of ER-positive tumor cells against estrogen loss or ER-blockade may be attributed to the remnants of genome stabilizer processes to achieve self-directed tumor death instead of to an aggressive willpower for tumor survival [13].

CONCLUSION

The erroneous practice of using antiestrogens over the long term has revealed that the majority of women with less than optimal genetic reserve capacities do not respond well to antiestrogen treatment, which experience however has mistakenly been regarded as a de novo antiestrogen resistance. Even among genetically proficient breast cancer cases that show good tumor response initially, long-lasting antiestrogen treatment leads to the exhaustion of compensatory mechanisms with manifestations of toxic symptoms and progression of the disease. This phase is erroneously regarded as acquired antiestrogen resistance, but instead it is rather an almost complete blockade of ER-signaling.

Analyses of the results of genetic studies on tumor cells have shown that upregulation of ER-signaling induced by natural estrogen or antiestrogen is a beneficial defensive process even in tumor cells, promoting their domestication and elimination. The presented scheme of complex genome-stabilizing mechanisms demonstrates that cancer cells may preserve their residual capacity for the upregulation of ER-signaling, even if the efforts are not satisfactory. Activating mutations affecting ER functions have key roles in both the genome stabilizer machinery of healthy cells and the restoration of altered genetic pathways of DNA-repair in tumor cells.

REFERENCES

- [1] Brown SB, Hankinson SE. Endogenous estrogens and the risk of breast, endometrial, and ovarian cancers. *Steroids* 2015; 99(Pt A): 8-10.
- [2] Endogenous Hormones and Breast Cancer Collaborative Group. Sex hormones and risk of breast cancer in premenopausal women: a collaborative reanalysis of individual participant data from seven prospective studies. *Lancet Oncol* 2013; 14(10):1009-19.
- [3] Suba Z. Estrogen withdrawal by oophorectomy as presumed anticancer means is a major medical mistake. *J Family Med Community Health* 2016; 3(3): 1081-7.
- [4] Suba Z. Autoregulative Circle of ER Signaling Drives the Machinery of Genome Stabilization. *J Cancer Biol Res* 2016; 4(4): 1089-97.
- [5] Suba Z. Diverse pathomechanisms leading to the breakdown of cellular estrogen surveillance and breast cancer development: new therapeutic strategies. *Drug Design Devel Ther* 2014; 8: 1381-90.
- [6] Jordan VC, Dowse LJ. Tamoxifen as an antitumour agent: effect on oestrogen binding. *J Endocrinol* 1976; 68: 297-303.
- [7] Hayes DF. Tamoxifen: Dr. Jekyll and Mr. Hyde? *J Natl Cancer Inst* 2004; 96: 895-7.
- [8] Dixon JM. Endocrine resistance in breast cancer. *New J Science* 2014; ID 390618, 27 pp.
- [9] Larionov AA, Miller WR. Challenges in defining predictive markers for response to endocrine therapy in breast cancer. *Future Oncology* 2009; 5(9): 1415-28.
- [10] Suba Z. The pitfall of the transient, inconsistent anticancer capacity of antiestrogens and the mechanism of apparent antiestrogen resistance. *Drug Des Devel Ther* 2015; 9: 4341-53.
- [11] Suba Z. DNA stabilization by the upregulation of estrogen signaling in BRCA gene mutation carriers. *Drug Des Devel Ther* 2015; 9: 2663-75.
- [12] Suba Z. Causal Therapy of Breast Cancer Irrelevant of Age, Tumor Stage and ER-Status: Stimulation of Estrogen Signaling Coupled With Breast Conserving Surgery. *Recent Pat Anticancer Drug Discov* 2016; 11(3): 254-66.
- [13] Suba Z. Activating mutations of ESR1, BRCA1 and CYP19 aromatase genes confer tumor response in breast cancers treated with antiestrogens. *Recent Pat Anticancer Drug Discov.* 2017; 12(2): 136-147.
- [14] Maggi A. Liganded and unliganded activation of estrogen receptor and hormone replacement therapies. *Biochim Biophys Acta* 2011; 1812(8): 1054-1060.
- [15] Helguero LA, Faulds MH, Gustafsson JA, Haldosén LA. Estrogen receptors alfa (ERalpha) and beta (ERbeta) differentially regulate proliferation and apoptosis of the normal murine mammary epithelial cell line HC11. *Oncogene* 2005; 24(44): 6605-16.
- [16] Suba Z. Low estrogen exposure and/or defective estrogen signaling induces disturbances in glucose uptake and energy expenditure. *J Diabet Metab* 2013; 4: 272-81.
- [17] Suba Z. Circulatory estrogen level protects against breast cancer in obese women. *Recent Pat Anticancer Drug Discov* 2013; 8(2): 154-67.
- [18] Suba Z. Interplay between insulin resistance and estrogen deficiency as co-activators in carcinogenesis. *Pathol Oncol Res* 2012; 18(2): 123-33.

- [19] Baldassarre A, Andrea Masotti A. Long Non-Coding RNAs and p53 Regulation. *Int J Mol Sci* 2012; 13(12): 16708-17.
- [20] Gupta RA, Shah N, Wang KC, Kim J, Horlings HM, Wong DJ, et al. Long non-coding RNA HOTAIR reprograms chromatin state to promote cancer metastasis. *Nature* 2010; 464(7291): 1071-6.
- [21] Venkitaraman AR. Cancer susceptibility and the functions of BRCA1 and BRCA2. *Cell* 2002; 108: 171–182.
- [22] Hosey AM, Gorski JJ, Murray MM, Quinn JE, Chung WY, Stewart GE. et al. Molecular basis for estrogen receptor alpha deficiency in BRCA1-linked breast cancer. *J Natl Cancer Inst* 2007; 99:1683–94.
- [23] Gorski JJ, Kennedy RD, Hosey AM, Harkin DP. The complex relationship between BRCA1 and ERalpha in hereditary breast cancer. *Clin Cancer Res* 2009; 15(5):1514-8.
- [24] Wang L, Di LJ. BRCA1 and estrogen/estrogen receptor in breast cancer: where they interact? *Int J Biol Sci* 2014; 10(5):566-75.
- [25] Zheng L, Annab LA, Afshari CA, Lee WH, Boyer TG. BRCA1 mediates ligand-independent transcriptional repression of the estrogen receptor. *Proc Natl Acad Sci USA* 2001; 98:9587–92.
- [26] Suba Z. Bidirectional crosstalk between estrogen signaling pathways and DNA repair machinery. Chapter 1. In: *Advances in Carcinogenesis Research*. Ed: Yamaguchi M. Nova Science Publishers Inc. New York, 2017.
- [27] Bhan A, Mandal SS. Estradiol-Induced Transcriptional Regulation of Long Non-Coding RNA, HOTAIR. *Methods Mol Biol* 2016; 1366: 395-412.
- [28] Tomita S, Abdalla MO, Fujiwara S, Matsumori H, Maehara K, Ohkawa Y, Iwase H, Saitoh N, Nakao M et al. A cluster of noncoding RNAs activates the ESR1 locus during breast cancer adaptation. *Nat Commun* 2015; 6: 6966.
- [29] Kininis M, Chen BS, Diehl AG, Isaacs GD, Zhang T, Siepel AC. et al. Genomic analyses of transcription factor binding, histone acetylation, and gene expression. *Mol Cell Biol* 2007; 27: 5090–104.
- [30] Jeffy BD, Hockings JK, Kemp MQ, et al. An estrogen receptor-alpha/p300 complex activates the BRCA-1 promoter at an AP-1 site that binds Jun/Fos transcription factors: repressive effects of p53 on BRCA-1 transcription. *Neoplasia* 2005; 7: 873–82.
- [31] Kinoshita Y, Chen S. Induction of aromatase (CYP19) expression in breast cancer cells through a nongenomic action of estrogen receptor alpha. *Cancer Res* 2003; 63(13): 3546-55.
- [32] Catalano S, Barone I, Giordano C, Rizza P, Qi H, Gu G, et al. Rapid estradiol/ER alpha signaling enhances aromatase enzymatic activity in breast cancer cells. *Mol Endocrinol* 2009; 23(10): 1634-45.
- [33] Wang C, Fan S, Li Z, Fu M, Rao M, Ma Y, et al. Cyclin D1 antagonizes BRCA1 repression of estrogen receptor alpha activity. *Cancer research* 2005; 65: 6557–67.
- [34] Fan S, Wang J, Yuan R, et al. BRCA1 inhibition of estrogen receptor signaling in transfected cells. *Science* 1999; 284: 1354–6.

- [35] Simpson ER, Dowsett M. Aromatase and its inhibitors: significance for breast cancer therapy. *Recent Prog Horm Res* 2002; 57: 317–338.
- [36] Ghosh S, Lu Y, Katz A, Hu Y, Li R. Tumor suppressor BRCA1 inhibits a breast cancer-associated promoter of the aromatase gene (CYP19) in human adipose stromal cells. *Am J Physiol Endocrinol Metab* 2007; 292: 246–252.
- [37] Schumacher A, Costa SD, Zenclussen AC. Endocrine factors modulating immune responses in pregnancy. *Front Immunol* 2014; 5: 196.
- [38] Farmer H, McCabe N, Lord CJ, Tutt AN, Johnson DA, Richardson TB, Santarosa M, Dillon KJ, Hickson I, Knights C, Martin NM, Jackson SP, Smith GC, Ashworth A. Targeting the DNA repair defect in BRCA mutant cells as a therapeutic strategy. *Nature* 2005; 434: 917–21.
- [39] Spillman M., Bowcock A. BRCA1 and BRCA2 mRNA levels are coordinately elevated in human breast cancer cells in response to estrogen. *Oncogene* 13: 1639–45, 1996.
- [40] Russo J, Russo IH. Toward a unified concept of mammary carcinogenesis Aldaz M.C. Gould M.N. McLachlan J. Slaga T.J. eds. *Progress in Clinical and Biological Research*: 1–16, Wiley-Liss New York 1997.
- [41] Fan S, Ma YX, Wang C, et al. p300 Modulates the BRCA1 inhibition of estrogen receptor activity. *Cancer Res* 2002; 62: 141–51.
- [42] Suzuki T, Miki Y, Akahira JI, Moriya T, Ohuchi N, Sasano H. Review: Aromatase in human breast carcinoma as a key regulator of intratumoral sex steroid concentrations. *Endocr J* 2008; 55: 455–63.
- [43] Sasano H., Miki Y., Nagasaki S., Suzuki T. In situ estrogen production and its regulation in human breast carcinoma: From endocrinology to intracrinology. *Pathol Int* 2009; 59: 777–89.
- [44] Suba Z. Circulating and Local Estrogen Concentrations are Protective against Breast Cancer in Obese Women. In: *Topics in Anti-Cancer Research*. Patents eBook Series. Vol. 4. Eds: Rahman A & Zaman K. Pp. 476–523 (48). Online.
- [45] Suba Z. Defective estrogen signaling is the highest risk for breast cancer. *Int J Cancer Res Prev* 2015; 8(2): 209–225.
- [46] Liu S, Ruan X, Schultz S, Neubauer H, Fehm T, Seeger H, et al. Oestrol stimulates proliferation and oestrogen receptor expression in breast cancer cell lines: comparison of four oestrogens. *Eur J Contracept Reprod Health Care* 2015; 20(1): 29–35.
- [47] Stoica GE, Franke TF, Moroni M, Mueller S, Morgan E, Iann MC, et al. Effect of estradiol on estrogen receptor- α gene expression and activity can be modulated by the ErbB2/PI 3-K/Akt pathway. *Oncogene* 2003; 22(39): 7998–8011.
- [48] Garcia M, Derocq D, Freiss G, Rochefort H. Activation of estrogen receptor transfected into a receptor-negative breast cancer cell line decreases the metastatic and invasive potential of the cells. *Proc Natl Acad Sci USA* 1992; 89(23): 11538–42.
- [49] Evans TRJ, Rowlands MG, Silva MC, Law M, Coombes RC. Prognostic significance of aromatase and estrone sulfatase enzymes in human breast cancer. *J Steroid Biochem Mol Biol* 1993; 44: (4–6): 583–7.

- [50] Mukhopadhyay KD, Liu Z, Bandyopadhyay A, Kirma NB, Tekmal RR, Wang S, Sun LZ. Aromatase expression increases the survival and malignancy of estrogen receptor positive breast cancer cells. *PLoS One* 2015; 10(4): e0121136.
- [51] Santen RJ, Song RX, Zhang Z, Kumar R, Jeng M-H, Masamura A et al. Long-term estradiol deprivation in breast cancer cells up-regulates growth factor signaling and enhances estrogen sensitivity. *Endocrine-Related Cancer* 2005; 12: S61–S73.
- [52] Holst F, Stahl PR, Ruiz C, Hellwinkel O, Jehan Z, Wendland M, et al. Estrogen receptor alpha (ESR1) gene amplification is frequent in breast cancer. *Nat Genet* 2007; 39(5): 655–60.
- [53] Merenbakh-Lamin K, Ben-Baruch N, Yeheskel A et al. “D538G mutation in estrogen receptor- α : a novel mechanism for acquired endocrine resistance in breast cancer,” *Cancer Res* 2013; 73(23): 6856–6864.
- [54] Zhang J, Zhang P, Wang L, Piao HL, Ma L. Long non-coding RNA HOTAIR in carcinogenesis and metastasis. *Acta Biochim Biophys Sin (Shanghai)* 2014; 46(1): 1-5.
- [55] Lu L, Zhu G, Zhang C, Deng Q, Katsaros D, Mayne ST, et al. Association of large noncoding RNA HOTAIR expression and its downstream intergenic CpG island methylation with survival in breast cancer. *Breast Cancer Res Treat* 2012; 136(3): 875–83.
- [56] Jordan VC, Fritz NF, Langan-Fahey S, Thompson M, Tormey DC. Alteration of endocrine parameters in premenopausal women with breast cancer during long-term adjuvant therapy with tamoxifen as the single agent. *J Natl Cancer Inst* 1991; 83(20): 1488-91.
- [57] Depypere HT, Bolca S, Bracke M, Delanghe J, Comhaire F, Blondeel P. The serum estradiol concentration is the main determinant of the estradiol concentration in normal breast tissue. *Maturitas* 2015; 81(1): 42-5.
- [58] Metindir J, Aslan S, Bilir G. Ovarian Cyst Formation in Patients Using Tamoxifen for Breast Cancer. *Japanese J Clin Onc* 35: (10): 607-611.
- [59] Braithwaite RS, Chlebowski RT, Lau J, George S, Hess R, Col FN: Meta-analysis of vascular and neoplastic events associated with tamoxifen. *J General Int Med* 2003; 18: 937-947.
- [60] Skoog L, Humla S, Axelsson M, Frost M, Norman A, Nordenskjöld B, et al. Estrogen receptor levels and survival of breast cancer patients. A study on patients participating in randomized trials of adjuvant therapy. *Acta Oncol* 1987;26(2):95-100.
- [61] Kunovac KT, Baumgart J, Stavreus EA, Sundström PI, Moby L, Kask K, et al. Higher than expected estradiol levels in aromatase inhibitor-treated, postmenopausal breast cancer patients. *Climacteric*. 2012; 15(5): 473-80.
- [62] Bollet MA, Savignoni A, De Koning L, et al. Tumor aromatase expression as a prognostic factor for local control in young breast cancer patients after breast-conserving treatment. *Breast Cancer Res* 2009; 11(4): R54.
- [63] Santen RJ, Song RX, Masamura S, Yue W, Fan P, Sogon T, Hayashi S et al. Adaptation to estradiol deprivation causes up-regulation of growth factor pathways and hypersensitivity to estradiol in breast cancer cells. *Adv Exp Med Biol* 2008; 630: 19-34.

- [64] Liu H, Lee ES, Gajdos C, Pearce ST, Chen B, Osipo C, Loweth J, McKian K, De Los Reyes A, Wing L, Jordan VC. Apoptotic action of 17beta-estradiol in raloxifene-resistant MCF-7 cells in vitro and in vivo. *J Natl Cancer Inst.* 2003; 95(21):1586-97.
- [65] Kim C, Tang G, Pogue-Geile KL, Costantino JP, Baehner FL, Baker J, et al. Estrogen receptor (ESR1) mRNA expression and benefit from tamoxifen in the treatment and prevention of estrogen receptor-positive breast cancer. *J Clin Oncol* 2011; 29(31): 4160-7.
- [66] Robinson DR, Wu Y-M, Vats P, Su F, Lonigro RJ, Xuhong Cao X et al. Activating *ESR1* mutations in hormone-resistant metastatic breast cancer. *Nature Genetics.* 2013; Vol: 45: pp: 1446–51.
- [67] Gelsomino L, Gu G, Rechoum Y, Beyer AR, Pejerrey SM, Tsimelzon A, et al. ESR1 mutations affect anti-proliferative responses to tamoxifen through enhanced cross-talk with IGF signaling. *Breast Cancer Res Treat* 2016; 157(2): 253-65.
- [68] Catalano S, Giordano C, Panza S, Chemi F, Bonofiglio D, Lanzino M et al. Tamoxifen through GPER upregulates aromatase expression: a novel mechanism sustaining tamoxifen-resistant breast cancer cell growth. *Breast Cancer Res Treat.* 2014; 146(2): 273-85.

Chapter 75

MOTOR ABILITIES ARE RELATED TO SPECIFIC GENOTYPES IN RETT SYNDROME

***R. A. Fabio^{1,*}, T. Capri¹, M. Lotan^{2,3,4},
G. E. Towey¹ and G. Martino¹***

¹Cognitive Empowerment Laboratory,
Department of Cognitive Science, University of Messina,
Via Concezione, Messina, Italy

²Israeli Rett Center, National Evaluation Team,
Chaim Sheba Medical Center, Tel HaShomer, Ramat Gan, Israel

³Department of Physical Therapy, Ariel University, Ariel, Israel

⁴Association for Children at Risk,
An Israeli Program for Autistic Children, Tel-Aviv, Israel

ABSTRACT

Rett syndrome (RTT) is a rare, neurodevelopmental genetic disorder that develops in early childhood and influences many functions within neurobehavioural domains. The core of phenotype symptoms includes severe linguistic and motor impairments. The onset of RTT is characterised by a gradual or sudden loss of speech and hand function followed by a slow decrease in acquired gross motor skills with subsequent severe functional dependence. RTT is associated primarily with mutations in MECP2, a gene located on the long arm of the X chromosome (Xq28). The severity of impairments depends not only on genotype, but on the extent of X inactivation. However, CDKL5 and FOXP1 gene mutations have been also identified in girls affected by atypical RTT.

Despite sharing neurological features, subjects with RTT present considerable clinical variability. Research on effects of genotype of RTT is expanding in many directions. The current chapter, will discuss the correlations between genotype and motor abilities in subjects with RTT.

The main aim of this chapter is to relate functional outcomes, in particular motor impairments, to mutation type in patients with RTT. This chapter begins with a theoretical overview on genetic alterations in RTT and then focuses on motor specific impairments.

* Corresponding Author's Email: rafabio@unime.it (Tel: +39090344831).

In the second part of this chapter, we propose a preliminary research which analyzes the correlation between motor phenotype and specific genotype.

1. GENERAL INTRODUCTION

Establishing the genotype–phenotype correlations is a central goal in order to acquire a deeper understanding of genetic disorders. Studies on Rett Syndrome (RTT) are analyzing how abnormalities in specific genes can cause disease in motor abilities. Consequently, the main aim of the present study is to correlate the specific genotype mutation with different parameters of motor impairment. In the first paragraph, entitled “Genetics and Clinical Features of RTT”, we consider new findings in the genetics field related to RTT. In the second paragraph, entitled “Motor features in RTT”, we discuss motor specific impairments in RTT. In the third paragraph, we present a study in which a large sample (195 girls with RTT) with specific motor disability (in walking, use of hands, problems related to the feet and scoliosis) is analyzed in relationship with their specific genotype.

1.1. Genetics and Clinical Features of RTT

RTT is an X-linked genetic disorder affecting almost exclusively females. It is caused by mutations in the gene encoding the methyl CpG binding protein 2 (MECP2) (Hagberg, Aicardi, & Ramos, 1983; Amir, Van den Bellver, Wan, Tran, Francke, & Zoghbi, 1999). Clinical features of classical RTT include a characteristic developmental regression involving impairment of mobility and loss of learnt speech and skilled intentional hand movements, accompanied by stereotypic hand movement automatisms (Ross et al., 2016). Associated features, such as microcephaly, respiratory/autonomic abnormalities, seizures, growth deficits and early hypotonia are highly present (Neul et al., 2010; Fabio et al., 2014; 2016).

The course of this syndrome is divided into four characteristic stages. After an apparent normal development during infancy, the girls with RTT go through a period of developmental regression (Fabio, Antonietti, Marchetti, & Castelli, 2009a; 2009b). Once developmental regression begins, the individuals with RTT usually progress through four stages: early onset deceleration stage, rapid destructive stage, pseudostationary stage, and motor deterioration stage. After the regression most children are unable to walk, talk, or use their hands for functional activities, but there is considerable variability in this regression, mainly determined by genotype (Sigafos, Green, Schlosser, O’Eilly, Lancioni, Rispoli, et al., 2009). The girls who have some, but not all of the necessary criteria to be diagnosed with RTT, are categorized as atypical or as variant forms (Hagberg & Witt-Engerström, 1986).

The term ‘variant’ was coined to describe an onset of Rett like phenotypes that deviate from the classical clinical presentation of this syndrome (Leonard, Cobb & Downs, 2017). These phenotypes included: fruste forms, congenital variant (mostly caused by mutations in FOXG1; Ariani et al., 2008) early seizure variant (mainly caused by mutation in cyclin-dependent kinase-like 5 (CDLK5; Fehr, 2013), male, late childhood, regression and preserved speech variant (mostly associated with Arg133Cys mutation; Urbanowicz, Downs, Girdler, Ciccone & Leonard, 2015) or a C-terminal deletion (Zappella, 1992; De Bona et al., 2000).

The fruste form presents a later age of onset compared with the classical form, with regression occurring between 1 to 3 years of age, hand use is sometimes preserved with minimal stereotypic movements (Chahrouh1 & Zoghbi, 2007). The preserved speech variant is characterized by the ability to speak a few words, although not necessarily in context, and by normal head size (Zappella et al., 2001). The congenital form, the more severe variant, is characterized by the lack of the early period of normal development. Another severe variant is a form of classical RTT with onset of seizures before the age of 6 months, and was associated with mutations in CDKL5 (Ariani et al., 2008; Evans, Archer, Colley, Ravn, Nielsen, Kerr, et al., 2005; Larsson, 2013; Pini, Bigoni, Engerström, Calabrese, Felloni, Scusa et al., 2012; Rajaei, Erlandson, Kyllerman, Albage, Lundstrom, & Karrstedt et al., 2011).

Considering the heterogeneity of RTT, it is important to investigate and establish the genotype-phenotype correlations in order to obtain a clear clinical picture of RTT.

1.2. MECP2 Mutations and Functions

In 1999, a mutation was found on the gene MECP2 locus Xq28, which represent the main underlying cause of RTT (Amir et al., 1999). In fact, in the majority of cases RTT is caused by de novo mutations in the MECP2 gene. This gene is a transcriptional repressor involved in chromatin remodeling and the modulation of RNA splicing (Chahrouh, & Zoghbi, 2007). It encodes methyl-CpG binding protein 2 (Amir, Van den Bellver, Wan, Tran, Francke, & Zoghbi, 1999), an abundant nuclear protein that is considered to be important in chromatin-level regulation of transcription (Lyst & Bird, 2015).

Moreover, MeCP2 mediates transcriptional inhibition by binding to methylated CpG and CpA dinucleotides in the genome (Guo et al., 2014; Gabel et al., 2015) and recruiting co-repressor complexes (Nan, 1998; Kokura, 2001). MeCP2 may also function as an activator of transcription (Li et al., 2013) amongst other functions (Lyst & Bird, 2015).

MECP2 comprises of four exons that code for two different isoforms of the protein, due to alternative splicing of exon 2. The more abundant MeCP2-e1 isoform contains 24 amino acids encoded by exon 1 and lacks the 9 amino acids encoded by exon 2. The start site for the MeCP2-e2 isoform is in exon 2 (Chahrouh1 & Zoghbi, 2007; Dragich et al., 2007; Kriaucionis and Bird, 2004; Mnatzakanian et al., 2004).

MeCP2 is essential for normal brain function. Aberrations in this gene result in a subset of the commonly observed symptoms of RTT. Precisely, the loss of MeCP2 disrupts the given brain region or system from which it is deleted. Deletion from GABAergic circuits produces a near-complete Mecp2-null phenotype, including motor and cognitive impairments (Chao et al., 2010). Postnatal deletion of Mecp2 causes RTT-like phenotypes (Cheval et al., 2012).

It is possible to conclude that disruptions in MECP2 are the main cause of classic and variant forms of RTT. However, MeCP2 depletion studies have revealed that MeCP2 mutations are associated with other neuropsychiatric disorders, such as learning disabilities, autism spectrum disorders (Carney et al., 2003; Lam et al., 2000), bipolar disease (Klauck et al., 2002) and juvenile-onset schizophrenia (Cohen et al., 2002). In addition, mutations in MECP2 gene can also cause severe mental retardation with epilepsy and Angelman-like syndrome in females (Milani et al., 2005; Watson et al., 2001).

Given this genetic and clinical variability, the need for genotype-phenotype studies is a great opportunity for new discoveries.

1.3. CDKL5 and FOXG1 Gene Mutations

In literature, MECP2, CDKL5, and FOXG1 have been reported to be the causative genes of RTT. Mutations in the CDKL5 gene were often associated with a rare congenital disorder, characterized by intractable early-onset seizures and RTT-like features (Chahrour1 & Zoghbi, 2007). However, FOXG1 mutations are the rarest and least studied. According to the FOXG1 variant database of the International Rett Syndrome Foundation only 20 pathogenic variants have been found to date (Christodoulou, Grimm, Maher, & Bennetts, 2003; Christine, Lee, Lim, Kim, Hwang & Chae, 2015).

Kortum and colleagues (2011) have delineated the phenotype of FOXG1 mutation in positive patients. The FOXG1 phenotype is characterized by severe mental retardation, hypotonia, absent language, postnatal microcephaly, dyskinesia, corpus callosum hypogenesis, developmental epilepsy, mental retardation, and severe speech impairment (Christine, Lee, Lim, Kim, Hwang & Chae, 2015; Kortum et al., 2011; Philippe et al., 2010). This phenotype is called congenital variant of RTT. Although, many features of patients with FOXG1 mutation are similar to RTT, other clinical features differ from the classical presentation of RTT. For example, the presence of dyskinesias and brain imaging abnormalities, as well as the lack of regression and lack of respiratory arrhythmia (Kortum et al., 2011). Based on this onset of symptoms and the combination of developmental and brain imaging features observed in individuals with FOXG1 mutations, Kortum and colleagues (2011) have argued that it is possible to identify the FOXG1 phenotype as a FOXG1 syndrome and not as a RTT variant.

1.4. Severity and Genotype-Phenotype Relationships

The MECP2 mutation is related to clinical severity and influences many aspects of the phenotype. The severity of impairments depends not only on genotype, but on the extent of X inactivation. Phenotype-genotype correlation studies have shown that the most severe mutations are associated with Arg270X, Arg255X and Arg168X, whereas Arg133Cys, Arg294X and C-terminal deletions are associated with less-severe phenotypes (Leonard, Cobb & Downs, 2017). Precisely, the girls with severe mutations are less able to walk, retain hand use or use words, and tend to be diagnosed at an earlier age (Fehr, 2011). Whereas individuals with C-terminal deletions show later loss of skills and onset of stereotypies (Bebbington et al. 2010). Moreover, it has been found that R133C mutation causes a milder phenotype (Kerr et al., 2006; Leonard et al., 2003; Neul et al., 2007), whereas the R270X mutation is related to increased mortality (Bienvenu & Chelly, 2006).

2. MOTOR FEATURES IN RETT SYNDROME

As reported previously RTT is a disorder that causes a neurological and developmental arrest which manifests itself in a variety of disabilities, such as loss of functional hand use, loss of acquired speech, apraxia, ataxia, autonomic system dysfunction, epilepsy, breathing abnormalities, failure to thrive and muscle tone irregularities (Hagberg, 1993; Lotan, 2006; Kerr, & Julu, 1999). With regard to motor abilities, patient present alterations in the motor

system that are observed in the first months of life (Einspieler, Kerr, & Prechtl, 2005; Nomura, & Segawa, 1990); despite the fact that some girls gain the ability to walk, eventually they suffer from a motor deterioration that impairs gait, mobility and balance and compromises both upper and lower extremities.

2.1. Motor Characteristics According to Rett Syndrome Stages

Seen as the symptoms follow a certain course, the syndrome is described as divided into stages. Some of these overlap, and in some cases they do not present classic signs that fulfill the necessary criteria, therefore are classified as atypical (Trevathan, 1989). In later revisions one of the main aspects that has been excluded is the “sudden” onset of the pathology (Hagberg, 1993; Hagberg, & Witt Engerstrom, 1986); indeed the regression is not always clearly visible, and may be more subtle and gradual (Fehr et al., 2011; Jackowski et al., 2011; Lotan, 2006). Many revisions of the clinical criteria have been made in the last decade, allowing a clarification of the “variant” forms, such as the “early seizure onset variant” (CDKL5 disorder; Fehr, et al. 2013) and the the *Zappella* or preserved speech variant (Zappella, 1992).

Despite the fact that the classification of the stages is not as restrictive as before, there are common features, especially with regard to how motor skills are affected; this section will summarize the characteristic of each phase with particular emphasis on the motor abilities.

Stage I: Onset. Most children at this stage are not yet diagnosed as having RTT; it begins around the age of six to eighteen months, and most often, patients are diagnosed as having hypotonia, cerebral palsy, autism or both (Leonard, Bower & English, 1997). The child will display a slow rate of motor development, and may exhibit attempts at walking, but, because of the flaccid muscle tone, many don't reach this major milestone.

Stage II: The rapid destructive phase. The regression and the loss of acquired developmental milestones is the significant characteristic of stage II. The child (around one to three years old) may still not be diagnosed as having RTT or a final diagnosis may still be pending. The symptoms of this stage can include seizures, deceleration of head growth, irregular breathing or hyperventilation and initiation of spinal asymmetry. The main characteristic of this phase is the loss of purposeful hand use, which is replaced by the stereotypic hand movements.

Stage III: Pseudo-stationary stage. Is characterized by a relatively calm period, nevertheless, during this stage, most of the individuals with RTT develop significant problems, such as deformities and contractures (Larsson, & Engerstrom, 2001). The difficulty in motor planning is affected by ataxia and apraxia, and deambulation is limited because of spasticity and scoliosis.

Stage IV: Late motor deterioration. Significant decrease and lack of mobility is characteristic of stage IV. Many of the individuals with RTT, although non-ambulatory, may still be able to participate in a supported transfer (e.g., moving from a wheel-chair to a bed). Individuals who constantly use a wheel-chair will typically present a worse phenotypic expression, consequentially requiring more care. In some cases, however, additional medical issues are encountered. Such problems may include worsening scoliosis and contractures, lower extremity atrophic changes (Lotan, 2006), and growth retardation will usually be more apparent

(Motil, Schultz, Wong & Glaze, 1998). Often patients may present Parkinsonian features (Hagberg, 2005; Roze et al., 2007).

2.2. Neurological and Muscle Symptoms

Since the early description of RTT, it was clear that it referred to neurodevelopmental disorders and not a progressive encephalopathy (Larsson, 2013). The atrophy or the reduced dendritic/synaptic development causes the brain to be smaller. Studies which refer specially to the motor abilities have found a link between the reduction in size of the cerebellum and motor deficits in RTT. For example, in studies carried out on modified mice including *Mecp2*-knockout lines with early motor problems, major deterioration was found especially in the motor control regions, e.g., prefrontal and motor areas, with the involvement of noradrenergic and serotonergic systems (Santos, et al., 2010). Julu et al. (2008) have associated most of the clinical aspects of the syndrome with an immature brain, and some of these present an extra-pyramidal origin.

Different neurological features may impact motor function, in particular hypotonia and weakness in the early years and dystonia and bradykinesia in the later years (Hagberg, 2002). Fehr et al. (2011) have found that in the early stages of the syndrome, patients may learn to sit and walk, but the subsequent alteration of the muscle tone (e.g., bradykinesia) influences the retention of upright posture, and impedes deambulation. Difficulties in the motor coordination and the confusion reported concerning RTT patients, is due to the ataxia and gait apraxia (Larsson, 2013). These symptoms encumber the production of what we perceive as the simplest actions, such as sitting down or lowering from kneeling to heel sitting. As with other characteristics of RTT, variability among individuals is the norm. The infant with RTT will typically show hypotonia or the muscle tone might be within the low margin of the norm. With age, most individuals with RTT change from being hypotonic to becoming hypertonic (mostly rigidity, usually starting from lower extremities). 30% of adults with RTT remain hypotonic, 40% show rigidity as their main tonal characteristic and 30% become dystonic (Kerr, 1992). These changes influence the handling of the child and therefore necessitate ongoing evaluation of muscle tone by the therapist.

When muscle tone of the individual with RTT gradually changes from low to high, the change might cause a unique phenomenon. In contrast to children with cerebral palsy (toe walking), the child with RTT might present an asymmetrical reaction, causing her to side-tilt the trunk, leading to initial development of asymmetry of the spine, eventually resulting in scoliosis (Figure 1). Another unique reaction of individuals with RTT to the shortening of the Achilles tendon could be the tilting of the pelvis backwards, causing the child to lose the ability to walk independently. When an appropriate foresight planning is performed, such events should be expected and might be dealt with in order to prevent their severe consequences.

As has been mentioned above, because of the abnormal muscle coordination and the neurological impairment, individuals with RTT stiffen in time and the presence of muscle spasms increase the possibility of developing deformities, such as scoliosis. It seems that earlier onset is present in the more severe mutations, such as Arg255X, and the progression is worse in patients who are unable to walk. The presence of deformities (e.g., Figure 2) has been noted since the early studies which reported as a common feature the presence of asymmetry on the right side (Cass et al., 2003; Hagberg, & Romell, 2002). One of the major problems of scoliosis

is that is usually progressive; the prognosis of scoliosis development in RTT is worse, when the scoliosis appears before age five, when severe hypotonia exists from childhood, when there is an inability to walk or when walking was gained but lost at an early age (Hagberg, 1993). Scoliosis and kyphosis often begin with muscle tone problems, which are brought on by the child's inability to correctly process her body scheme (misinterpretation of proprioceptive input by the brain). Active exercise and passive range of movement routines are helpful.



Figure 1. X-ray taken in a suspended position showing scoliosis (Lotan, Merrick, & Carmeli, 2005).



Figure 2. Feet deformities (Lotan, 2006).

Recent studies on RTT mouse models, have allowed to pinpoint genotyping differentiation of specific motor impairments. For example patients with mutation p.R270X or p.R168X present a much more severe phenotype, while those with p.R133C, p.R294X, and C-terminal deletions are more likely to preserve walking abilities (Bebbington et al., 2008; Neul et al., 2008). A recent study that had the aim of validating a gross motor scale, outlined a relationship between the severity of the syndrome and the gross motor abilities, in particular found that milder phenotype (e.g., p.Arg133Cys, p.Arg294 or p.Arg306Cys) showed better performances in abilities such as sitting and walking, whereas mutations such as the p.Arg270 present greater impairment (Downs, et al., 2016).

As outlined above it is important to study more in depth those abilities and the precise genetic factors that can influence them. Consequently, the aim of the present study is to correlate the specific genotype mutation with different parameters of motor impairment. In this study a large sample (195 girls with RTT) with specific motor disability (in walking, use of hands, problems related to the feet and scoliosis) were analyzed in relationship with their specific genotype.

3. METHODS

3.1. Participants

190 girls with a diagnosis of RS, ranging from age 4 to 31 (mean = 18,34 years, SD = 5,36), took part in the experiment. Their families had been contacted by the Italian Rett Association which asked them to participate in the study. All of the participants have been diagnosed with RTT following the guidelines established by the Criteria Work Group and mutation analysis of the methyl-CpG binding protein 2 gene (Table 1).

A general assessment was carried out by a psychologist through the Vineland Adaptive Behavior Scale (VABS) (Sparrow, Balla, & Cicchetti, 1984) and the standardized test Rett Assessment Rating Scale (RARS, Fabio et al., 2005; Vignoli et al., 2010).

3.2. Material

Functional scales. The Vineland Adaptive Behavior Scales is used for diagnosis of intellectual and developmental disabilities. The Scales are organized in four domains: Communication (Receptive, Expressive, Written); Daily Living (Personal, Domestic, Community); Socialization (Interpersonal Relationships, Play and Leisure Time, Coping Skills); and Motor Skills (Gross, Fine).

The Rett Assessment Rating Scale (RARS) is a standardized scale used to evaluate subjects with Rett syndrome (Fabio, Martinazzoli & Antonietti, 2005). It is constructed following the diagnostic criteria for RTT proposed by DSM-IV-TR (APA, 2010) and recent research and clinical experience. It follows a structure similar to that used for the diagnosis of the pervasive developmental disorders included in the same nosographical category as RTT (i.e., Childhood Autism Rating Scale, CARS).

Table 1. Types and frequencies of MECP2 mutations among persons with Rett Syndrome

MECP2 mutation type	N
P152R	7
P251L	1
P302L	1
P322A	1
P376S	1
P403X	1
PK135E	1
R106W	1
R133C	6
R168X	8
R225X	2
R255X	8
R270X	14
R294X	9
R306C	7
R453X	1
T158M	16
W126Y	1
Y141X	2
A257X	1
delexon3	1
K22X	1
119-120dAG	1
387del2	1
431delA	1
CDKL5	1
1152de41	1
1156de33	1
1157de31	1
1157de41	2
1159de44	1
1163de33	1
1163de44	1
1165d69i21	1
1188de32	1
Unknown or not specified	90
Total	195

A total of 31 items was generated as representative of the profile of RTT. Each item concerns a specific phenotypic characteristic and describes four increasing levels of severity. Each item is provided with a brief glossary explaining its meaning in a few words. Each item is rated on a 4-point scale, where 1 = within normal limits, 2 = infrequent or low abnormality, 3 = frequent or medium- high abnormality, and 4 = strong abnormality. Intermediate ratings are possible; for example, an answer between 2 to 3 points is rated as 2.5. For each item, the evaluator circles the number corresponding to the best description of the patient. After a patient

has been rated on all 31 items, a Total score is computed by summing the individual ratings. This Total score allows the evaluator to identify the level of severity of RTT, conceptualized as a continuum ranging from mild symptoms to heavy deficits.

3.2.1. Motor Scale

The Motor subscale of the RARS was obtained by summing up each of the four items. The items considered are walking, use of hands, scoliosis and problems related to feet. Below the precise scale for each item is presented.

X. Walking

1	1.5	2	2.5	3	3.5	4
The girl is able to maintain an erect position and to walk by herself. She knows how to look where she is going, walk on any surface and go up and down stairs.		The girl is able to maintain an erect position but sometimes needs support when walking.		The girl is able to maintain an erect position but always needs support when walking.		The girl is unable to maintain an erect position or walk by herself. She needs a wheelchair or a pram in order to move around.

XI. Hands:

1	1.5	2	2.5	3	3.5	4
Functional use of the hands is only slightly affected; the girl manages to grip and hold objects for long enough to enable her to make ample movements. She has good hand-eye coordination. Stereotypes do not influence the intentional use of the hands.		Functional use of the hands is more compromised; the girl manages to touch, push or hit objects but not to grip them. Hand-eye coordination is scarce. The girl is inhibited in moving by stereotypes.		Functional use of the hands is compromised; the girl is not always able to use her hands intentionally; she does not always manage to touch objects, even if strongly motivated to do so, especially in the presence of marked stereotypes.		Functional use of the hands is almost completely compromised; the girl is not able to use her hands intentionally; she does not manage to touch objects even if strongly motivated to do so, especially because of insistent stereotypes.

XII. Scoliosis:

1	1.5	2	2.5	3	3.5	4
The girl shows no signs of scoliosis.		The girl shows signs of minor scoliosis.		The girl shows signs of major scoliosis.		The girl shows signs of very serious scoliosis.

XIII. Feet:

1	1.5	2	2.5	3	3.5	4
The girl has no problems with her feet.		The girl has slightly small feet, with minor circulation problems.		The girl has small, valgus or different sized feet, with circulation problems.		The girl has problems with her feet, to the extent that she is unable to walk.

3.3. Results

Results are presented in relation to the general motor subscale of the RARS and the specific scale of walking, use of hands, scoliosis and problems related to the feet.

With reference to the first analysis, to calculate the general motor subscale of the RARS, the sum of each type of motor item was analyzed. With reference to the second analysis, the four parameters of the motor scale were used.

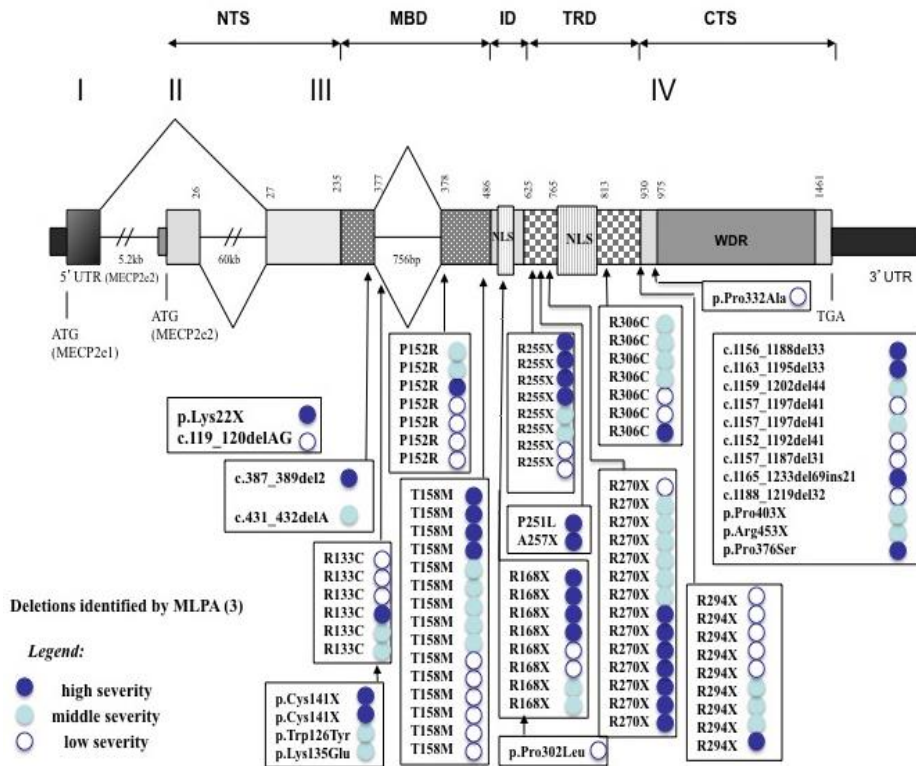


Figure 3. Genomic structure of MECP2 gene and localization of general motor in the coding regions.

To proceed with the genotype-phenotype correlation, the general motor scale score was divided into three categories namely low, medium and high level of severity on the basis of the 33rd and 66th and 100th percentiles for the variable. Based on low, medium and high level of severity, participants were placed within one of the three types of categories. Because some of the girls with RTT show only clinical features and not mutation in MECP2, only 105 patients were included in the analysis (Table 1).

As shown in Figure 3 and 8, patients with a truncating mutation after NLS manifested a lower degree of impairment on general motor scale than patients with a truncating mutation within NLS $\chi^2 = 5.74$, $p < .03$.

With reference to the second analysis, the genotype-phenotype correlation was carried out with the four parameters of the general motor scale: walking, use of hands, scoliosis and problems related to the feet. To calculate these correlations, the walking scale score was divided into three categories namely low, medium and high level of severity on the basis of the 33rd and 66th and 100th percentiles for each variable. Based on low, medium and high level on high levels of severity, participants were placed within one of the three types of categories.

As shown in Figure 4 and 9, patients with a truncating mutation after NLS manifested a lower degree of impairment on walking scores than patients with a truncating mutation within NLS $\chi^2 = 3.11$, $p < .05$.

To further calculate these correlations, the use of hands score was divided into three categories namely low, medium and high level of severity on the basis of the 33rd and 66th and 100th percentiles for each variable. Based on low, medium and high levels of severity, participants were placed within one of the three categories.

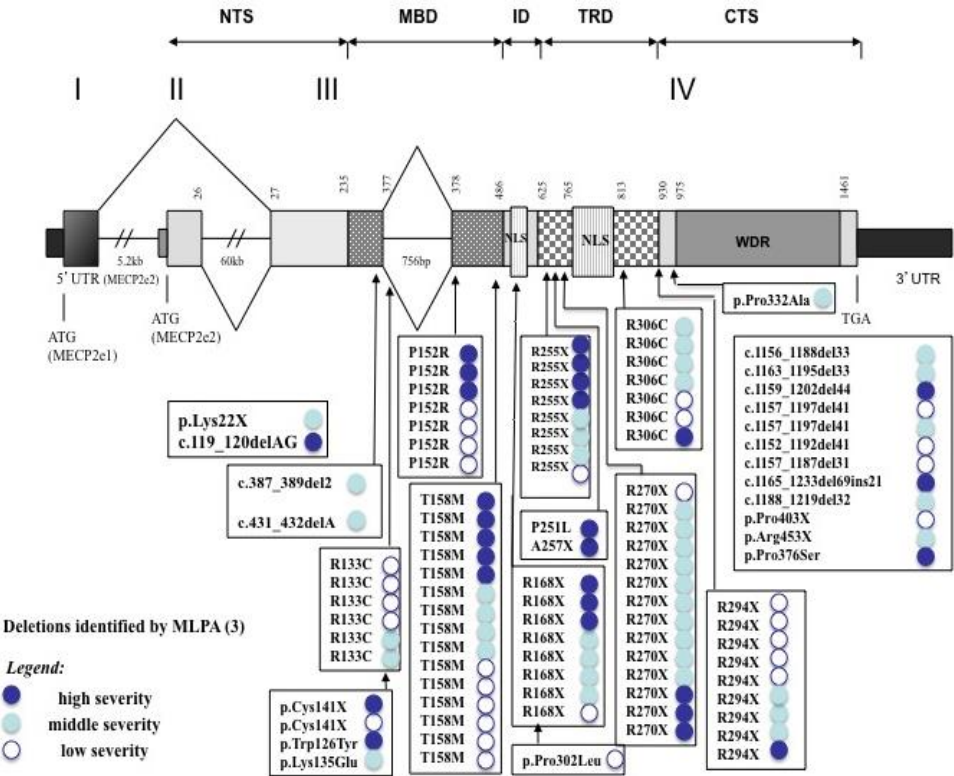


Figure 4. Genomic structure of MECP2 gene and localization of use of walking produced in the coding regions.

As shown Figure 8, patients with a truncating mutation after NLS manifested a lower degree of impairment on general motor scale than patients with a truncating mutation within NLS $\chi^2 = 0.74$, $p < .44$.

Furthermore the scoliosis score was divided into three categories namely low, medium and high level of severity on the basis of the 33rd and 66th and 100th percentiles for each variable. Based on low, medium and high levels of severity, participants were placed within one of the three categories.

As shown in Figure 6, patients with a truncating mutation after NLS manifested the same degree of impairment on scoliosis than patients with a truncating mutation within NLS $\chi^2 = 1.74$, $p < .23$.

Finally the feet problem score was divided into three categories namely low, medium and high level of severity on the basis of the 33rd and 66th and 100th percentiles for each variable. Based on low, medium and high levels of severity, participants were placed within one of the three categories.

As shown in Figure 7 and 10, patients with a truncating mutation after NLS manifested a lower degree of impairment on the feet problem score than patients with a truncating mutation within NLS $\chi^2 = 3.74$, $p < .05$.

Based on the same categories above presented (low, medium and high levels of severity) participants within the most numerous genotypes of our sample were placed in each precise genotype. Figure 11 shows their precise level of impairment. From these data patients with R168, R255 and R270 show the most serious impairments.

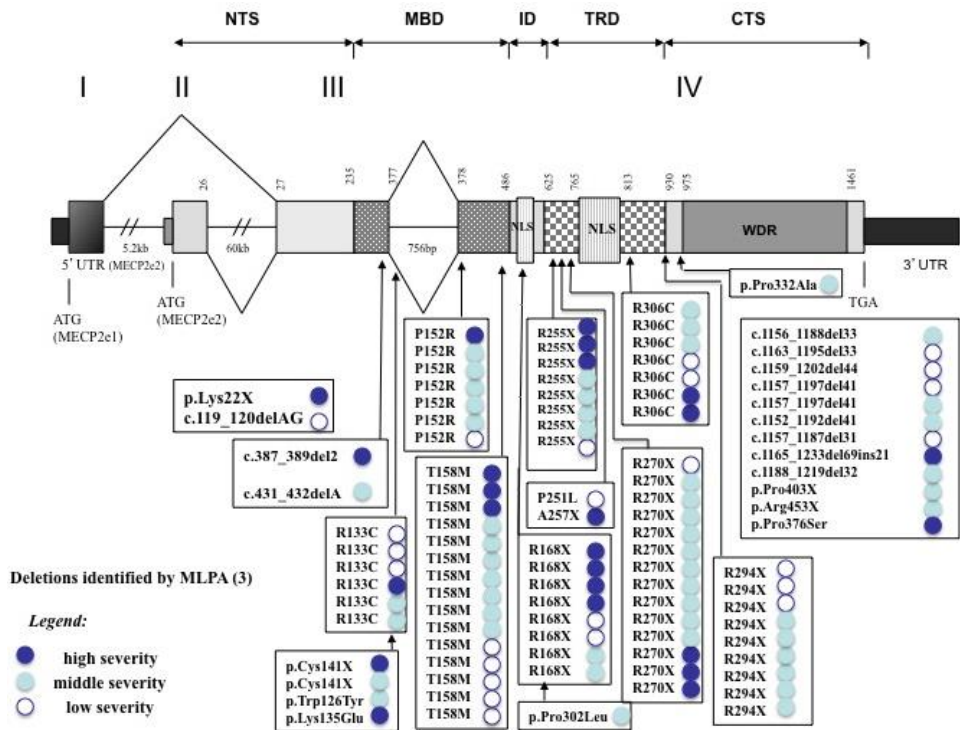


Figure 5. Genomic structure of MECP2 gene and localization of use of hand score produced in the coding regions.

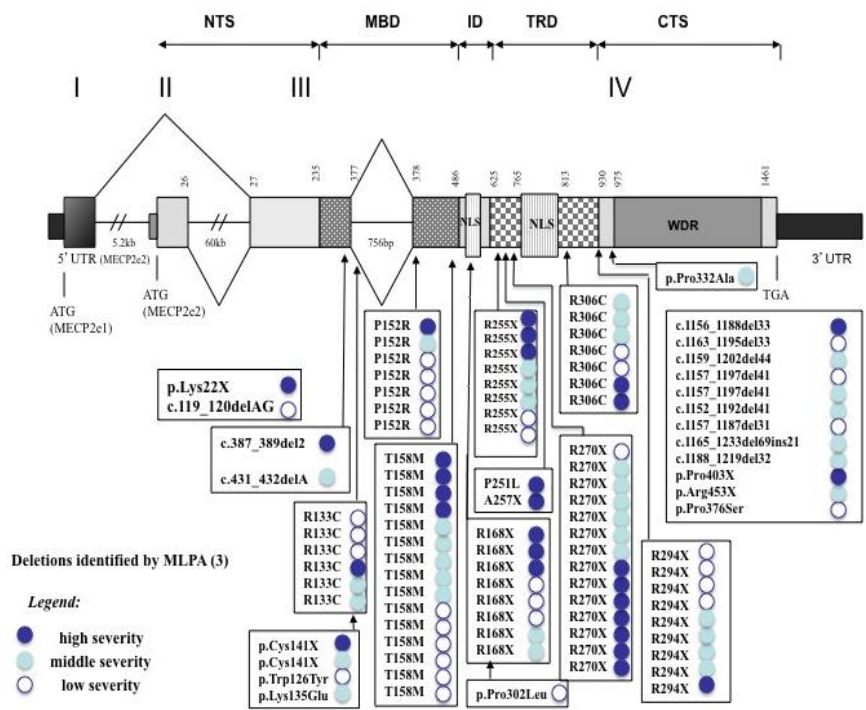
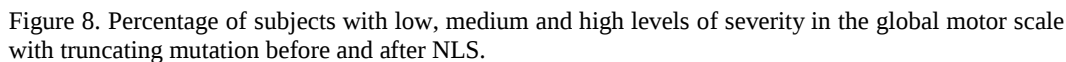
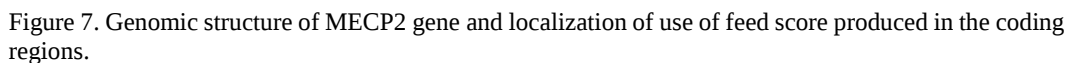


Figure 6. Genomic structure of MECP2 gene and localization of scoliosis score produced in the coding regions.



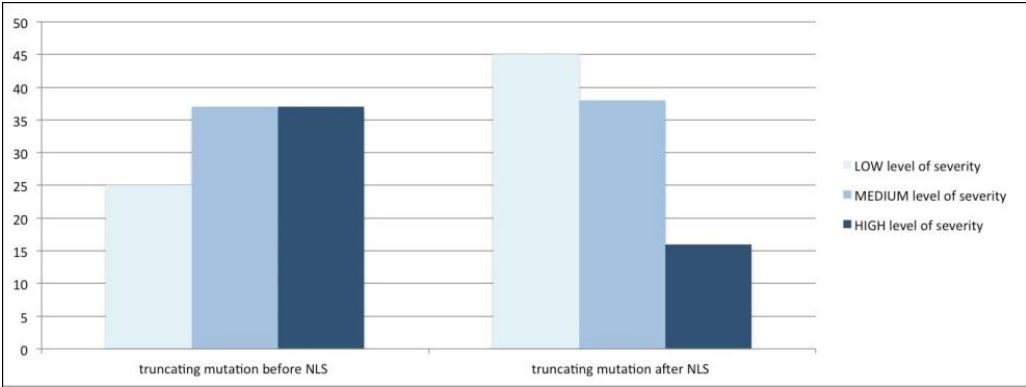


Figure 9. Percentage of subjects with low, medium and high levels of severity in the walking scale with truncating mutation before and after NLS.

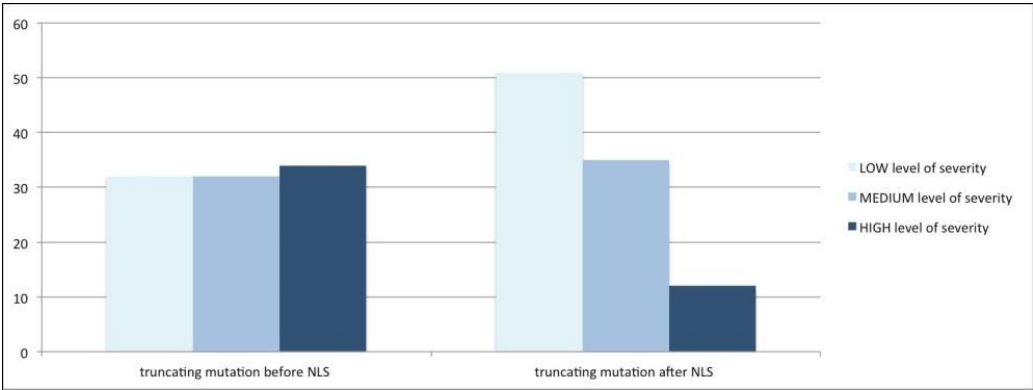


Figure 10. Percentage of subjects with low, medium and high levels of severity in the feet problems scale with truncating mutation before and after NLS.

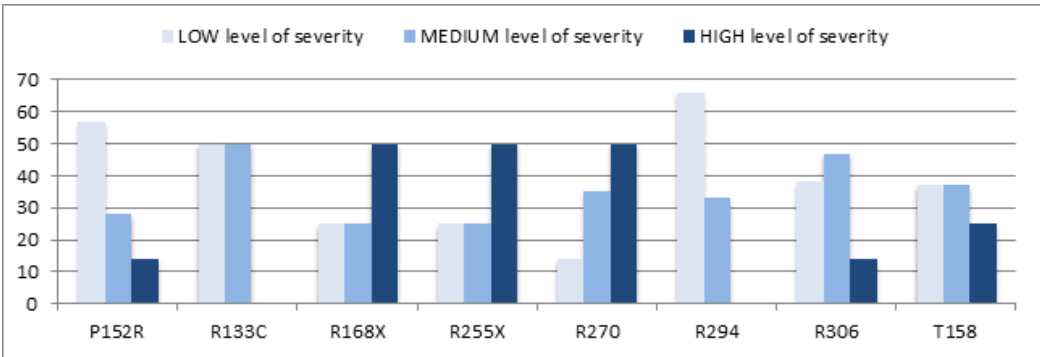


Figure 11. Percentage of subjects with low, medium and high levels of severity in the global motor scale within the most numerous genotypes of our sample.

CONCLUSION

In this study, disease-causing mutations have been related to the intensity of the parameters of the motor impairments measured by the subscale of the RARS (Antonietti et al., 2007) namely the general motor scale, the walking subscale, the use of hands subscale, the scoliosis subscale and the feet problems scale. Comparisons between the truncating mutations differently affecting functional domains support the idea that the crucial factor that leads to different phenotypes is the integrity of NLS.

As in the study of Fabio et al. (2014) and Falzone et al. (2015) we suppose that if the protein can penetrate into the nucleus and link to Methylated CpG, it maintains a residual role causing a milder clinical damage.

The two mutations produced differential phenotypic effects in the five domains here analyzed. Our data confirms the preexisting literature (e.g., Bebbington et al., 2008; Downs, et al., 2016; Neul et al., 2008), in particular genotype-phenotype correlation showed that patients with p.R270X, p.R255X or p.R168X present a much more severe phenotype in terms of specific motor impairment, in particular for walking and problems related to the feet, whilst patients with R306 present a medium level of impairment, confirming earlier classifications (medium-high level of impairment; Fabio et al., 2014). Those with p.R133C and p.R294X show milder motor deterioration. Many studies have associated scoliosis to the more severe mutations (e.g., Arg255X; Leonard, Cobb, & Downs, 2016), while in the present study there were not significant differences between patients with a truncating mutation after NLS or within NLS.

The most important innovation introduced in the study was the use of the correlation between motor scale in relation to the specific genotype.

REFERENCES

- Amir, R. E., Van den Bellver, I. B., Wan, M., Tran, C. Q., Francke, U. & Zoghbi, H. Y. (1999). Rett Syndrome is caused by mutations in X-linked MECP2, encoding methyl-CpG-binding protein 2. *Nature Genetics*, 23(2), 185-8.
- Amir, R. E., Van Den Veyve, I. B., Schultz, R., Malicki, D. M., Tran, C. Q., Dahle, E. J., et al. (2000). Influence of mutation type and X chromosome inactivation on Rett Syndrome phenotypes. *Annals of Neurology*, 47, 670-9.
- Ariani, F., et al. (2008). FOXP1 is responsible for the congenital variant of Rett syndrome. *Am. J. Hum. Genet.*, 83, 89-93.
- Bebbington, A., Anderson, A., Ravine, D., Fyfe, S., Pineda, M., de Klerk, N., Ben-Zeev, B., Yatawara, N., Percy, A., Kaufmann, W. E. & Leonard, H. (2008). Investigating genotype-phenotype relationships in Rett syndrome using an international data set. *Neurology*, 70(11), 868-875.
- Bebbington, A., et al. (2010). Updating the profile of C-terminal MECP2 deletions in Rett syndrome. *J. Med. Genet.*, 47, 242-248.
- Cass, H., Reilly, S., Owen, L., Wisbeach, A., Weekes, L., Slonims, V., Wigram, T. & Charman, T. (2003). Findings from a multidisciplinary clinical case series of females with Rett syndrome. *Developmental Medicine & Child Neurology*, 45(5), 325-337.

- Chao, H. T., et al. (2010). Dysfunction in GABA signaling mediates autism-like stereotypies and Rett syndrome phenotypes. *Nature*, 468, 263–269.
- Cheval, H., et al. (2012). Postnatal inactivation reveals enhanced requirement for MeCP2 at distinct age windows. *Hum. Mol. Genet.*, 21, 3806–3814.
- Christine, K. B., Lee, J. S., Lim, B. C., Kim, K. J., Hwang, Y. S. & Chae, J. H. (2015). FOXG1 Mutation is a Low-Incidence Genetic Cause in Atypical Rett Syndrome. *Child Neurology Open*, 1-5.
- Christodoulou, J., Grimm, A., Maher, T. & Bennetts, B. (2003). RettBASE: The IRSA MECP2 variation database—a new mutation database in evolution. *Hum Mutat*, 21(5), 466-472.
- De Bona, C., et al. (2000). Preserved speech variant is allelic of classic Rett syndrome. *Eur. J. Hum. Genet.*, 8, 325–330
- Einspieler, C., Kerr, A. M. & Prechtl, H. F. (2005). Abnormal general movements in girls with Rett disorder: the first four months of life. *Brain and Development*, 27(1), S8-S13.
- Evans, J. C., Archer, H. L., Colley, J. P., Ravn, K., Nielsen, J. B., Kerr, A., et al. (2005). Early onset seizures and Rett-like features associated with mutations in CDKL5. *European Journal of Human Genetics*, 13, 1113-1120.
- Fabio, R. A., Antonietti, A., Marchetti, A. & Castelli, I. (2009a). Attention and communication in Rett syndrome. *Research in Autism Spectrum Disorders*, 3, 329–335.
- Fabio, R. A., Giannatiempo, S., Antonietti, A. & Budden, S. (2009b). The role of stereotypies in overselectivity processes in Rett Syndrome. *Research in Developmental Disabilities*, 30, 136–145.
- Fabio, R. A., Giannatiempo, S., Oliva, P. & Murdaca, A. M. (2011). The increase of attention in Rett Syndrome. A pretest-post test research design. *Journal of Developmental and Physical Disability*, 23, 99-111
- Fabio, R. A., Colombo, B., Russo, S., Cogliati, S. F., Masciandri, M., Antonietti, A. & Tavian, D. (2014). Recent Insights into Genotype-Phenotype Relationships in Patients with Rett Syndrome Using a Fine Grain Scale, *Research in Developmental Disabilities*, 35(11), 2976–2986.
- Fabio, R. A., Billeci, L., Crifaci, G., Troise, E., Tortorella, G. & Pioggia, G. (2016). Cognitive training modifies frequency EEG bands and neuropsychological measures in Rett syndrome, *Research in Developmental Disabilities*, 54, 73-85.
- Falzone, A., Gangemi, A. & Fabio, R. A. (2015). Correlations Between Linguistic Phenotype and Genetic Alterations in Rett Syndrome. In: *Advances in Genetic Research*, vol. 14, chapter id 32719. New York: Nova Science Publisher.
- Fehr, S., Bebbington, A., Nassar, N., Downs, J., Ronen, G. M., De Klerk, N. & Leonard, H. (2011). Trends in the diagnosis of Rett syndrome in Australia. *Pediatric Research*, 70, 313-319.
- Fehr, S., et al. (2011). Altered attainment of developmental milestones influences the age of diagnosis of Rett syndrome. *J. Child Neurol.*, 26, 980–987.
- Fehr, S., Wilson, M., Downs, J., Williams, S., Murgia, A., Sartori, S., Vecchi, M., Ho, G., Polli, R., Psoni, S., Bao, X., De Klerk, N., Leonard, H. & Christodoulou, J. (2013). The CDKL5 disorder is an independent clinical entity associated with early-onset encephalopathy. *European Journal of Human Genetics*, 21, 266-273.
- Guo, J. U., Su, Y., Shin, J. H., Shin, J., Li, H., Xie, B., Zhong, C., Hu, S., Le, T., Fan, G., et al. (2014). Distribution, recognition and regulation of non-CpG methylation in the adult mammalian brain. *Nature neuroscience*, 17, 215-222.

- Gabel, H. W., Kinde, B., Stroud, H., Gilbert, C. S., Harmin, D. A., Kastan, N. R., Hemberg, M., Ebert, D. H. & Greenberg, M. E. (2015). Disruption of DNA-methylation-dependent long gene repression in Rett syndrome. *Nature*, 522, 89-93.
- Hagberg, B. (1993). Clinical Criteria, Stages and Natural History. In B. Hagberg (Ed.), *RettSyndrome-Clinical & Biological Aspects*, (pp. 4-20). London: Mac Keith Press.
- Hagberg, B. (1999). *Rett syndrome: Clinical and biological aspects*. London: Mac Keith Press.
- Hagberg, B. (2002). Clinical manifestations and stages of Rett syndrome. *Mental retardation and developmental disabilities research reviews*, 8(2), 61-65.
- Hagberg, B. (2005). Rett syndrome: long-term clinical follow-up experiences over four decades. *Journal of Child Neurology*, 20, 722-727.
- Hagberg, B., Aicardi, J., Dias, K. & Ramos, O. (1983). A progressive syndrome of autism, dementia, ataxia, and loss of purposeful hand use in girls: Rett's syndrome: report of 35 cases. *Ann Neurol*, 14, 471-479.
- Hagberg, B. & Witt-Engerstrom, I. (1986). Rett syndrome: a suggested staging system for describing impairment profile with increasing age towards adolescence. *American journal of medical genetics. Supplement*, 1, 47-59.
- Hagberg, B. & Romell, M. (2002). Rett females: patterns of characteristic side-asymmetric neuroimpairments at long-term follow-up. *Neuropediatrics*, 33(6), 324-326.
- Jackowski, A. P., Laureano, M. R., Del'Aquila, M. A., de Moura, L. M., Assunção, I., Silva, I. & Schwartzman, J. S. (2011). Update on Clinical Features and Brain Abnormalities in Neurogenetics Syndromes. *Journal of Applied Research in Intellectual Disabilities*, 24(3), 217-236.
- Julu, P., Witt Engerström, I., Hansen, S., Apartopoulos, F., Engerström, B., Pini, G., Delamont, R. S. & Smeets, E. (2008). Cardiorespiratory challenges in Rett syndrome. *The Lancet*, 371, 1981-1983.
- Kerr, A. M. (1992). *The future for Rett syndrome girls*. International Rett Syndrome Association, Winter, 13-4.
- Kerr, A. M. & Julu, P. O. (1999). Recent insights into hyperventilation from the study of Rett Syndrome. *Archives of Disease in Childhood*, 80, 384-7.
- Kokura, K., Kaul, S. C., Wadhwa, R., Nomura, T., Khan, M. M., Shinagawa, T., Yasukawa, T., Colmenares, C. & Ishii, S. (2001). The Ski protein family is required for MeCP2-mediated transcriptional repression. *The Journal of biological chemistry*, 276, 34115-34121.
- Kortum, F., Das, S., Flindt, M., et al. (2011). The core FOXP1 syndrome phenotype consists of postnatal microcephaly, severe mental retardation, absent language, dyskinesia, and corpus callosum hypogenesis. *J Med Genet.*, 48(6), 396-406.
- Larsson, G. (2013). *Rett syndrome, motor development, mobility and orthostatic reactions. Loss of function, difficulties and possibilities*. Licentiate thesis. Umeå University: Umeå.
- Larsson, G. & Engerstrom, I. W. (2011). Gross motor ability in Rett Syndrome-the power of expectation, motivation and planning. *Brain and Development*, 23(Suppl 1), S77-81.
- Leonard, H., Cobb, S. & Downs, J. (2017). Clinical and biological progress over 50 years in Rett syndrome. *Rare Disease*, 13, 37-50.
- Leonard, H., Bower, C. & English, D. (1997). The prevalence and incidence of Rett syndrome in Australia. *European Child and Adolescent Psychiatry*, 6, 8-10.
- Li, Y., Wang, H., Muffat, J., Cheng, A. W., Orlando, D. A., Loven, J., Kwok, S. M., Feldman, D. A., Bateup, H. S., Gao, Q., et al. (2013). Global transcriptional and translational

- repression in human embryonic-stem-cell-derived Rett syndrome neurons. *Cell Stem Cell*, 13, 446-458.
- Lotan, M. (2006). *Management for Rett syndrome*. Tel Aviv: Israel Rett Syndrome Center.
- Lotan, M. (2006). Rett syndrome. Guidelines for individual intervention. *Scientific World Journal*, 6, 1504-1516.
- Lyst, M. J. & Bird, A. (2015). Rett syndrome: a complex disorder with simple roots. Nature reviews. *Genetics*, in press.
- Motil, K. J., Schultz, R. J., Wong, W. W. & Glaze, D. G. (1998). Increased energy expenditure associated with repetitive involuntary movement does not contribute to growth failure in girls with Rett Syndrome. *Journal of Pediatrics*, 132(2), 228-33.
- Nan, X., Ng, H. H., Johnson, C. A., Laherty, C. D., Turner, B. M., Eisenman, R. N. & Bird, A. (1998). Transcriptional repression by the methyl-CpG-binding protein MeCP2 involves a histone deacetylase complex. *Nature*, 393, 386-389.
- Neul, J. L., Kaufmann, W. E., Glaze, D. G., Christodoulou, J., Clarke, A. J., Bahi Buisson, N., Leonard, H., Bailey, M. E., Schanen, N. C., Zappella, M., et al. (2010). Rett syndrome: revised diagnostic criteria and nomenclature. *Ann Neurol*, 68, 944-950.
- Neul, J. L., Fang, P., Barrish, J., Lane, J., Caeg, E. B., Smith, E. O., Zoghbi, H., Percy, A. & Glaze, D. G. (2008). Specific mutations in methyl-CpG-binding protein 2 confer different severity in Rett syndrome. *Neurology*, 70(16), 1313-21.
- Nomura, Y. & Segawa, M. (1990). Clinical features of the early stage of the Rett syndrome. *Brain and Development*, 12(1), 16-19.
- Philippe, C., Amsellem, D., Francannet, C., et al. (2010). Phenotypic variability in Rett syndrome associated with FOXP1 mutations in females. *J Med Genet.*, 47(1), 59-65.
- Pini, G., Bigoni, S., Engerström, I. W., Calabrese, O., Felloni, B., Scusa, M. F., et al. (2012). Variant of Rett Syndrome and CDKL5 Gene: Clinical and Autonomic Description of 10 Cases. *Neuropediatrics*, 43, 37-43.
- Rajaei, S., Erlandson, A., Kyllerman, M., Albage, M., Lundstrom, I., Karrstedt, E. L., et al. (2011). Early Infantile Onset "Congenital" Rett Syndrome Variants: Swedish Experience Through Four Decades and Mutation Analysis. *J Child Neurol*, 26, 65-71.
- Ross, P. D., Guy, J., Selfridge, J., Kamal, B., Bahey, N., Tanner, E., Gillingwater, T., Jones, R. A., Loughrey, C. M., McCarroll, C. S., Bailey, M. E. S., Bird, A. & Cobb, S. (2016). Exclusive expression of MeCP2 in the nervous system distinguishes between brain and peripheral Rett syndrome-like phenotypes. *Human Molecular Genetics*, 4-46.
- Roze, E., Cochen, V., Sangla, S., Bienvenu, T., Roubergue, A., Leu-Se-menescu, S. & Vidaihet, M. (2007). Rett syndrome: An overlooked diagnosis in women with stereotypic handmovements, psychomotor retardation, Parkinsonism, and dystonia? *Movement Disorders*, 22, 387-389.
- Santos, M., Summavielle, T., Teixeira-Castro, A., Silva-Fernandes, A., Duarte- Silva, S., Marques, F., Martins, L., Dierssen, M., Oliveira, P., Sousa, N. & Maciel, P. (2010). Monoamine deficits in the brain of methyl-CpG binding protein 2 null mice suggest the involvement of the cerebral cortex in early stages of Rett syndrome. *Neuroscience*, 170(2), 453-467.
- Trevathan, E. (1989). Rett syndrome. *Pediatrics*, 83, 636-637.
- Urbanowicz, A., Downs, J., Girdler, S., Ciccone, N. & Leonard, H. (2015). Aspects of speech-language abilities are influenced by MECP2 mutation type in girls with Rett syndrome. *Am. J. Med. Genet*, 16, 354-362.

- Vignoli, A., Fabio, R. A., La Briola, F., Giannatiempo, S., Antonietti, A, Maggiolini, S. & Canevini, M. P. (2010). Correlations between neurophysiological, behavioral, and cognitive function in Rett syndrome. *Epilepsy & Behavior*, *17*, 489–496.
- Zappella, M. (1992). The Rett girls with preserved speech. *Brain and Development*, *14*, 98-101.

Chapter 76

**A LONGITUDINAL STUDY ON
THE DEVELOPMENT OF A BOY WITH FRAGILE X
SYNDROME: DEVELOPMENTAL TRENDS AND
ADAPTIVE CHANGES IN A SINGLE CASE REPORT**

***María Auxiliadora Robles-Bello^{1,2*}, PhD,
Nieves Valencia-Naranjo², PhD
and David Sánchez-Teruel³, PhD***

¹Asociación Síndrome de Down de Jaén y provincia

²Department of Psychology, University of Jaén, Jaén, Spain

³Department of Psychology, University of Córdoba, Córdoba, Spain

ABSTRACT

Fragile X syndrome (FXS) is the most common inherited cause of intellectual disability. However, findings reported in cross-sectional studies on this population are heterogeneous. This chapter focuses on the longitudinal assessment of a boy with FXS from 12 months through to 6 years of age using two tests: one that assesses psychomotor development and another that assesses neuropsychological maturity. The child had attended an Early Childhood Development Intervention Center-ECDIC since 12 months old. He was administered the *Brunet-Lézine Revised Scale of Psychomotor Development in Early Childhood* until 36 months and underwent CUMANIN neuropsychological testing from age 3 to 6. The results obtained allow us to observe trends in the boy's psychomotor development and neuropsychological maturity over time. Significant commonalities between these results and those of previous cross-sectional studies are discussed. Furthermore, some conclusions are drawn that may prove valuable to professionals and researchers interested in this syndrome.

* Corresponding Author's Email: marobles@ujaen.es.

1. INTRODUCTION

Fragile X syndrome (FXS) is the leading known inherited cause of intellectual disability in the world and the second most common chromosomal abnormality after Down syndrome, affecting approximately 1 per 2,633 males and 1 per 4,000 females (Kidd et al., 2014; Fernández-Carvajal et al., 2009). Given that this is a disorder linked to the X chromosome, it mostly affects males, while females are carriers (Bailey, Hatton, Tassone, Skinner & Taylor, 2001; Wilding, Cornish & Munir, 2002). Fragile X syndrome is an X-linked monogenic disease caused by a mutation in the FMR-1 gene located at Xq27.3 (Bagni, Tassone, Neri & Hagerman, 2012).

The behavioral phenotype presents itself in a myriad of ways among this population (Verkerk et al., 1991). However, it has been observed that FXS involves cognitive functioning that suggests a common profile in this group of people. According to Brun-Gasca (2006), we can find general characteristics that do not appear in every affected person, and which may vary depending on sex and mutation type. Specifically, clear sex-dependent differences are found (Murphy & Abbeduto, 2007). Cognitive deficits are therefore less evident in girls than in boys; between 30 and 50% of females with a full mutation have intellectual disability ranging from mild to moderate, whereas only a small number have severe disability. One-third of females display normal intelligence and, in some cases, learning disorders (especially in mathematics) as well as social and relationship problems. Combined with less noticeable effects in terms of physical features, all this makes an early diagnosis of this syndrome in girls difficult. Regarding the mutation type in boys with permutation, the phenotype is very similar to that of boys with full mutation (Cornish et al., 2008).

Given these sex-related differences, the data discussed herein refer to boys with FXS. Regarding linguistic abilities, the majority of studies have focused on language expression deficits, while studies examining language reception have yielded conflicting results (Abbeduto et al., 2003; Abbeduto, Brady & Kover, 2007). Fürgang (2001) found vocabulary level to be the least affected by auditory processing problems and a delay in syntactic development. However, Brun-Gasca and Artigas-Pallarés (2001) hold the view that those with FXS grasp syntax with relative ease. Children with FXS also present language deficits when it comes to organizing and sequencing information, exhibiting perseveration, echolalia, and poor communication skills. Therefore, they struggle to know when it's their turn to speak and maintain eye contact with the speaker; they also find it difficult to keep up with the topic of conversation. What is more, the use of tangential language frequently occurs. Typical behavior in individuals with FXS, namely avoiding eye contact, usually surfaces when speech development begins and not at earlier stages (Van der Molen et al., 2010). FXS also affects oral-motor functions (e.g., low muscle tone, dyspraxia, intelligibility problems) as well as voice problems (Fürgang, 2001).

Intellectual functioning heterogeneity is highly characteristic of this syndrome, which speaks to the importance of conducting a proper individual evaluation of all areas involved (Abbeduto et al., 2007; Van der Molen et al., 2010). In fact, some features frequently reported in males with FXS include ongoing deficits in short-term memory (Munir, Cornish & Wilding, 2000a); visual spatial memory (Munir et al., 2000a); executive function (Hooper et al., 2008; Munir, Cornish & Wilding, 2000b; Scerif, Cornish, Wilding, Driver & Karmiloff-Smith, 2007); and difficulties in processing sequential and abstract information (Van der Molen et al., 2010).

Males with FXS are usually shy and socially anxious; they avoid eye contact or do so selectively, more often addressing people they know rather than people they don't know (Lewis et al., 2006). They do not avoid social situations at an early age, quite the opposite in fact: they usually seek them out and expose themselves to these scenarios (Robles-Bello & Sánchez-Teruel, 2013a).

As can be observed, there is a significant body of scientific literature on FXS. However, the majority of studies have applied cross-sectional methodologies, meaning that outcomes based on longitudinal data are limited; these could provide a more epigenetic understanding of this disorder. From this perspective, the current study proposes two specific aims: first, to obtain a full description of a boy with FXS which serves to increase knowledge about the syndrome's underlying features focusing on individual data covering a six-year period; and second, to acknowledge the importance of individualized early childhood intervention, particularly given the heterogeneity within this disorder.

2. METHOD

2.1. Subject

X is a boy born in January 2008 who was referred to an Early Childhood Development Intervention Center-ECDIC) (*Centro de Desarrollo Infantil y Atención Temprana-CDIAT*) (Robles-Bello & Sánchez-Teruel, 2013b) with a pediatric diagnosis of psychomotor delay and communication disorder with hyperactivity, plus a potential diagnosis of FXS taking into account phenotypic and neuropsychological traits. This diagnosis was later confirmed, further indicating that X was a carrier of a full mutation with an expansion of 250 CGG triple repeats (PCR to amplify the region where CGG expansion of the FMR-1 gene occurs—at the FRAXA locus—and capillary electrophoresis).

2.2. Measures

Testing instruments varied over the course of the evaluation process on the basis of the child's chronological age. The measures used are described below.

Brunet-Lézine Revised Scale of Psychomotor Development in Early Childhood (Josse, 1997; n.d.). This scale applies to toddlers and infants aged 0 to 30 months and aims to measure the child's maturity level in the four areas under examination: postural-motor control; hand-eye coordination (cognition-perception); language/communication; and sociability/autonomy. It allows us to obtain a development age and an overall development quotient for the child, as well as a partial assessment of the child's development age and development quotient, for each of the aforementioned areas. The measure has adequate reliability (stability coefficients—test/retest at 15-day intervals in infant samples at ages 6, 12 and 18 months—of 0.70 and higher; there is also item homogeneity, reporting Cronbach's alpha of between 0.69 and 0.87). Internal validity of the different scales in relation to the general scale ranged from 0.49 to 0.67 (Ramos-Martín, Sancho-García, Cachero-Sanz, Vara-Arias & Iturria-Matamala, 2009).

The *Child Neuropsychological Maturity Questionnaire-CUMANIN* (Portellano, Mateos & Martínez, 2000) is an individually administered test that assesses neuropsychological maturity in children aged between 3 and 6 years (36 months to 78 months). The questionnaire takes between 30 and 50 minutes to complete. The CUMANIN comprises main scales (psychomotricity, language articulation, language expression, language comprehension, spatial structuring, visual perception, iconic memory and rhythm) and subscales (attention, verbal fluency, reading, writing and laterality). The Cronbach's alpha reliability level varied from scale to scale, reporting 0.71 (psychomotricity), 0.92 (language articulation), 0.73 (language expression), 0.72 (language comprehension), 0.81 (spatial structuring), 0.91 (visual perception), 0.57 (iconic memory) and 0.72 (rhythm).

2.3. Procedure

The evaluation process took place at the start and end of each age period (1, 2, 3, 4, 5 and 6 years). The results obtained allow us to observe trends in the child's development over time. He was administered the *Brunet-Lézine Revised Scale of Psychomotor Development in Early Childhood* (Josse, 1997; n.d.) until 36 months and took the CUMANIN neuropsychological test from age 3 through 6. Throughout the entire evaluation process, X regularly attended daycare (in the first three years) and later school. Furthermore, X was enrolled in an early intervention program for the first three years of his life under Public Health Law (Consejería de Salud, 2017).

The principal aim of "Early Childhood Intervention" is to promote a child's well-being and family development by enabling boys and girls with developmental disorders—or those at risk for exhibiting them—to achieve fuller integration within family, school and social life (Candel, 2005). This early child intervention from birth up until age 6 focuses, among other things, on enhancing cognitive performance, autonomy, language and communication, and motor skills at an Early Childhood Development Intervention Center-ECDIC. In some European countries, Spain included, the process begins at birth and is carried through until 6 years of age (until 3 in some autonomous regions of Spain). It is also free to users and co-funded by the national public healthcare system (Robles-Bello & Sánchez-Teruel, 2013).

The direct work sessions with X lasted 45 minutes, with the boy attending the ECDIC three times a week. The interdisciplinary team made up of physiotherapists, speech therapists and psychologists, in collaboration with the parents, adjusted the program to meet the needs of the child and his family. These work objectives were fulfilled using the program developed by Zulueta, Mollá, Martínez, Lago de Lanzos and Arrieta (2004). The family's session attendance rate was 100%.

3. RESULTS

The results shown in Table 1 below correspond to the evaluation period spanning 3 to 6 years using the CUMANIN test, reflecting the raw and centile scores for each evaluated area.

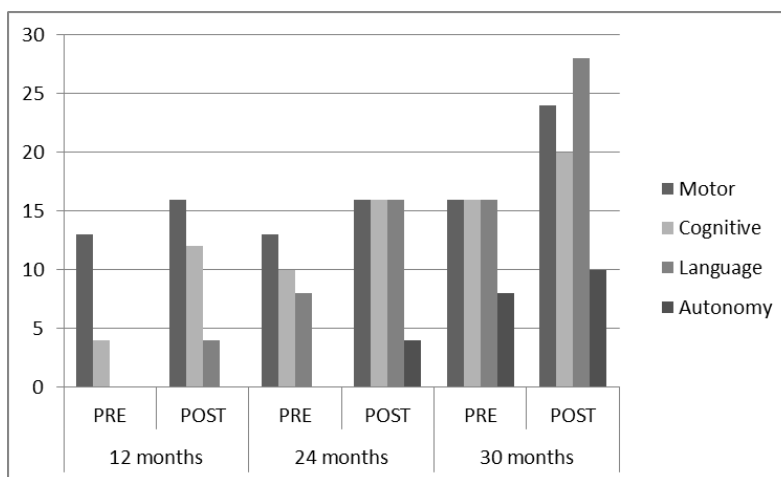


Figure 1. Development level attained on the Brunet-Lézine scale before and after each treatment year up until 30 months.

Table 1. Raw and centile scores on the Child Neuropsychological Maturity Questionnaire (CUMANIN)

	Age	3 years		4 years		5 years		6 years	
Subscale	Scores	PRE	POST	PRE	POST	PRE	POST	PRE	POST
Language Articulation	RS	6	7	7	8	7	8	11	12
	CS	55	60	35	40	20	25	20	25
Language Expression	RS	0	0	0	1	2	2	3	3
	CS	25	5	5	5	20	10	10	10
Language Comprehension	RS	2	3	4	6	6	8	8	9
	CS	70	60	50	75	75	95	95	99
Verbal Fluency	RS	0	0	0	1	2	2	4	6
	CS	15	5	1	3	3	2	2	5
Verbal Development	RS	8	10	11	15	15	18	22	24
	CS	40	30	25	30	20	15	50	70
Psychomotoricity	RS	3	4	5	4	6	7	10	10
	CS	25	20	15	5	15	20	80	80
Visual Perception	RS	0	2	2	2	7	10	10	12
	CS	45	45	15	10	30	40	40	65
Iconic Memory	RS	1	2	2	2	7	7	9	9
	CS	40	15	20	20	75	60	95	95
Rhythm	RS	0	0	0	1	2	2	2	3
	CS	30	25	15	25	30	20	20	35
Spatial Structuring	RS	2	3	4	6	8	9	9	10
	CS	20	10	10	30	35	40	40	60
Nonverbal Development	RS	6	11	13	15	29	34	40	44
	CS	15	10	15	25	25	25	60	85
Attention	RS	2	4	4	6	8	12	12	16
	CS	90	60	55	55	40	30	30	35
Development Quotient	T	75	75	75	75	107	110	107	110

RS = Raw score; CS = Centile score; T = Total

4. DISCUSSION

This study focused on the developmental period spanning 12 months to 6 years of a boy diagnosed with fragile X syndrome. Overall, the development of X shows a trend to improvement during the first three years of his life; however, this does not continue in the same way from age 3 onwards. Functional impairment is more pronounced in the last three years. His neuropsychological profile is characterized by peaks and dips in learning, with language being the area where most progress is made and autonomy the most representative in the first three years. Although this child's trajectory cannot be generalized, his follow-up data do not coincide with findings reported in cross-sectional studies (Roberts, Hatton & Bailey, 2001).

Following the first year of intervention, family and daycare subjectively perceive a positive trend in X's acquisitions. This assessment holds true in subsequent regular contact sessions. The raw scores obtained using the psychomotor development test (Josse, 1997) confirm this trend across the different assessed areas (motor, cognitive, language and autonomy).

Within the context of this positive progression profile, we sought to identify whether acquisition increase between years 0 and 3 maintained a similar rate of growth. The gains in raw scores for each period suggest positive growth in motor and language acquisition skills. Regarding language, similar gains are observed in scores across each period, whereas age 3 is seen as a stage of significant gains in motor acquisition. These results are not consistent with those obtained by Roberts et al., (2001), who identified language as the area where least progress is made, possibly due to the working memory problems discussed in some previous studies (Baker et al., 2011). However, growth rates in cognition and autonomy slow according to subsequent evaluations.

The following set of results correspond to the developmental period spanning 3 to 6 years, where learning is assessed using the CUMANIN test which allows for a more in-depth assessment of the different areas of development. Rice, Warren and Betz (2005) argue that language skills in children with FXS present particular challenges when it comes to the expression component, consistent with the development of nonverbal cognitive abilities. Furthermore, they report voice and speech problems that make it difficult to understand their expressions. The results obtained for X partially concur with this conclusion. When comparing the ability to repeat previously heard expressions (language expression) and speaking skills (language articulation) with nonverbal maturation (nonverbal development), his performance is comparable to what is expected between the ages of 3 and 5 and under 6. The comparison with verbal development points to a similar trend, although expression impairment is observed from age 3 onwards.

These outcomes regarding his language skills vary in terms of comprehension. In this case, the comparison made with the child's level of nonverbal and verbal maturity suggests that his ability to extract information from a previously heard story is one of the boy's strengths. His performance level across the different evaluated periods (3, 4, 5 and 6 years) is similar to that observed in a population matched on chronological age. Estigarribia et al., (2011) also found that recall of a previously heard story did not differ significantly from that of the norm group after having controlled for nonverbal development level and short-term memory. The discrepancies found between language dimensions (comprehension, expression) when comparing performance, either with verbal or nonverbal maturity level, supports the conclusion drawn by Price et al., (2008). According to these authors, the link

between cognitive and linguistic abilities (syntax expression and comprehension) in FXS cases—and possibly in other conditions such as Down syndrome—is more flexible than that observed in cases involving norm groups.

Another point of interest in this discussion is the examining of X's progress based on raw scores (Van der Molen et al., 2010). This trend in abilities is always positive across all evaluation periods, although they differ at the growth curve observed. The repeat of previously heard expressions emerges when the child approaches the end of age 4, increasing very slowly yet progressively. Growth in this ability (based on standard deviations in relation to the norm group mean) remains relatively constant. A different pattern is observed for speech production (language articulation) which develops very slowly from age 3 onwards, achieving greater gains between age 5 and 6. Despite this, the discrepancy in performance among children of a similar chronological age increases gradually with advancing age. Taking into account the raw scores, this positive trend is also observed in the child's ability to understand the elements included in a previously heard story. His capacity to comprehend is present from the age of 3, generally maintaining a positive developmental trajectory until 6 years.

Overall nonverbal development (nonverbal development) is progressive but very slow and remains at the lower functioning range between the ages of 3 and 5. Evaluations at age 6 report significant performance improvements, which are especially influenced by psychomotricity development (age 6) and ability gains in iconic memory (age 5). However, auditory memory ability, measured through rhythm, maintains a low profile across the different evaluation stages.

The visual perception task involves copying geometric shapes and is defined by Van der Molen (2010) as a performance measure using abstract items. For this researcher, performance in FXS cases is poor (e.g., lower than expected for nonverbal development level) on tasks that contain abstract items as opposed to tasks that contain concrete items, that is, visual perception skills and better processing of meaningful information. X's visual perception performance does not completely follow the pattern suggested by Van der Molen (2010). The discrepancy between task functioning and the child's level of nonverbal maturity is noticeable at age 6; however, this performance—from approaching the end of age 3 up until 4 and 5 years—assumes values close to what might be expected, albeit inconsistently. In terms of the trend in raw scores, X's ability to copy shapes arises at age 3 and holds stable until age 5. His performance improves at this age, yielding a significant 5-point gain. He makes favorable progress in year 6, achieving scores close to those obtained by the norm group mean. Despite this, the growth rate at age 6 is slower than at age 5.

Spatial organization ability (spatial structuring) is found to be in line with the level of nonverbal maturity across all evaluated periods; it is even slightly above expectations at age 5. The trend in raw scores across all age groups is continual and positive, yielding an improvement in centile position, especially from the end of age 4 to the end of age 6, a period in which performance level is similar to that found in the norm group. This good performance may be linked to the spatial structuring task features covered in the CUMANIN assessment test, where a large proportion of the task's items require commands to be carried out, taking into account the axis of corporal symmetry. Kogan et al., (2009) identified egocentric spatial organization as a strength that children with FXS exhibit when compared with performance on tasks where the reference axis for spatial organization is located outside the body. A salient aspect in relation to this task is a profile similar to the one observed in the visual perception task. Kogan

et al., (2009) suggested that visual perceptual ability in children with this disorder may find themselves guided more by the spatial organization of objects than by the features of these stimuli. Hence, capacity development for spatial structuring that occurs at the end of year 4 may have facilitated visual perception task performance.

Psychomotricity is measured using an implicit procedural learning task. Bussy, Charrin, Brun, Curie and Des Portes (2011) examined this type of learning in children with FXS. These researchers define procedural learning as the acquisition and storing of rules associated with motor skills. The children identified these features slowly and unconsciously, drawing on past experience. The acquisition of implicit procedural learning in children without difficulties starts early and reaches good performance levels at age 7 (Thomas & Nelson, 2001, taken from Bussy et al., 2011). The findings reported by Bussy et al., (2011) suggest that implicit procedural learning functions efficiently in FXS. In the case of X, his performance coincides with the pattern established by Bussy et al., (2011): it is efficient across all assessed age groups, falling within expectations in nonverbal maturity and even exceeding them somewhat at age 6. However, when compared with the norm group in terms of percentile position, X finds himself in the first quartile up until age 6, an age at which he performs very well and reports efficiency levels similar to those observed in the norm group. When taking into account raw scores, performance begins at age 3 and maintains positive yet slow growth up until age 6, which is when the highest growth in these scores is observed (a 3-point gain).

The two tasks covered in the evaluation of nonverbal development level are related to memory capacity, specifically (1) rhythm, an auditory memory task; and (2) iconic memory, a visual memory task featuring concrete and meaningful material. Rhythm-related performance between the ages of 3 and 5 reveals a low profile aligned with nonverbal maturity level. The discrepancy between both scores is seen at age 6, a time when X's nonverbal development improves substantially while his ability to repeat rhythmic sequences develops at a much slower rate. The difficulties reported in this auditory task contrast with performance on the visual memory (iconic memory) task. In the latter, the overall profile is one of a development level comparable to his nonverbal development between the ages of 3 and 4 and higher between years 5 and 6. Progress in terms of raw scores suggests that rhythm performance is delayed until the end of year 4 and improves at later stages albeit very slowly (the gain stands at approximately 1 point per year).

The recall of meaningful visual material begins from the end of age 3 and holds stable until age 5, the age at which substantial improvements in the number of recalled items are observed. Progress is maintained at age 6, although less growth is seen (2-point gain). When comparing X's performance data with that of a similar age norm group, a significant deficit in both abilities between the ages of 3 and 4 is observed. However, visual memory performance between the ages of 5 and 6 improves substantially, with values approaching those obtained by the norm group; meanwhile, the significant gap in auditory task performance is maintained. Hence, the difference between both types of memory coincides with the greater difficulties in auditory memory as identified by Baker, Hooper, Skinner, Hatton, Schaaf, Ornstein and Bailey (2011), although not in the persistent trend of these results during development.

Attentional control refers to the ability to choose specific stimuli to learn about, remain attentive for a long period of time, regulate and supervise (monitor) behavior and actions, and control impulses. Attentional control is one of the domains within the executive control system. The study conducted by Cornish, Cole, Longhi, Karmiloff-Smith and Scerif (2012) showed that FXS children's accuracy in an attentional control test was lower, albeit comparable, to that

observed in children (3–10 years) of a similar mental age. This good performance coincides with the pattern observed in X up until age 5. During this period, his performance somewhat exceeds the expected level, taking into account his nonverbal development when compared with the norm group mean. However, from this age onwards, a distancing in relation to this norm group emerges. In their review of attentional control capacity in control preschoolers, Anderson and Reidy (2012) suggested that accuracy in attentional control tasks improves significantly between ages 4.5 and 5. This trend in the control group may explain the differences observed in X's performance, who continues to show progress in his ability albeit more slowly. Thus, when examining the raw scores, X's performance reveals a 2-point gain at ages 3 and 4, with a growth curve increasing trend at 5 and 6 years of age (4-point gain per year).

Another test that relates to executive functioning is that of verbal fluency (Anderson & Reidy, 2012). In this task, the child has to make a meaningful sentence based on a single word. Hooper et al., (2008) identified an overall deficit when examining executive function in males with FXS, reporting major difficulties across different domains of executive function (e.g., inhibition, working memory, set-shifting and planning), with the exception of processing speed. The data corresponding to X's cognitive flexibility in verbal fluency are very poor, irrespective of the comparison criteria (norm group, verbal maturity and nonverbal maturity). In terms of raw scores, floor-level performance is seen at the end of year 4 and rises slightly at age 5 (1-point gain). The greatest development is observed at age 6 with a 4-point gain.

An important aspect to acknowledge is the performance difference between the verbal fluency and attention tasks. Both tasks are associated with the concept of executive function. However, a significant decline is observed in verbal fluency which is not identified in attention until age 6. A possible reason for this could be the use of a visual pathway in the attention task, in line with the results from Baker et al.,'s (2011) study.

One of the core objectives of any intervention aimed at improving children's cognitive abilities is to align the treatment programs with their own features, designing the tasks around their strengths and compensating for their particular difficulties; on many occasions, this also involves introducing different types of support. In this regard, it is of interest to identify the skills and gaps in two main processing areas: auditory/verbal and visual. The scientific literature that documents FXS features gives, in some instances, contradictory information about verbal and visual processing abilities in this disorder. These gaps may be associated with the difficulty in capturing the wide range of factors that exert an influence on both these principal operating areas; the study of a population undergoing a learning/acquisition process; the different reference methods used and even individual variability, to name a few of the general factors.

The progress made by children with FXS in terms of ability acquisition varies from study to study. In their literature review, Hahn, Brady, Warren and Fleming (2015) established three possible mean trajectories (overall) of adaptive behavior functioning (e.g., communication, socialization, daily routine and motor skills) characterized by (a) a decline over time; (b) progressive development up until approximately age 10 and subsequent stabilization or decline; and (c) a positive path from infancy to age 12. Their findings suggest that the development of children with FXS is not unique. Approximately half of the children develop well over the entire evaluation period (between 30 and 120 months), while the remaining half (56%) lose adaptive skills in relation to their chronological peers as well as in absolute values, that is, when taking into account the raw scores. In cases where a decline in adaptive functioning was observed, this occurred approaching approximately 7 years of age.

Although adaptive behavior has not been examined directly in this study, it can be assumed that this behavior is influenced, among other things, by children's skill sets. From this perspective, X's overall development profile would likely correspond to a progressive form of development. Given that progress monitoring stopped before the child reached 7, the age that Hahn et al., (2015) established as a potential critical point in children's development, conclusions cannot be drawn about future progress; this especially holds true when we see that no significant differences were observed in different variables among children progressing well and those experiencing ability decline. What is more, in cases where a decline occurred, Hahn et al., (2015) observed that this could be attributed to certain domains of adaptive functioning. In the case of X, areas of particular concern could be those that we have previously described as performance levels showing less favorable trends and persistent difficulties. These may be target areas for intervention and/or for developing compensatory mechanisms.

To summarize, as to be expected, X exhibited idiosyncratic functioning: it cannot be entirely extrapolated to functioning profiles extracted from groups. Despite this, convergences with common features in FXS are seen. These include language and cognitive difficulties in general terms and, more specifically, language expression difficulties, short-term memory problems, deficits in visual spatial processing and psychomotricity, as well as difficulties in tasks related to executive functioning. However, it is also possible to identify some of these more individual elements. In the case of X, there are two main ages in which his performance differs notably from this general deficit pattern. The first period is age 3, which is when X demonstrates good performance in speech production and language comprehension; attentional control capacity, even surpassing the mean norm group; and efficient visual perception levels. The early intervention program which X and his family had actively participated in almost since his birth is considered one of the most relevant factors.

The other significant stage is the period spanning ages 5 through 6. During this period, X makes very good progress in language comprehension, outstripping learning growth recorded at earlier stages. He also performs very well in psychomotricity (especially at age 6) and in iconic memory (at age 5). However, progress is less pronounced in other abilities such as visual perception and spatial structuring. Alongside persistent difficulties in language expression, verbal fluency and rhythm, we even observe a certain loss of ability when compared with the norm group and/or his nonverbal maturity level, as in the case of attentional control. Despite all this, it should be pointed out that an examination of X's raw scores reveals positive progress, although not at the rate identified in the comparison group which shows no difficulties.

REFERENCES

- Abbeduto, L., Brady, N. & Kover, S. T. (2007). Language development and Fragile X syndrome: profiles, syndrome-specificity, and within-syndrome differences. *Mental Retardation and Developmental Disabilities*, 13: 36–46. doi: 10.1002/mrdd.20142.
- Abbeduto, L., Murphy, M. M., Cawthon, S. W., Richmond, E. K., Weissman, M. D., Karadottir, S. & O'Brien, A. (2003). Receptive language skills of adolescents and young adults with Down or Fragile X syndrome. *American Journal on Mental Retardation*, 108: 149–160. doi: 10.1352/0895-8017(2003)108%3C0149:RLSOAA%3E2.0.CO;2.

- Anderson, P. J. & Reidy, N. (2012). Assessing executive function in preschoolers. *Neuropsychological Review*, 22: 345–360. doi: 10.1007/s11065-012-9220-3.
- Bagni, C., Tassone, F., Neri, G. & Hagerman, R. (2012). Fragile X syndrome: Causes, diagnosis, mechanisms, and therapeutics. *Journal of Clinical Investigation*, 122 (XII); 314–22. doi: 10.1172/JCI63141
- Bailey, D. B., Hatton, D. D., Tassone, F., Skinner, M. & Taylor, A. K. (2001). Variability in FMRP and early development in males with fragile x syndrome. *American Journal on Mental Retardation*, 106: 16–27. doi: 10.1352/08958017(2001)106<0016:VIFAED>2.0.CO;2.
- Baker, S., Hooper, S., Skinner, M., Hatton, D., Schaaf, J., Ornstein, P. & Bailey, D. (2011). Working memory subsystems and task complexity in young boys with Fragile X syndrome. *Journal of Intellectual Disability Research*, 55 (I): 19–29. doi: 10.1111/j.1365-2788.2010.01343.x.
- Brun-Gasca, C. (2006). El fenotipo cognitivo-conductual. In M. I. Tejada (dir.), *Síndrome X Frágil. Libro de consulta para familias y profesionales [Fragile X syndrome. Books for families and professionals]* (pp. 31-36). Madrid: Real Patronato sobre Discapacidad. Retrieved from: <http://www.fraxa.org/pdfs/LibroSXF.PDF>.
- Brun-Gasca, C. & Artigas-Pallarés, J. (2001). Aspectos psicolingüísticos en el síndrome del cromosoma X frágil [*Psycholinguistic aspects of fragile X chromosome syndrome*]. *Revista de Neurología*, 33 (supl.1), 29-32.
- Bussy, G.1., Charrin, E., Brun, A., Curie, A. & des Portes, V. (2011). Implicit procedural learning in fragile X and Down syndrome. *Journal of Intellectual Disability Research*, 55(5), 521-8. doi: 10.1111/j.1365-2788.2011.01410.x.
- Candel, I. (2005). Elaboración de un programa de atención temprana [Elaboration of a program of early intervention]. *Electronic Journal of Research in Educational Psychology*, 3(7), 151-192. Retrieved from: http://www.investigacionpsicopedagogica.org/revista/articulos/7/espanol/Art_7_98.pdf.
- Consejería de Salud, (2017). Atención Temprana. Sevilla. Junta de Andalucía. [*Early attention. Seville. Junta de Andalucía*]. Retrieved from: <http://www.juntadeandalucia.es/temas/salud/infantil/temprana.html>.
- Cornish, K. M., Li, L., Kogan, C. S., Jacquemont, S., Turk, J., Dalton, A. Hagerman, R. J. & Hagerman, P. J. (2008). Age-dependent cognitive changes in carriers of the fragile Xsyndrome. *Cortex*, 44(6), 628-636. doi:10.1016/j.cortex.2006.11.002.
- Cornish, K., Cole, V., Longhi, E., Karmiloff-Smith, A. & Scerif, G. (2012). Does attention constrain developmental trajectories in fragile x syndrome? A 3-year prospective longitudinal study. *American Journal on Intellectual and Developmental Disabilities*, 117(2), 103-20. doi: 10.1352/1944-7558-117.2.103.
- Estigarribia, B., Martin, G. E., Roberts, J. E., Spencer, A., Gucwa, A. & Sideris, J. (2011). Narrative Skill in Boys with Fragile X Syndrome with and without Autism Spectrum Disorder. *Applied Psycholinguistics*, 32(2), 359-388. doi: 10.1017/S0142716410000445.
- Fernández-Carvajal, I., Walichiewicz, P., Xiaosen, X., Pan, R., Hagerman, P. J. & Tassone, F. (2009). Screening for expanded alleles of the FMR1 gene in blood spots from newborn males in a Spanish population. *Journal of Molecular Diagnostics*, 11(4), 324-329. doi:10.2353/jmoldx.2009.080173.

- Fürgang, R. (2001). La terapia del Lenguaje en el Síndrome X Frágil [Language therapy in fragile X syndrome]. *Revista de Neurología*, 33, 82-87. Retrieved from: <http://www.slideshare.net/davidpastorcalles/la-terapia-dellenguajeenelsndromefragilx>.
- Hahn, L. J., Brady, N. C., Warren, S. F. & Fleming, K. K. (2015). Do Children with Fragile X Syndrome Show Declines or Plateaus in Adaptive Behavior? *American Journal on Intellectual and Developmental Disabilities*, 120(5), 412-32. doi: 10.1352/1944-7558-120.5.412.
- Hooper, S. R., Hatton, D., Sideris, J., Sullivan, K., Hammer, J., Schaaf, J., Bailey, D. B. Jr. (2008). Executive functions in young males with fragile X syndrome in comparison to mental age-matched controls: baseline findings from a longitudinal study. *Neuropsychology*, 22, 36-47. doi: 10.1037/0894-4105.22.1.36.
- Josse, D. (1997). *Brunet-Lézine Révisé: Échelle de développement psychomoteur de la petite enfance*. Paris: Établissements d'Applications Psychotechniques. [Brunet-Lézine Revised: Psychomotor Development Scale of Early Childhood. Paris: Psychotechnical Applications Establishments].
- Josse, D. (n./d.). *Escala de desarrollo psicomotor de la primera infancia Brunet-Lézine-Revisado [Brunet-Lézine-Revised scale of psychomotor development in early childhood]*. Madrid: Psymtéc.
- Kidd, S. A., Lachiewicz, A., Barbouth, D., Blitz, R. K., Delahunty, C., McBrien, D., Berry-Kravis, E. (2014). Fragile X Syndrome: A Review of Associated Medical Problems. *Pediatrics*, 134 (5), 995-1005 doi: 10.1542 / peds.2013-4301.
- Kogan, C. S., Boutet, I., Cornish, K., Graham, G. E., Berry-Kravis, E., Drouin, A. & Milgram, N. W. (2009). A comparative neuropsychological test battery differentiates cognitive signatures of Fragile X and Down syndrome. *Journal of Intellectual Disability Research*, 53(2), 125-42. doi: 10.1111/j.1365-2788.2008.01135.x.
- Lewis, P., Abbeduto, L., Murphy, M., Richmond, E., Giles, N., Bruno, L. & Schroeder, S. (2006). Cognitive, language and social-cognitive skills of individuals with Fragile X syndrome with and without autism. *Journal of Intellectual Disability Research*, 50, 532-545. doi: 10.1111/j.1365-2788.2006.00803.x.
- Munir, F., Cornish, K. M. & Wilding, J. (2000a). Nature of the working memory deficit in Fragile X syndrome. *Brain and Cognition*, 44, 387-401. doi:10.1006/brcg.1999.1200.
- Munir, F., Cornish, K. M. & Wilding, J. (2000b). A neuropsychological profile of attention deficits in young males with Fragile X syndrome. *Neuropsychologia*, 38, 1261-1270. doi: 10.1016/S0028-3932(00)00036-1.
- Murphy, M. M. & Abbeduto, L. (2007). Gender differences in receptive language in Fragile X syndrome. *Journal of Intellectual Disability Research*, 45, 387-400. doi: 10.1111/j.1365-2788.2006.00888.x.
- Portellano, J. A., Mateos, R. & Martínez, R. (2000). *Cuestionario de Madurez Neuropsicológica Escolar [School Neuropsychological Maturity Questionnaire]*. Madrid: Tea Ediciones.
- Price, J. R., Roberts, J. E., Hennon, E. A., Berni, M. C., Anderson, K. L. & Sideris, J. (2008). Syntactic complexity during conversation of boys with fragile X syndrome and Down syndrome. *Journal of Speech, Language, and Hearing Research*, 51(1), 3-15. doi: 10.1044/1092-4388(2008/001).
- Ramos-Martín, J. M., Sancho-García, M. J., Cachero-Sanz, P., Vara-Arias, M. T., & Iturria-Matamala, B. (2009). El temperamento infantil en el ámbito de la prevención primaria.

- Relación con el cociente de desarrollo y su modificabilidad [Child Temperament and Development Quotient in Early Prevention]. *Clínica y Salud*, 20(1), 67-78.
- Rice, M., Warren, S. & Betz, S. (2005). Language symptoms of developmental language disorders: An overview of autism, Down syndrome, fragile X, specific language impairment, and Williams síndrome. *Applied Psycholinguistics* 26, 7–27. doi: 10.1017/S0142716405050034.
- Roberts, J. E., Hatton, D. D. & Bailey, D. B. (2001). Development and behaviour of male toddlers with Fragile X syndrome. *Journal of Early Intervention*, 24(3), 207-223. doi:10.1177/10538151010240030601.
- Robles-Bello, M. A. & Sánchez-Teruel, D. (2013a). Tratamiento del síndrome del cromosoma X frágil desde la atención infantil temprana en España [Fragile X Syndrome Treatment from early child attention in Spain]. *Clínica y Salud*, 24(1), 19-26. doi: 10.5093/cl2013a3.
- Robles-Bello, M. A. & Sánchez-Teruel, D. (2013b). Early Childhood Intervention in Spain. *Papeles del Psicólogo*, 34(1), 3-13.
- Scerif, G., Cornish, K., Wilding, J., Driver, J. & Karmiloff-Smith, A. (2007). Delineation of early attentional control difficulties in Fragile X syndrome: Focus on neurocomputational changes. *Neuropsychologia*, 45, 1889–1898. doi: 10.1016/j.neuropsychologia.2006.12.005.
- Thomas, K. M. & Nelson, C. A. (2001). Serial reaction time learning in preschool- and school-age children. *Journal of Experimental Child Psychology*, 79(4), 364-87. doi: 10.1006/jecp.2000.2613.
- Van der Molen, T. (2010). Co-morbidities of COPD in primary care: frequency, relation to COPD, and treatment consequences. *Primary Care Respiratory Journal*, 19(4), 326-34. doi: 10.4104/pcrj.2010.00053.
- Van der Molen, M. J. W., Huizinga, M., Huizinga, H. M., Ridderinkhof, K. R., Van der Molen, M. W., Hamel, B. J. C., Ramakers, G. J. A. (2010). Profiling Fragile X Syndrome in males: strengths and weakness in cognitive abilities. *Research in Developmental Disabilities*, 31, 426-439. doi: 10.1016/j.ridd.2009.10.013.
- Verkerk, A. J. M., Pieretti, M., Sutcliffe, J. S., Fu, Y. H., Kuhl, D. P. A., Pizzuti, A., Warren, S. T. (1991). Identification of a gene (FMR-1) containing a CGG repeat coincident with a breakpoint cluster region exhibiting length variation in fragile X syndrome. *Cell*, 65, 905-914. doi: 10.1016/j.ridd.2009.10.013.
- Wilding, J., Cornish, K. & Munir, F. (2002). Further delineation of the executive deficit in males with fragile X syndrome. *Neuropsychologia*, 40, 1343–1349. doi:10.1016/S0028-3932(01)00212-3.
- Zulueta, M. I., Mollá, M. T., Martínez, A. M., Lago de Lanzos, C. & Arrieta, F. (2004). *Programa para la estimulación del desarrollo infantil [Program for the stimulation of child development]*. Madrid: Editorial Cepe.

Chapter 77

CORNELIA DE LANGE SYNDROME: A REVIEW

Elizabeth A. M. Frost*, MD

Icahn Medical Center at Mount Sinai,
New York, NY, US

ABSTRACT

Cornelia de Lange syndrome (CdLS), also known as Bushy syndrome, Amsterdam dwarfism and Brachmann- de Lange syndrome is a genetic multi system disorder, usually caused by spontaneous mutation. Although present from birth, it may not always be immediately diagnosed. The estimated incidence is about 1:10,000-30,000 births. Both sexes are affected. Mortality is high early in life and neurosensory, craniofacial, musculoskeletal, cardiac and gastrointestinal abnormalities are all apparent. There is no known cure and treatment is supportive, requiring a team support system.

Keywords: genetic mutation, rare disorders, respiratory failure, dwarfism, Cornelia de Lange syndrome

INTRODUCTION

Cornelia de Lange syndrome (CdLS), is a complex syndrome of multiple congenital anomalies that vary in severity. It is genetically heterogeneous and sporadic, with an estimated prevalence of 1 in 10,000 to 30,000. Winfried Robert Clemens Brachmann (1888–1969), a German physician documented the first case when he described the features and autopsy results of a 19 day old patient who had died of pneumonia in 1916 [1]. He described, mainly, characteristics of the upper limbs and wrote on the facial symptoms less specifically. Some years later, in 1933, Cornelia Catharina de Lange, a Dutch pediatrician after whom the disorder has been named documented the cases of two girls with unusual facies and mental retardation,

* Corresponding Author's Email: Elzfrost@aol.com (Clinical Professor of Anesthesia).

one 17 months and the other 6 months, admitted within weeks of each other to Emma Children's Hospital [2]. The first child had pneumonia and during her first year had suffered from many feeding difficulties. She was very small for her age and microcephalic. The second child had a similar medical history. De Lange noted that the resemblance to each other was remarkable. She termed the condition "un type nouveau de dégénération (typus Amstelodamensis)." After she presented a further case to the Amsterdam Neurological Society, the disorder gained recognition by 1941 [3].

In a review in 1985, John Marius Opitz commented: "Brachmann's paper is a classic of Western Medical iconography, deserving to be commemorated in the eponym "Brachmann- de Lange syndrome." The conjoined description is now generally accepted, although the term "de Lange" or Cornelia de Lange's syndrome is also common [3].

To date there is no known cure although the syndrome can be managed by treating associated clinical symptoms. Sixty six percent of CdLS individuals die before the first year of life [4]. However, for those who survive infancy, life expectancy is relatively normal, reaching to the 6th decade. Nevertheless, certain features of this condition, particularly severe malformations of the heart or throat, decrease life expectancy dramatically. Mortality occurs primarily from aspiration in infancy and from infection and bowel obstruction in later years [4].

Overall, CdLS is probably under-diagnosed and frequently misdiagnosed and subsequently mismanaged.

CLINICAL FEATURES

As occurs in many other genetically flawed syndromes, strong resemblance is seen between people with CdLS. Distinctive facial features include thin confluent eyebrows (synophrys), long eyelashes, a short upturned nose, and thin downturned lips [5-7]. Prenatal and postnatal growth deficiency, psychomotor delay, and upper limb malformations are common. Birth weight is typically below 5lbs. Small stature and macrocephaly are also common. Although almost all organ systems can be affected, individuals with CdLS most notably display deficits in the development of neurosensory, craniofacial, musculoskeletal, cardiac, gastrointestinal and genitourinary systems. Medical concerns are also common and include gastro- esophageal reflux disease, heart defects, seizures, feeding difficulties, vision problems, and hearing loss. But CdLS is not a "one size fits all" condition. An individual may have many of the following traits, or only a select few. Geneticists and clinicians can establish the diagnosis after evaluating all the criteria. The Human Phenotype Ontology (HPO) provides a list of features that have been reported in people with CdeLS. Much of the information in the HPO comes from Orphanet, a European rare disease database. The list includes a rough estimate of the frequency of a feature. Frequencies are based on a specific study and may not be representative of all studies (Table 1).

Table 1. The approximate rate of occurrence of features found in CdeLS.
Adapted from (<https://rarediseases.info.nih.gov/diseases/10109/cornelia-de-lange-syndrome>) November 2017

Signs and Symptoms	Approximate number of patients (when available)
Abnormally low-pitched voice	Very frequent (present in 80%-99% of cases)
Anteverted nares	Very frequent (present in 80%-99% of cases)
Atresia of the external auditory canal	Very frequent (present in 80%-99% of cases)
Brachycephaly	Very frequent (present in 80%-99% of cases)
Curly eyelashes	Very frequent (present in 80%-99% of cases)
Poor dentition, delayed eruption	Very frequent (present in 80%-99% of cases)
Skeletal maturation slow	Very frequent (present in 80%-99% of cases)
Downturned corners of lips and mouth	Very frequent (present in 80%-99% of cases)
Gastroesophageal reflux	Very frequent (present in 80%-99% of cases)
Generalized hirsutism	Very frequent (present in 80%-99% of cases)

More organ specific diagnostic criteria are as follows.

Neuropsychiatric and Neurosensory

Hearing loss, including both sensorineural and conductive loss occur in up to 60% [6, 7]. Ear canals are narrowed or stenotic, making otitis media and sinusitis common complications [8, 9, 10]. Peripapillary pigmentation, high myopia, ptosis, microcornea, and blepharitis are also ophthalmologic problems [11]. Nasolacrimal duct obstruction, nystagmus, cataract and glaucoma are rarer findings.

Growth failure is apparent in over 95% beginning in utero and most obvious by six months of age with mean height and weight remain below the 5th percentile [6, 7]. Developmental delay, often severe, follows a similar path [12]. Overall IQ values range from below 30 to 102, with an average of 53 [13]. Patients with mild forms of CdLS have higher functioning with IQ that may be normal to borderline but are usually diagnosed with learning disabilities [12, 13, 14]. Disabilities in speech and language are common often leading to behavioral issues that may be secondary to frustration from an inability to communicate. Behavior is seen as consistent with depression and attention deficit hyperactivity disorder, and display obsessive-compulsive behavior, autistic behavior, including self- destructive tendencies, defiance, extreme shyness and avoidance of social interactions.

Craniofacial

As noted above, facial features are the most distinctive clinical feature. The characteristic facial phenotype is recognizable with the Facial Dysmorphology Novel Analysis (FDNA) technology [15]. The head is small, with a low hairline on the forehead and posterior neck. Eyebrows are well defined and confluent, and are highly arched in 98% of patients [5]. Eyelashes are long and thick with an exaggerated upward curve of the upper eyelashes and an exaggerated downward curve of the bottom eyelashes. Hypertelorism combined with an antimongoloid slant of the eyes is common. The midface is flattened and associated with a short nose and ante-verted nares. The nasal bridge is broad and/or depressed. The philtrum is long, smooth, and prominent. The lips are characteristically thin with downturned corners. Micrognathia and a high arched palate and cleft palate occur in 30% [5]. Dental anomalies are common and include widely spaced, small, absent or displaced teeth. Ears are low set, posteriorly rotated, and often hirsute.

Musculoskeletal

Specific extremity findings are a further aid to diagnosis. Although lower extremity findings are less common, over 80% of affected individuals have partial syndactyly of toes 2 and 3 [5]. Hands and feet are small in over 90%, and single palmar creases can be seen in 50% along with fifth finger clinodactyly in 75%, and brachydactyly [5]. The first metacarpal is often shortened with a proximally placed thumb [6]. Upper extremity malformations are seen in up to 30% of patients, and range from oligodactyly to ulnar deficiency to absent forearm. Misplaced diaphyses may develop distal to the elbow [6]. Radial head dislocation with abnormal elbow extension, clubbed feet and poikilothermia are also common. Pectus excavatum, scoliosis and hip dislocation or dysplasia have also been described. Missing arms, forearms, and/or fingers are found in 25% of patients with CdLS.

Cardiovascular and Pulmonary Findings

Congenital heart disease may be diagnosed in as many as 20 – 30%, compared to 0.8% for all births [5]. The most common abnormalities include (in descending order): ventricular septal defects, atrial septal defects, pulmonic stenosis, tetralogy of Fallot, hypoplastic left heart syndrome, and tricuspid aortic valve [16]. While some heart anomalies have obvious signs and symptoms at birth that prompt evaluation by a pediatric cardiologist other defects are more subtle and are not always recognized immediately, delaying detection. Children suspected of having CdLS should be screened by echocardiography.

Pulmonary hypoplasia and lobular anomalies predispose CdLS patients to respiratory infections. The most common causes of death in such patients are acute pneumonia and bronchitis. Pulmonary hypertension may result from repeated episodes of upper airway obstruction due to micrognathia and macroglossia exacerbating hypoxic and hypercapnic episodes

Gastrointestinal

The most common gastrointestinal complication is gastroesophageal reflux disease (GERD) occurring in over 90%⁶. Esophagitis, aspiration, chemical pneumonitis, and irritability are complications of GERD that can be avoided by diagnosis and treatment in the neonatal period. Pyloric stenosis with vomiting may contribute to malnutrition and poor weight gain during the neonatal period. Malrotation of the gut is less common but is still seen in at least 10% and may result in the life threatening situation of volvulus [6]. Congenital diaphragmatic hernia may also occur but the reported incidence may be underestimated as these infants often die in the perinatal period.

Genitourinary

Structural kidney and/or urinary tract anomalies, the most common being vesiculoureteral reflux, pelvic dilation and renal dysplasia associated with renal dysfunction, are found in up to 40% of individuals with CdLS [17]. Genitalia malformations include cryptorchidism occurring in up to 73% of males, with hypoplastic and micropenis in 57% and hypospadias [6]. Females may have small labia majora and abnormally formed uteri.

PATHOGENESIS AND CAUSES

Most cases of CdLS are due to spontaneous genetic mutations. Such mutations affecting the cohesin complex and multiple genes have been associated with CdLS. Researchers at the Children's Hospital of Philadelphia and the University of Newcastle upon Tyne identified a gene (NIPBL) on chromosome 5 that causes CdLS when it is mutated [18, 19]. In July 2012, a fourth "CdLS gene"—HDAC8—was announced. HDAC8 is an X-linked gene, meaning it is located on the X chromosome. Individuals with CdLS who have the gene change in HDAC8 make up only a small portion of all people with CdLS [20].

These genes, (NIPBL, SMC1A, SMC3), are all involved in sister chromatid cohesion. Cohesion proteins are required for chromosome segregation, regulation of gene expression, DNA repair and maintenance of genome stability. Mutations in NIPBL on chromosome 5 account for ~ 50% of CdLS cases, while mutations in SMC1A on the inactivated X chromosome, and SMC3 on chromosome 10 account for ~5% [21]. NIPBL and SMC3 mutations are both believed to have an autosomal dominant inheritance. SMC1A mutations may have an X-linked dominant pattern of inheritance, however males and females are affected similarly [22]. The genotype-phenotype correlation reveals that mutations in NIPBL result in more severe phenotypes than mutations in SMC1A and SMC3 genes that may often result in less severe expression [23]. Associated phenotype in NIPBL mutations increases in severity as the severity of the mutation increases. More severe mutations occur in NIPBL deletions or truncations, while milder forms of CdLS occur in patients with NIPBL missense mutations. (SMC1A and SMC3 gene mutations are predominantly missense and small in-frame deletions). The phenotype in patients with SMC1A and SMC3 mutations is milder, resulting mainly in mild to moderate mental retardation, with little growth retardation and limb or systemic

involvement [23]. However, a recent study indicated that CDL caused by a SMC 1A variant may be a responsible for neglected syndromic craniosynostosis [24]. Although mutations in these genes are implicated in 65% of patients, the pathogenesis of most cases of CdLS is sporadic and dominant [21].

Nevertheless this complex multisystem developmental disorder is caused by mutations in cohesin subunits and regulators, the precise molecular mechanisms are not well defined. However, identification points to a global deregulation of the transcriptional gene expression program. Cohesin is associated with the boundaries of chromosome domains and with enhancer and promoter regions connecting the three-dimensional genome organization with transcriptional regulation. A recent study that connected gene communities, and structures emerging from the interactions of non-coding regulatory elements and genes in the three-dimensional chromosomal space, provided a molecular explanation for the pathophysiology of CdLS associated with mutations in the cohesin-loading factor *NIPBL* and the cohesin subunit *SMC1A* *NIPBL* [25]. Genes deregulated in CdLS are positioned within reach of *NIPBL*- and cohesin-occupied regions through promoter-promoter interactions. The researchers offer a dynamic model where *NIPBL* loads cohesin to connect genes in communities, offering an explanation for the gene expression deregulation in the CdLS.

Other genes are probably also responsible when they mutate making it possible to understand why CdLS varies so widely from one individual to another and what can be done to improve the quality of life for people with the syndrome as they age [26].

Table 2. The latter two genes seem to correlate with a milder form of the syndrome (adapted from Wikipedia, accessed November 2017)

Name	OMIM	Gene	Appx. %	Notes
CDLS1	122470	NIPBL	50%	A gene responsible for CdLS on chromosome 5 was discovered in 2004 jointly by researchers at the Children’s Hospital of Philadelphia, and researchers at Newcastle University, UK.[7]
CDLS2	300590	SMC1A	5%	In 2006, a second gene, on the X chromosome, was found by Italian scientists.
CDLS3	610759	SMC3	1%	A third gene discovery was announced in 2007. The gene is on chromosome 10 and was also discovered by the research team in Philadelphia.

DIAGNOSIS

The diagnosis of CdLS is primarily a clinical one, based on medical signs that are evident from a medical history, physical examination, and laboratory tests. Since 2006, testing for *NIPBL* and *SMC1A* has been available through the University of Chicago [27]. Such testing is best accomplished through referral to a genetics specialist or clinic.

TREATMENT

There is no known cure. Because CdLS affects so many different systems of the body, medical management requires a team approach. Treatment varies based on the signs and symptoms presented. For example, poor growth after birth may require supplemental formulas and/or gastrostomy tube placement to meet nutritional needs and/or to decrease pulmonary complications from gastric aspiration. Ongoing physical, occupational, and speech therapies are recommended to improve developmental potential. Medications may be required to prevent or control seizures.

Surgery may be necessary to treat skeletal abnormalities, gastrointestinal problems, congenital heart defects and other health problems. General anesthesia is usually indicated because of poor cooperation and mental retardation especially in children and should be provided by an individual who is experienced in this condition as many problems associated with the procedure have been identified [28]. Antibiotic coverage and aspiration prophylaxis are essential. Structural renal defects are found in 40% and alter drug excretion. Securing the airway may be difficult and a smaller size tube indicated with the immediate availability of a difficult airway cart. Sedatives used as premedication may cause airway obstruction. Response to many drugs may be unpredictable secondary to endocrine disorders.

Support organization information should be given to the family whenever a diagnosis is made: the CdLS Foundation, 1-800-753-2357, www.CdLSusa.org. The CdLS foundation offers cards that are age specific of treatment options as follows [29, 30]:

1. *Infancy and at the time of diagnosis.* A karyotype should be obtained (blood chromosomes analysis sent and evaluated), although it will typically be normal. If a diagnosis is made, several studies and services are recommended including;
2. Echocardiogram, renal ultrasound, pediatric ophthalmologic evaluation with cycloplegic refraction, audiology evaluation (otoacoustic emissions, or brainstem auditory evoked response if audiology is abnormal), upper GI series to rule out malrotation and reflux, evaluation for GERD including pH probe and/or endoscopy, if positive and may require prokinetics or surgical correction with Nissen fundoplication or insertion of a gastrostomy tube, developmental assessment in infancy and continuing every one to three years, early intervention services initiated and continued as long as needed. Growth assessment should use appropriate CdLS growth charts. Treatment with high calorie formulas is often suggested, and may help with weight gain, but, individuals with CdLS appear to grow at their own pace with a high metabolic rate. Molecular testing should be available if parents are interested in further pregnancies and prenatal diagnosis options.
3. *Early Childhood (one to eight years old)* Regular evaluations and immunizations with the primary care provider are indicated. Cryptorchidism should be repaired by 18 months. Developmental services, with school placement and therapy issues should be individualized as most individuals will benefit from physical, occupational and speech therapy. The use of sign language facilitates oral communication. Growth should be monitored via CdLS-specific growth charts. Pediatric dentistry evaluation is

recommended q 6 months. Pediatric ophthalmology evaluation once or annually, should be done if indicated by findings on the first examination. Audiology testing is required every two to three years. Should there be clinical suspicion of worsening or initial signs of GERD, a repeat evaluation should be performed. Endoscopy has the greatest yield, but pH probe could be considered. Signs of potential volvulus such as bilious emesis or bilious withdrawal from the gastrostomy tube or sudden acute abdominal pain is an emergency and may require surgery. All subspecialties should become involved.

4. *Late Childhood (eight years – puberty)* Regular care through the primary care provider is indicated. Orthopedic involvement may be needed for joint contractures, hip complications, bunions, development of scoliosis, or orthotic use. Behavioral assessment includes assessment for ADHD and self-injurious behavior. Ongoing developmental services, with school placement and therapy issues as individualized are indicated. Again, most individuals benefit from physical, occupational and speech therapy. The use of sign language is encouraged to facilitate oral communication. Growth via CdLS-specific growth charts should be monitored as well as periodic dental audiologic and visual assessments. With any clinical suspicion of worsening or initial signs of GERD, a repeat evaluation should be performed. As before, endoscopy will often have the greatest yield, but pH probe could be considered.
5. *Adolescence (puberty – 20 years)* Regular care through the primary care provider is again emphasized with ongoing developmental services. School placement and therapy issues should be individualized. Plans should be initiated early for school or workplace placement after high school with job training and/or higher education. Pelvic examination with Pap smear regularly, or at least every three years, depending on sexual activity, from late adolescence throughout adulthood is indicated for females. Hormonal therapy with patient and family, both from the pregnancy prevention point of view, and management of menstruation (individualized to specific patient and family) should be discussed and recurrence risks identified if developmentally appropriate. Orthopedic involvement may be needed for joint contractures, hip complications, bunions, development of scoliosis, or orthotic use. Behavioral assessment includes ADHD and self-injurious behavior. Physical, occupational and speech therapy are still indicated. Growth via CdLS-specific growth charts should be monitored. Dental, eye and hearing evaluation should continue every 6months- 1 year as indicated. Worsening of GI symptoms must be immediately evaluated.
6. *Adulthood* Regular evaluations with primary care provider should continue, following blood pressure, baseline EKG, routine breast, or testicular and prostate examination as per usual medical guidelines. Job training or work issues, higher education should be discussed as well as behavioral or psychiatric assessment, including ADHD, obsessive-compulsive symptoms, self-injurious behavior, depression. A DEXA scan can rule out osteoporosis. Dental evaluation should be made every four to six months, depending on compliance. Regular pelvic examination with Pap smear at least every

three years, would depend on sexual activity, from late adolescence throughout adulthood. Hormonal therapy with patient and family should be available. Orthopedic involvement may be needed for joint contractures, hip complications, bunions, development of scoliosis, or orthotic use. Ongoing developmental services, with school placement and therapy issues should be individualized and individuals will benefit from physical, occupational and speech therapy. Growth monitoring should continue as well as dental, eye and auditory monitoring with appropriate subspecialist consultation as needed.

SUPPORT SERVICES

The Cornelia de Lange Syndrome (CdLS) Foundation is a nonprofit, family support organization based in Avon, Connecticut, that exists to ensure early and accurate diagnosis of CdLS, promote research into the causes and manifestations of the syndrome, and help people with a diagnosis of CdLS, and others with similar characteristics, to make informed decisions throughout their lives [29]. More information may be obtained from <https://ghr.nlm.nih.gov/condition/cornelia-de-lange-syndrome>.

CONCLUSION

Although specific gene mutations can be found in some patients, genetic testing is usually reserved to confirm an already highly suspected CdLS diagnosis. At present there is no cure for CdLS, which can have wide expression. Treatment is symptomatic and therapy based. Early intervention by means of medical and surgical care is necessary for feeding difficulties, congenital heart disease, urinary, auditory and visual abnormalities, as well as psychomotor delay. The CdLS foundation can provide invaluable support and management to patients and their families for management of this syndrome.

REFERENCES

- [1] Brachmann, W. Ein Fall von symmetrischer Monodaktylie durch Ulnadefekt, mit symmetrischer Flughautbildung in den Ellenbeugen, sowie anderen Abnormitäten (Zwerghaftigkeit, Halsrippen, Behaarung) (A case of symmetrical monodactyly, representing ulnar deficiency, with symmetrical antecubital webbing and other abnormalities, (dwarfism, cervical ribs, hirsutism)). *Jahrbuch fuer Kinderheilkunde und physische Erziehung* 84: 225-235, 1916.
- [2] de Lange, C. Sur un type nouveau de degenerescence (typus Amstelodamensis). [On a new type of degenerescence (Amstelodamensis typus)]. *Arch. Med. Enfants* 36: 713-719, 1933.
- [3] <http://www.whonamedit.com/synd.cfm/1080.html> (accessed Nov 1st 2017).

- [4] Sargent, WW. Anesthetic management of a patient with Cornelia de Lange syndrome. *Anesthesiology* 1991; 74: 1162-1163.
- [5] Jackson L, Kline AD, Barr MA, Koch S. De Lange syndrome: a clinical review of 310 individuals. *American Journal of Medical Genetics* 1993; 47: 940–946.
- [6] Kline AD, Krantz ID, Sommer A, Kliever M., Jackson LG, Fitzpatrick DR, Levin AV, Selicorni A. Cornelia de Lange syndrome: Clinical review, diagnostic and scoring systems, and Anticipatory guidance. *American Journal of medical genetics part A* 2007; 143A: 1287-1296.
- [7] “CdLS Foundation - Characteristics of CdLS”. Cornelia de Lange Syndrome Foundation. November 2017.
- [8] Marchisio P, Selicorni A, Bianchini S, et al. Audiological findings, genotype and clinical severity score in Cornelia de Lange syndrome. *Int J Pediatr Otorhinolaryngol.* 2014; 78(7):1045-8. [Medline].
- [9] Marchisio P, Selicorni A, Pignataro L, et al. Otitis media with effusion and hearing loss in children with Cornelia de Lange syndrome. *Am J Med Genet A.* 2008; 146A(4):426-32.
- [10] Eiason MJ, Melzer JM, Gallagher TQ Int Cornelia de Lange syndrome: What every otolaryngologist should know. *J Trichology.* 2017; 9(3):105-107. doi: 10.4103/ijt.ijt_14_17.
- [11] Wygnanski-Jaffe T, Shin J, Perruzza E, et al. Ophthalmologic findings in the Cornelia de Lange syndrome. *J AAPOS.* 2005. 9:407-10.
- [12] Kline AD, Stanley C, Belevich J, et al. Developmental data on individuals with the Brachmann-de Lange syndrome. *American Journal of Medical Genetics* 1993; 47: 1053–8.
- [13] Saal HM, Samango-Sprouse CA, Rodnan LA, Rosenbaum KN, Custer DA. Brachmann-de Lange syndrome with normal IQ. *American Journal of Medical Genetics* 1993; 47: 995–8.
- [14] Moeschler JB, Graham JM Jr. Mild Brachmann-de Lange syndrome. Phenotypic and developmental characteristics of mildly affected individuals. *American Journal of Medical Genetics* 1993; 47: 969–76.
- [15] Basel-Vanagaite L, Wolf L, Orin M, et al: DOI: 10.1111/cge.12716 *Clinical Genetics* 2015.
- [16] Chatfield KC, Schrier SA, Li J, Clark D, Kaur M, Kline AD, Deardorff MA, Jackson LS, Goldmuntz E, Krantz ID (2012), “Congenital heart disease in Cornelia de Lange syndrome: Phenotype and genotype analysis,” *Am J Med Genet A*, 158A (10): 2499–2505, PMC 3551981, PMID 22965847, doi:10.1002/ajmg.a.35582.
- [17] Selicorni A, Sforzini C, Milani D, et al. Anomalies of the kidney and urinary tract are common in de Lange syndrome. *Am J Med Genet A.* 2005. 132:395-7.
- [18] Krantz ID, McCallum J, DeScipio C, et al. “Cornelia de Lange syndrome is caused by mutations in NIPBL, the human homolog of Drosophila melanogaster Nipped-B.” *Nature Genetics.* 2004; 36 (6): 631–5. PMID 15146186.doi:10.1038/ng1364.

- [19] Tonkin E, Wang TJ, Lisgo S, Bamshad MJ, Strachan T (2004). "NIPBL, encoding a homolog of fungal Scc2-type sister chromatid cohesion proteins and fly Nipped-B, is mutated in Cornelia de Lange syndrome." *Nature Genetics*. 36 (6): 636–641. PMID 15146185. doi:10.1038/ng1363.
- [20] "HDAC8 FAQ Sheet" (PDF). CdLS Foundation Web site. Cornelia de Lange Syndrome Foundation. Retrieved November 2017.
- [21] Revekova E, Focarelli ML, Susani L et al. Cornelia de Lange mutations in SMC1A or SMC3 affect binding to DNA. *Human Molecular Genetics* 2009; 18: 418-427.
- [22] Musio A, Selicorni A, Focarelli ML et al. X-linked Cornelia de Lange syndrome owing to SMC1L1 mutations. *Nature Genetics* 2006; 38: 528-530.
- [23] Deardoff M, Kaur M, Yaeger D et al. Mutations in cohesion complex members SMC3 and SMC1A cause a mild variant of Cornelia de Lange syndrome with predominant mental retardation. *American Journal of Human Genetics* 2007; 80: 485-494 16.
- [24] Xu Y, Sun S, Li N et al. Identification and analysis of the genetic causes in nine unrelated probands with syndromic craniosynostosis. *Gene*. 2017 Oct 13. pii: S0378-1119(17)30886-7. doi: 10.1016/j.gene.2017.10.041. [Epub ahead of print].
- [25] Boudaoud I, Fournier É, Baguette A et al. Connected Gene Communities Underlie Transcriptional Changes in Cornelia de Lange Syndrome. *Genetics*. 2017;207(1):139-151. doi: 10.1534/genetics.117.202291. Epub 2017 Jul 5.
- [26] Moss J, Penhallow J, Ansari M et al. Genotype-phenotype correlations in Cornelia de Lange syndrome: Behavioral characteristics and changes with age. *Am J Med Genet A*. 2017. doi: 10.1002/ajmg.a.38228.
- [27] <http://dnatesting.uchicago.edu/> accessed November 2017.
- [28] Vestergaard L, Dey N, Winding R. Anesthetic considerations in a patient with Cornelia de-Lange syndrome *J Anaesthesiol Clin Pharmacol*. 2015;31(3):419-20. doi: 10.4103/0970-9185.161732.
- [29] Cornelia de Lange Syndrome (CdLS) Foundation, Inc. <http://www.cdlsusa.org/> accessed November 2017.
- [30] <http://www.cdlsusa.org/docs/treatment-guidelines.pdf>.

Further Reading

- Aitken DA, Ireland M, Berry E, et al. Second-trimester pregnancy associated plasma protein-A levels are reduced in Cornelia de Lange syndrome pregnancies. *Prenat Diagn*. 1999; 19(8):706-10.
- Allanson JE, Hennekam RC, Ireland M. De Lange syndrome: subjective and objective comparison of the classical and mild phenotypes. *J Med Genet*. 1997; 34(8):645-50.
- Berney TP, Ireland M, Burn J. Behavioural phenotype of Cornelia de Lange syndrome. *Arch Dis Child*. 1999; 81(4):333-6.
- Bhuiyan ZA, Stewart H, Redeker EJ, Mannens MM, Hennekam RC. Large genomic rearrangements in NIPBL are infrequent in Cornelia de Lange syndrome. *Eur J Hum Genet*. 2007; 15(4):505-8.

- Borck G, Zarhrate M, Bonnefont JP, Munnich A, et al. Incidence and clinical features of X-linked Cornelia de Lange syndrome due to SMC1L1 mutations. *Hum Mutat.* 2007;. 28(2):205-6.
- Braddock SR, Lachman RS, Stoppenhagen CC, et al. Radiological features in Brachmann-de Lange syndrome. *Am J Med Genet.* 1993. 47(7):1006-13.
- Bull MJ, Fitzgerald JF, Heifetz SA, Brei TJ. Gastrointestinal abnormalities: a significant cause of feeding difficulties and failure to thrive in Brachmann-de Lange syndrome. *Am J Med Genet.* 1993; 47(7):1029-34.
- Fitzpatrick DR, Kline AD. Cornelia de Lange syndrome. Cassidy SB, Allanson JE, eds. *Management of Genetic Syndromes.* New York, NY: John Wiley and Sons; 2005.
- Fryns JP. Partial trisomy 4p and Brachmann-de Lange syndrome. *Am J Med Genet.* 2000. 95(4):406.
- Gervasini C, Parenti I, Picinelli C, et al. Molecular characterization of a mosaic NIPBL deletion in a Cornelia de Lange patient with severe phenotype. *Eur J Med Genet.* 2013; 56(3):138-43.
- Gervasini C, Picinelli C, Azzollini J, Rusconi D, Masciadri M, Cereda A, et al. Genomic imbalances in patients with a clinical presentation in the spectrum of Cornelia de Lange syndrome. *BMC Med Genet.* 2013; 14:41.
- Gillis LA, McCallum J, Kaur M, et al. NIPBL mutational analysis in 120 individuals with Cornelia de Lange syndrome and evaluation of genotype-phenotype correlations. *Am J Hum Genet.* 2004. 75:610-23.
- Gil-Rodriguez MC, Deardorff MA, Ansari M, et al. De Novo Heterozygous Mutations in SMC3 Cause a Range of Cornelia de Lange Syndrome-Overlapping Phenotypes. *Hum Mutat.* 2015; 36(4):454-62.
- Goodban MT. Survey of speech and language skills with prognostic indicators in 116 patients with Cornelia de Lange syndrome. *Am J Med Genet.* 1993. 47(7):1059-63.
- Huisman SA, Redeker EJ, Maas SM, Mannens MM, Hennekam RC. High rate of mosaicism in individuals with Cornelia de Lange syndrome. *J Med Genet.* 2013; 50(5):339-44.
- Hyman P, Oliver C, Hall S. Self-injurious behavior, self-restraint, and compulsive behaviors in Cornelia de Lange syndrome. *Am J Ment Retard.* 2002. 107:146-54.
- Hyman P, Oliver C. Causal explanations, concern and optimism regarding self-injurious behaviour displayed by individuals with Cornelia de Lange syndrome: the parents' perspective. *J Intellect Disabil Res.* 2001; 45:326-34.
- Ireland M, Donnai D, Burn J. Brachmann-de Lange syndrome. Delineation of the clinical phenotype. *Am J Med Genet.* 1993 Nov 15. 47(7):959-64.
- Janek KC, Smith DF, Kline AD, et al. Improvement in hearing loss over time in Cornelia de Lange syndrome. *Int J Pediatr Otorhinolaryngol.* 2016;. 87:203-7.
- Kline AD, Barr M, Jackson LG. Growth manifestations in the Brachmann-de Lange syndrome. *Am J Med Genet.* 1993 47(7):1042-9.
- Kline AD, Grados M, Sponseller K, et al. Natural history of aging in Cornelia de Lange syndrome. *Am J Med Genet C Semin Med Genet.* 2007 Aug 15. 145C(3):248-60.
- Kozma C. Autosomal dominant inheritance of Brachmann-de Lange syndrome. *Am J Med Genet.* 1996; 66(4):445-8.
- Manouvrier S, Espinasse M, Vaast P, et al. Brachmann-de Lange syndrome: pre- and postnatal findings. *Am J Med Genet.* 1996; 62(3):268-73.

- Mende RH, Drake DP, Olomi RM, Hamel BC. Cornelia de Lange Syndrome: A Newborn with Imperforate Anus and a NIPBL Mutation. *Case Rep Genet.* 2012. 2012:247683.
- Schrier SA, Sherer I, Deardorff MA, Clark D, Audette L, Gillis L, et al. Causes of death and autopsy findings in a large study cohort of individuals with Cornelia de Lange syndrome and review of the literature. *Am J Med Genet A.* 2011; 155A(12):3007-24.
- Wilmink FA, Papatsonis DN, Grijseels EW, Wessels MW. Cornelia de lange syndrome: a recognizable fetal phenotype. *Fetal Diagn Ther.* 2009. 26(1):50-3.

Chapter 78

COGNITIVE-BEHAVIORAL PHENOTYPE IN KLINEFELTER SYNDROME (KS) AND OTHER SEX CHROMOSOMAL ANEUPLOIDIES (SCAs): CLINICAL VARIABILITY

**A. P. Verri*, C. D'Angelo, A. Cremante, F. Clerici,
A. Mauri and C. Castelletti**

IRCCS National Neurological Institute "C. Mondino" Foundation,
Laboratorio di Psicologia Cognitivo-Comportamentale,
Pavia, Italy

ABSTRACT

Background: SCAs are the most frequently occurring chromosomal abnormalities with an incidence of 1 in 400 births. As the number of X chromosomes increases, the phenotypic severity increases as well and it is estimated that cognitive abilities decrease by 10–15 IQ points for each additional X chromosome.

Aim of this paper is to illustrate clinical variability of cognitive-behavioral phenotype in the different SCAs.

Design and Methods: The sample was composed by 53 subjects (mean age = 21.16 yrs., range: 13-54) with karyotype 47, XXY (73%), 49, XXXXY (7%), 48, XXYY (9%), mosaicism 47, XXY/48, XXXY (2%), 47, XYY (5%), 48, XXXY (2%), 49, XXXYY (2%).

Only 5 subjects have been diagnosed prenatally (4 KS and 1 XXYY).

Primary caregivers completed a comprehensive questionnaire detailing birth, medical, developmental and psychological history. Cognitive and behavioral assessment was performed with clinical interviews using DSM 5 criteria and psychometric questionnaires (WISC-R, WAIS-R, CPM, Token Test, VABS, SCL90, SCQ).

Twenty-one sex and age matched subjects karyotypically normal were also evaluated from the behavioural point of view.

Results: Mean IQ in typical KS was 87.45 ± 2 ds (sd = 20.12) range 45-123, VIQ 91.74 (sd = 19.55) range 50-130 and PIQ 86.87 (sd = 20.87) range 50-126.

* Corresponding Author's Email: annapia.verri@mondino.it (Neuropsychiatrist and Neurologist).

Mean IQ in other SCAs was 68.71 (sd = 20.81) range 45-106, VIQ 69.36 (sd = 21.97) range 47-113 and PIQ 74.72 (sd = 21.70) range 45-112.

In CPM KS subjects scored 27.75 (range 13-36) and 31.50 in the Token Test (range 21-35) while in CPM the other SCAs subjects scored 22.27 (range 10-35) and 22.50 (range 9-31) in the Token Test ($p < .05$).

VABS scores documented more marked impairment on adaptive behavior in atypical SCAs subjects. SCL90 documented an elevation of paranoid scale in the 70% of KS subjects and 50% of other SCAs.

Autistic traits were present in 67% of the other SCAs subjects and in the 18% of KS at the SCQ.

Conclusion: A precise identification of the cognitive and behavioral phenotype in different SCAs may enhance the clinical treatment, anticipatory guidance, and care throughout the lifespan.

1. INTRODUCTION

Klinefelter's Syndrome (KS), described for the first time in 1942, is a genetic non-inherited pathology which is caused by an alteration of the number of sex chromosomes in male subjects. Patients affected by this syndrome have 47 chromosomes, due to the presence of one supernumerary X chromosome. This genetic anomaly influences sexual development and physical appearance, cognitive functions, motor and language development, and social skills. Cognitive abilities are typically in the average to low average range with weaknesses in verbal skills. Language difficulties are one of the most distinctive traits in cognitive functioning of people with KS. In fact, KS show an increased risk for developmental delays, speech-language disorders and learning disorders. Limitations in communication and behavioral aspects markedly affect social adaptation and the development of personality.

This prominent effects of the extra X chromosome on cognition, particularly in the language domain [1], has been documented also by fMRI studies. Measuring the patterns of brain activity during language processing in KS men, it was shown that language activity in the brain was less lateralized in the experimental group as compared to controls. It revealed an increased activity in the right hemisphere rather than reduced activity in the left hemisphere that causes a loss of asymmetric processing of language. The regions mostly involved was the Superior Temporal Gyrus (STG) and the supramarginal gyrus region, which is close to the posterior section of the STG and part of Wernicke's area. Reduced language laterality in the STG was highly correlated with the degree of disorganization of thought and language. Decreased functional asymmetry of language areas in the brain in XXY men may be secondary to abnormal X chromosomal inactivation [2].

Moreover, subjects with KS are predisposed to psychopathological risk. Behavior can include hyperactivity, attention problems, impulsivity, aggression, mood instability and autistic traits. In particular, Boks [3] observed the presence of a great variability of symptoms in a group of KS boys: learning disorders (65%), ADHD (63%), depressive disorders (24%), psychotic disorders (8%) and schizophrenia (2%). The risk of hospitalization for psychosis in adults KS is greater than in control subjects [4].

Some problems appear after the onset of puberty, when physical differences in KS become more evident and might result in body image disorders, sense of isolation and shame. Low self-esteem, anxiety, problems of socialization and mood disorders occur in boys with KS during

adolescence [5]. The presence of learning difficulties at school, mild cognitive impairment, problems in achieving good academic results often cause feelings of distrust even in childhood. The lack of integration within the peer group is the major source of anxiety and mood disorders. Many of the KS subjects seem to be more sensitive, anxious and insecure, and show a higher incidence of anxious-depressive disorders than the general population and an increased propensity to the use of drugs [5]. Some studies have emphasized that people with KS are friendly and open to interactions, do not usually have major problems with social interaction and adaptation, although they may be shy, sensitive and unassertive [6, 7]. Other studies have showed that males with KS have difficulties in the construction of satisfying social relationships, with antisocial behavior in adolescence and a more unstable occupational history, but a minority of them meet criteria for antisocial behavior disorder in adulthood [8].

KS is characterized by a constellation of physical symptoms: inadequate virilization, hypogonadism, azoospermia, infertility, gynecomastia, elevated average height (179.2 ± 6.2 cm) and increased plasma gonadotrophins [9], [10]. These problems are progressive and they begin to appear with more evidence during adolescence, in correspondence of sexual development.

Aneuploidy 47, XXY is the most common abnormality of sex chromosomes in humans, with an incidence equal to 1/450 male live births. About 80% of the KS patients show an XYY karyotype, but the 20% have other numeric sex chromosomes abnormalities (48, XXXY, 48, XXYY, 49, XXXXY, 46, XY/47, XXY mosaicism) or structurally abnormal sex chromosomes [11]. As the number of X chromosomes increases, the phenotypic severity increases as well and it is estimated that cognitive abilities decrease by 10–15 IQ points for each additional X chromosome [12].

The presence of one or more additional X chromosome(s) above the typical 46, XY in males leads to testicular dysgenesis and hypergonadotropic hypogonadism, and thus, 48, XXYY, 48, XXXY and 49, XXXXY are considered ‘variants’ of KS (47, XXY) because of these shared features. However, the increased risks for congenital malformations, additional medical problems and more complex psychological involvement in these other SCAs make distinction from 47, XXY important for these patients [12].

47, XXY (KS) is associated with tall stature, with studies reporting a mean adult height ranging from 179 to 188 cm. This characteristic is also of 48, XXXY and 48, XXYY males, in contrast, stature in males with 49, XXXXY syndrome is usually below average. A hypothesis is that the influence of extreme over dosage of sex chromosome genes in the pentasomic condition, affects multiple organ sites and growth pathways. Considering again the body habitus, other SCAs vary from being underweight to obese, and only approximately 30% have gynecomastia which is usually mild.

The degree of facial dysmorphism in all three syndromes is variable and often subtle, although dysmorphic features are typically more distinct in 49, XXXXY compared with 48, XXYY and 48, XXXY. Across all three conditions, common findings include hypertelorism, epicanthal folds, up-slanting palpebral fissures, hooded eyelids, significant dental problems. All of the findings above have also been described in 47, XXY, but they occur more frequently in other SCAs.

Developmental delays are common in infancy and early childhood, with speech delays, especially in expressive language, and motor delay associated with hypotonia, in almost all patients.

Visootsak et al., [7], using an adaptive functioning assessment (Vineland Adaptive Behavior Scales) in SCAs subjects, found that the mean standardized scores for adaptive functioning were in the disability range.

Other neurodevelopmental and psychological disorders are significant components of the phenotype of other SCAs and are typically more severe and/or complex when compared with typical KS. Developmental dyspraxia contributes to the early language and motor deficits. Attention Deficit Hyperactivity Disorder (ADHD) is present in over 70% of them, significantly higher than typical KS.

At the end, some emotional symptoms, as emotional immaturity, anxiety symptoms, obsessive-compulsive behaviours, impulsivity, behavioural dysregulation and tic disorders, are more commonly seen in these conditions compared with typical KS.

There is still little of knowledge about the cognitive behavioral phenotype of other SCAs. It's sure, anyway, that these patients are considered as a heterogeneous group with different and distinctive characteristics from typical KS. The focus of interest in the study of SCA was the genetics and the neurodevelopmental risk. Accumulating evidence that such genetic conditions not only impact physical development, but also psychological development, increased the awareness of the importance to study also psychological, emotional and behavioral problems [2]. The aim of this research is to compare a group of KS subjects and one with different SCAs, considering cognitive behavioral phenotype. In fact, some studies found that the number of supernumerary X negatively correlated with intellectual development and an increase of symptoms [13]. Moreover, we considered the history of epilepsy, autistic traits and psychosis as important variable of distinction between the two groups. We hypothesize that the other SCAs show more cognitive impairment and more developmental problems than typical KS.

2. PRESENT STUDY

The sample was composed by 53 adolescents and adults (mean age = 21.16 yrs, range: 13-54) with karyotype 47, XXY (73%), 49, XXXXY (7%), 48, XXYY (9%), mosaicism 47,XXY/48, XXXY(2%), 47,XXY (5%), 48, XXXY (2%), 49, XXXYY (2%) (Figure. 1; Table. 1).

These subjects were referred to the Laboratory of Cognitive Behavioral Psychology of the National Neurological Institute "C. Mondino" Foundation of Pavia, Italy. The inclusion criteria was first of all the genetic diagnosis of SCAs; moreover, the age: from all the Klinefelter subjects evaluated we chose patients from 13 years old. All the patients have been karyotyped using standard techniques, except two of them, who were diagnosed during a screening for intellectual disability, using the Array-Comparative Genomic Hybridation (CGH) molecular cytogenetic method.

Primary caregivers completed a comprehensive questionnaire detailing birth, medical, developmental and psychological history [14]. Cognitive and behavioral assessment was performed through a clinical interview made with the DSM-IV criteria and psychometric questionnaires. For the assessment of global cognitive functioning, Wechsler Scales (WISC-R - Wechsler Intelligence Scale for Children - Revised [15], WAIS-R – Wechsler Adult Intelligence Scale – Revised, [16] and Coloured Progressive Matrices (CPM) [17] were used.

The Token Test [18] was used to evaluate the ability of oral comprehension. The adaptive behaviour, the ability of personal and social self-sufficiency in real-life situations and the way in which cognitive abilities translate into management of self-autonomy in daily-life, were evaluated with The Vineland Adaptive Behaviour Scales (VABS) [19]. The Symptom Checklist (SCL-90-R) [20] is considered a valid instrument of general psychological distress in patients with experiencing a range of mental health and medical conditions. The Social Communication Questionnaire (SCQ) [21] offers a quick way to screen for Autism Spectrum Disorder. It was used to evaluate communicative, social and relational skills, and it was administered to parents. At last, (BSRI) [22] measured different aspects of psychological gender traits.

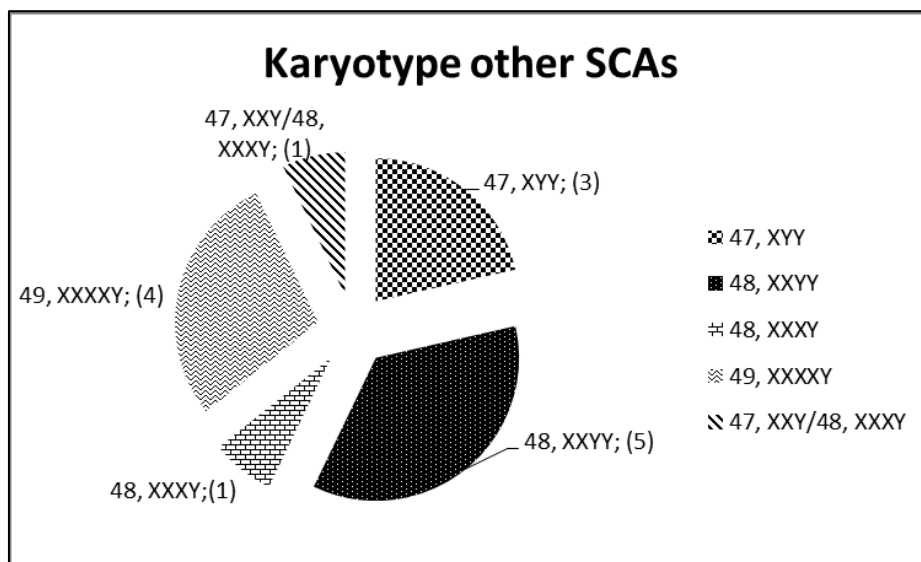


Figure 1. The graphic highlights other SCAs. Five subjects showed 48, XXYY karyotype, four 49, XXXXY, three 47, XYY and one 48, XXXY. One subject showed mosaicism (47, XXY/48, XXXY).

Table 1. Population

Karyotype	N	Mean age at first evaluation
47, XXY	39	24.43
Other SCAs	14	19.66
47, XYY	3	27
48, XXYY	5	20.8
48, XXXY	1	12
49, XXXXY	4	15.5
47, XXY/48, XXXY	1	23
Total	53	21.16

The table highlights mean age of the patients at first evaluation, grouped according to the karyotype.

Only 5 subjects have been diagnosed prenatally (4 KS and 1 XXYY) (Table 2 and 3).

Table 2. Age at diagnosis KS

Age at diagnosis	Percentage (n = 39)
Prenatal diagnosis	8,89% (4)
1-10 years	17,78% (8)
10- 18 years	35,56%(16)
18+ years	20%(9)

* 2 patients have not data.

Mean age of diagnosis across the entire group KS was 12, 89 years. The most frequent diagnosis was in the range of 10-18 years; prenatal diagnosis was the less frequent.

Table 3. Age at diagnosis other SCAs

Age at diagnosis	Percentage (n = 14)
Prenatal diagnosis	7,14% (1)
1-10 years	35,71% (6)
10- 18 years	42,86% (5)
18+ years	14,29% (2)

Mean age of diagnosis across the entire group other SCAs was 11,57 years. The most frequent diagnosis was in the range of 10-18 years; prenatal diagnosis was the less frequent.

A group of control, composed by twenty-one sex and age matched subjects, karyotypically normal, completed behavioral questionnaires (CPM, SCL90, BSRI). The statistical evaluation was carried out using the T-test and then we used the effect size Hedges' *g* to have a quantitative measure of the strength of the phenomenon, considering the different sizes of the groups.

3. RESULTS

3.1. Developmental and Clinical History

Anamnestic data (Table 4) confirmed the increased prevalence of psycho-motor delay in other SCAs as compared with typical KS.

The 70% of other SCAs show language delay, in contrast to only 30% of typical KS. Regarding as motor development, the 57% of other SCAs present a delay, compared with 15% of typical KS. Psychotic disorders emerged for the 40% of other SCAs, compared with 22,5% of typical KS. Epilepsy was diagnosed in the 25% of cases of other SCAs (66% of them were affected by generalized epilepsy, the others by focal) and in the 12,5% of typical KS (75% of them present generalized epilepsy, the others focal). Interestingly the 57% of subjects with epilepsy show also a psychotic disorder. On a statistic level, correlations with T-test confirmed a significant difference between the groups both for language delay ($t = .032$; $p < .05$) and for motor delay ($t = .013$; $p < .05$). The effect size for language delay was Hedges' *g* = 0.114765, while for motor delay was Hedges' *g* = 1.247012. Otherwise, for epilepsy and psychosis, between groups analysis didn't find out any significant difference (E: $t = .316$; P: $t = .158$), probably due to a large variability and heterogeneity into a too much small sample.

Table 4. Anamnestic data

* Neurodevelopmental and Previous Neuropsychiatric DIAGNOSES	KS (39)	“other SCAs” [14]
	%	%
Speech delay	33,33	64,28
Motor delay	17,95	50,01
Learning disability	35,90	57,14
Loneliness	41,02	64,28
epilepsy	12,50	25
ADHD	17,95	21,43
Impulse control disorder	30,77	35,71
Psychotic Disorder	22,50	40
Tic Disorder	10,26	14,28
Generalized anxiety disorder	35,90	35,71
Obsessive-compulsive disorder	15,38	42,86
Mood Disorders	35,90	21,43

*Anamnestic data.

The table highlights that other SCAs showed more frequent developmental problems compared to KS, except for generalized anxiety disorder and Mood Disorders, that were more frequent in KS.

3.2. DSM-IV Diagnosis

Considering the axis I disorders, 18% of typical KS and 21% of other SCAs present a diagnosis of ADHD, while the 36% of typical KS and the 58% of other SCAs show a diagnosis of Learning Disabilities. Psychotic Disorder emerged for the 40% of other SCAs, compared to 22,5% of typical KS. Finally, in the 50% of all patients, adaptive and behavioral problems and depressive or anxious traits were signaled.

On axis II, in the typical KS group, the 8% of patients are in the range of moderate-severe intellectual disability and the 17% of mild intellectual disability. Thirty-eight per cent of them showed an IQ level borderline, 24% were in the normal range and 13% present an IQ level higher than normal. In the other SCAs group, instead, the 31% of patients are in the range of moderate-severe intellectual disability and the 31% of mild intellectual disability. Twenty-three per cent of them showed an IQ level borderline, 15% were in the normal range and nobody presents an IQ level higher than normal.

3.3. IQ and Adaptive Behavior

Mean IQ in typical KS was $87,45 \pm 2$ sd (sd = 20,12) range 45-123, VIQ 91,74 (sd = 19,55) range 50-130 and PIQ 86,87 (sd = 20,87) range 50-126. Instead, mean IQ in other SCAs was 68, 71 (sd = 20,81) range 45-106, VIQ 69,36 (sd=21,97) range 47-113 and PIQ 74,72 (sd = 21,70) range 45-112 (Figures 2 and 3).

In CPM KS subjects obtained as mean score 27,75 (range 13-36, moda = 33) and 31,50 in the Token Test (range 21-35), while in CPM the SCAs subjects mean score was 22,27 (range 10-35, moda = 29) and 22,50 (range 9-31) in the Token Test.

Moreover, between groups analysis (t-test) identified significant differences between KS and other SCAs in mean IQ, lower in other SCAs ($p < .05$). The effect size for the mean IQ was Hedges' $g = 0.923239$, that means that the result went in the expected direction. In CPM, we found out a significant difference between KS/control group and other SCAs/control group ($p < .05$). In Token Test scores, a significant difference between groups was found ($p < .05$). The effect size was Hedges' $g = 6.121377$. Vineland scale scores documented more marked impairment in other SCAs subjects on adaptive behavior than in KS subjects ($p < .05$) (Figure 4).

In particular, for the other SCAs, communication, socialization and motor abilities resulted as weak points, compared to the average of normative population. Also for the typical KS, communication and motor abilities prove to be weak points, while socialization skills were worse than controls, but less compromised than in other SCAs. At the end, both for other SCAs and for typical KS, the daily-life abilities resulted as strong points.

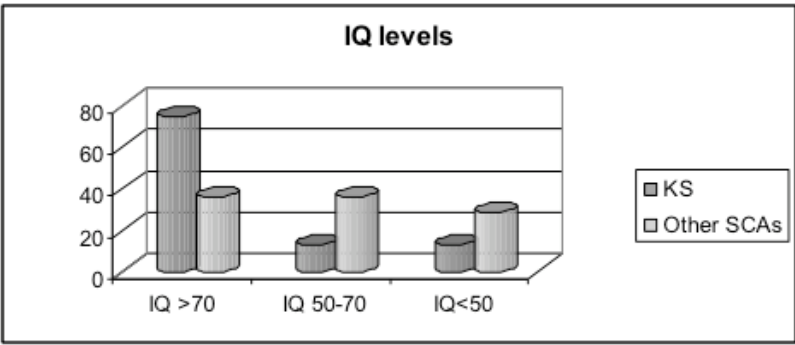


Figure 2. Percentages of IQ levels for KS and for other SCAs. IQ is significantly higher in KS than in other SCAs.

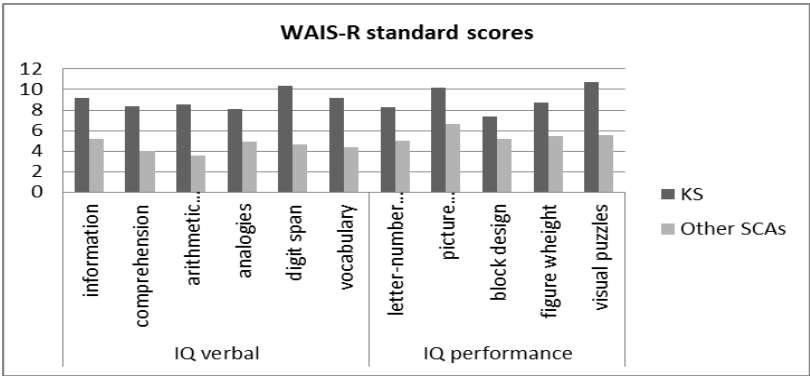


Figure 3. The graphic shows results at the WAIS-R. KS obtained higher scores in all scales. The difference between groups was more evident in the verbal IQ.

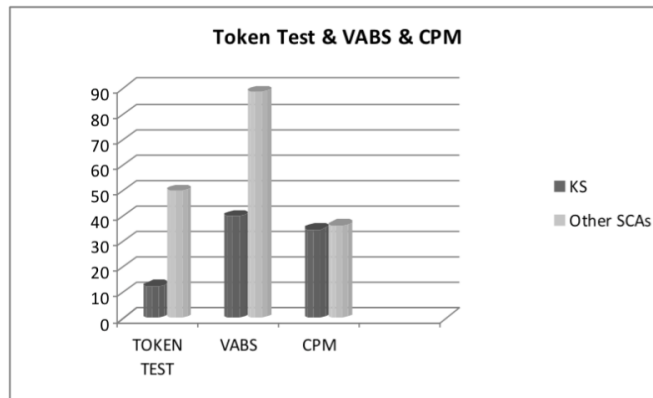


Figure 4. The figure shows results of groups at the Token Test, VABS and CPM. In all tests other SCAs showed worse scores.

3.3.1. SCL-90

SCL90 documented an elevation of psychotic traits in KS/SCAs (about 50%) subjects in comparison with the control group (5%). There was also an elevation in somatization scale in KS subjects (58% KS, 16% other SCAs and 27% control groups), and in paranoid scale (70% for KS; 50% both for other SCAs and control group). However, there wasn't a statistical significant difference between the groups (Figure. 5).

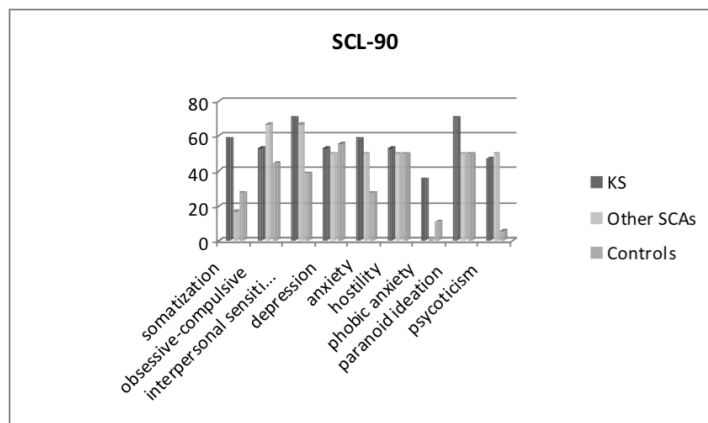


Figure 5. This figure shows the results of the SCL-90. KS scores were higher in somatization, interpersonal sensitivity, anxiety, hostility, phobic anxiety and paranoid ideation scales, compared to other SCAs and controls. Other SCAs' scores were higher than the other groups in the obsessive-compulsive scale and in the psychoticism scales. Other SCAs showed a score approximately to the zero in the phobic anxiety. Controls' scores were higher in the depression scale and lower in the psychoticism scale, compared to other SCAs and KS.

3.3.2. SCQ

Considering the comorbidity between KS and ASD, reported by the literature [23], the SCQ was used to verify the presence of mild autistic traits. They were, in fact, present in 67% of the other SCAs subjects and in the 18% of KS. A statistical significant difference was found between the two groups. (Figure 6). The effect size was Hedges' $g = 1.002256$.

3.4. Bem Sex Role Inventory

In the feminine and masculine scales of BSRI profiles there was a significant difference between KS and other SCAs ($p<.05$). KS subjects show a low masculine scale in comparison with both SCAs and controls, while feminine scores are similar to controls. In other SCAs subjects the feminine scores were very low, while they showed a very high masculine scale. (Figure 7).

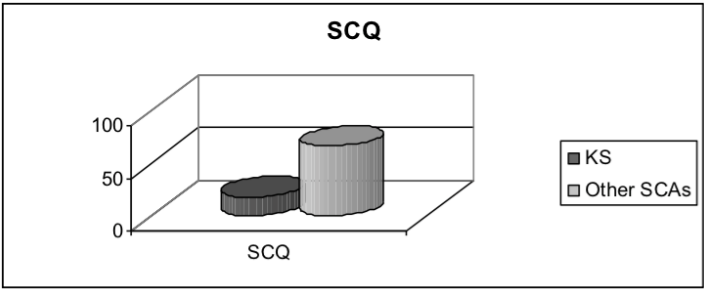


Figure 6. The graphic highlights that other SCAs showed more occurrence of autistic traits compared to KS.

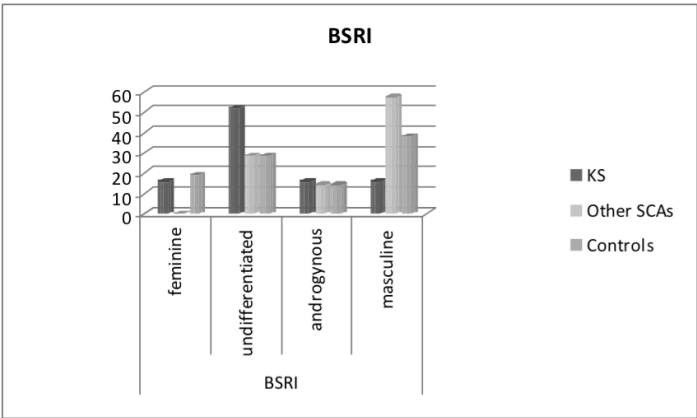


Figure 7. Graphic highlights that other SCAs showed higher scores in the masculine scale, but in the feminine scale they obtained a score approximately to the zero. In the undifferentiated scale and in the androgynous scales, other SCAs scores were similar to controls. KS scores were quite homogeneous in each scale, except for the undifferentiated scale, where they got a higher score than controls and SCAs. In the androgynous scale, scores were very similar between the three groups.

4. DISCUSSION

The aneuploidy of the sex chromosomes is not usually associated with intellectual disability, but it is characterized by the presence of specific cognitive profiles. However, the cognitive level in KS proves to be in mean ten points lower than those of their brothers or peers [11]. The typical cognitive profile is mainly characterized by the presence of the discrepancy between scores on performance tasks and those achieved in the verbal subtests, in favor of the

former. Some studies have shown that verbal IQ scores are 10 points below the average compared with those of performance IQ [24]. This discrepancy may change during the life stages: differences in verbal conceptual and non verbal reasoning skills may diminish over time, with the former no longer being such an area of deficit. In our sample, mean IQ in typical KS was 87,45, in a range of normality/low average. Unlike what suggested in literature, mean of VIQ (91,74) was higher than mean of PIQ (86,87). This could be due to the use of alternative strategies for problem solving that require the use of verbal reasoning ability, learned through experience to compensate for their present difficulties, or may even be linked to the effects of hormonal therapies [24]. On the other hand, in our other SCAs subjects, mean IQ was 68,71, VIQ was 69,36 and PIQ was 74,72, in line with previous studies. Moreover this data confirm our hypothesis: the other SCAs show a lower IQ than typical KS. Other SCAs present an highlighted language, motor and social delay during the development, and a greater incidence of seizures and psychotic disorders. These data are consistent with literature. In fact, it was found that the number of supernumerary X negatively correlated with intellectual disability and an increase of symptoms [13]. Through parents' interviews, we found that a motor delay in the acquisition of first steps was more marked for other SCAs (about 19 months) than typical KS (about 14 months) (Table 4).

Adaptive functioning skills were found to be significantly lower than IQ in most cases, with a mean adaptive functioning in the disability range. These findings indicate that overall daily functioning is often more impaired than would be expected based on cognitive (IQ) scores. While the factors involved in the discrepancy between cognitive abilities and adaptive functioning deficits are not fully understood, these deficits contribute to the disability and prevent many individuals from achieving academic and occupational success [12].

A frequent problem reported by the patients were the seizures, documented by clinical history. Epilepsy was diagnosed in the 25% of cases of other SCAs and in the 12,5% of typical KS. The incidence is higher for other SCAs in comparison to typical KS. In fact, it is known that epilepsy is a common health problem among people with Intellectual Disability (ID). The estimated prevalence of epilepsy in people with ID ranges from 15-30%, while the prevalence of epilepsy in the general population is estimated at 0.6–1% [25].

Seizures in KS and in other SCAs are rare [26-28]. Probably we have found a higher incidence because our laboratory is located in an Institute for neurology. Cognitive functioning of KS subjects is characterized by the difficulty in expressive language. Language difficulties have been identified in 70-80% of children with KS at an early age [29]. Language difficulties include delay in onset of first words and in acquisition of the main stages of language development, in reading, expression, writing and reasoning abilities in arithmetic. Children with KS show difficulties in expression, inability to communicate their thoughts, ideas and emotions; otherwise the comprehension ability is in the standard. Often, during the developmental age, these problems are framed in learning disabilities as dyslexia and dysorthography [30]. Receptive language deficits have also been noted. Problems with phonemic discrimination, processing speed and comprehension of grammatical and morphological aspects of language have been reported [31-34]. Limitations in material processing speed and memory of auditory verbal material, which are associated with problems in decoding words, have been found in individuals with KS.

In our sample, these data were confirmed already by parents' reports. Both typical KS and other SCAs showed a delay in the acquisition of first words in comparison with typical children (typical KS: 16 months; other SCAs 17 months). Consequently the production of the first

sentences results difficult and slow, in fact, in our sample, children of both groups have reached this ability at about 31 months of age. It's obvious that there is a deficit in expressive language already in early development. As reported above in the data of verbal IQ, this delay persists during the life span.

Early language difficulties influence social adaptation and behavior disorder, together with personality development [35].

Moreover, adults with KS displayed relative difficulties in discriminating emotions in tone of voice, and, to a lesser extent, in verbal content [2]. This finding suggests that the XXY chromosomal pattern may not only be associated with difficulties in semantic aspects of language, but with prosodic aspects, as well. This finding may contribute to the development of more comprehensive models addressing the role of the X chromosome in normal and abnormal development of social communication. In fact, the limitation in communication markedly affects social adaptation and behavioral aspects, as well as the development of personality, even if the literature is not unanimous in describing social traits and personality in KS.

The sex chromosome aneuploidies are considered a risk factor for psychosis and psychopathology [36-38]. A higher incidence of psychiatric disorders, as anxiety, depression, behavioral disorder and schizophrenia, has been documented in people with KS compared with general population [5, 39]. A recent study reported that 8% of KS meet criteria for psychotic disorder, 45% have isolated psychotic symptoms and 24% meet criteria for depressive disorder [40]. In our study, psychosis problems emerged for the 40% of other SCAs, and for the 22,5% of typical KS. The SCL-90 documented psychotic traits in about 50% of KS/SCAs subjects in comparison with 5% of control group. There was also an elevation in somatization and in paranoid scales. A risk for hospitalization for KS adults is higher than the controls [4]. Subjects with both KS and schizophrenia show structural and functional anomalies in the Central Nervous System [4]. Sex chromosome aneuploidy is considered a risk factor for the development of psychiatric disorder.

Neuroimaging studies have documented anomalies in the brain structures of boys and adults with KS, which correlated with the presence of psychosocial problems [41]. These psychopathological aspects may be partly explained by a psycho-neurological phenotype that includes grey matter deficits in the superior temporal gyrus, the orbitofrontal cortex and the inferior frontal gyrus, white matter anomalies, impaired executive functions with severe deficits in the inhibitory component, abnormal structure of amygdala, caudate and putamen [37, 42]. These alterations may be caused by excessive expression of genes that lie in the pseudo-autosomal regions of the X-chromosome [41]. Moreover, total brain volume in typical KS is 7-8% smaller than healthy age-matched controls, whereas a reduction of 20% in males with other SCAs was found [43]. These reductions suggests that the X chromosome influences overall brain volume to a greater extent than the number of sex chromosomes in total [43].

Considering the different aspects of personality, many of the KS people seem to be more sensitive, anxious and insecure, and show a higher incidence of anxious-depressive disorders than the general population and an increased propensity to the use of drugs [5]. Other studies have emphasize that people with KS are friendly and open to interactions, do not usually have major problems with social interaction and adaptation, although they may be shy, sensitive and unassertive [6, 44]. KS subjects have difficulties in the construction of satisfying social relationships, with antisocial behavior in adolescence and a more unstable occupational history, but a minority of them meet criteria for antisocial behavior disorder in adulthood [8]. Although

children with KS are reported by clinicians to be hyperactive and to show difficulties in concentration, some studies have showed that children with KS have a docile temperament and lower activity levels compared with unaffected peers [10].

Gives a measurement of psychological androgyny, or high levels of both masculinity and femininity, considering that gender roles may be defined as “expectations about what is appropriate behavior for each sex” [44, 45]. The masculine scale has items as aggressive, ambitious, analytical, competitive, dominant, forceful, individualistic, while the feminine scale’s items are for example affectionate, cheerful, compassionate, gentle, tender, sympathetic. Men with KS exhibit marked variations in phenotype, which may range from males with severe signs of androgen deficiency to normally virilized males. In our sample we found out an interesting statistical significant difference in the scores of BSRI between typical KS and other SCAs. In other SCAs subjects, the feminine scores were very low, quite approximately to the zero, while the masculine scores are high. On the contrary, typical KS presented a feminine scores quite similar to those of the controls, but they showed a low masculine scale.

Another important aspect in the KS phenotype, considering a neuropsychological perspective, is the impact on the area of social information processing. Men with KS are reported to have inaccurate perception of social-emotional cues and difficulties in expressing their emotions [46]. A study have showed that 27% of the boys with KS met criteria for autism spectrum disorders [40]. In fact, children with KS show significant impairments in social cognition, in particular, they present deficit in ToM (Theory of Mind) when compared with the typically developing children, with performance not different from children with ASD, independent of level of intellectual functioning, receptive and expressive language. Impaired ToM may result in social difficulties. Moreover, both KS and ASD, also show difficulties in facial affect recognition, specifically in identifying angry facial expressions [47]. However, this mentalizing deficit may be related to a different set of cognitive dysfunctions: a recent neuroimaging study, showed frontal deficits in KS group in contrast to amygdala deficits in the ASD group [47]. In our sample, autistic traits were present in 67% of the other SCAs subjects and in the 18% of KS at the SCQ, with a statistical significant difference between groups.

Taking in account of the variability of cognitive behavioral phenotype and considering the DSM-IV diagnosis of each patient, the incidence of diagnosis of Intellectual Disability, of ADHD and of Learning Disabilities, is higher for other SCAs than for typical KS. In the same way, also at a diagnostic level, the frame of SCAs subjects appears generally more compromised.

Our hypothesis has been confirmed: other SCAs show more impairment and more developmental problems than typical KS, taking into consideration their cognitive behavioral phenotype. Anyway the cognitive profile in KS is characterized by marked variability and this could be influenced also by an early diagnosis, useful in order to plan different types of rehabilitation, when the developmental disorders is evident from a clinical point of view. A treatment, when begins in early life, could prevent some difficulties and developmental risk, considered with specific regard to the language and subsequently with possible emotional and behavioral problems. In fact, KS subjects, who had prenatal diagnosis, develop learning and language disabilities in a lower proportion than patients diagnosed by chance [48].

A limit of the present study is to have not considered the correlation between the age of diagnosis and the different disabilities of patients. In relation to these medical and psychological different impairments, we did not analyze the influence of the medical complications (cardiological, gastric, bone problems) in the SCA subjects and their consequent

cognitive and behavioral phenotype [49]. Moreover, our sample is small and in the future it could be interesting to study more subjects. It could be also intriguing to consider the role of the medical treatment of the patients, when it is applied. In fact, the diagnosis and therapy of andrological diseases interact with fertility and sexuality that are more sensitive to psychological, educational, cultural, religious and social factor than any other body function. In KS patients, androgen effects on appearance and social characteristics are modulated by the androgen receptor CAGn polymorphism [41]. The CAGn length has a marked influence on the social status in Ks patients. Men with shorter CAGn, and so higher androgenic activity, present more fertility problems than endocrine disorders. These men are more likely to live with a partner because they are sufficiently virilized, and so then they could present with the desire for paternity. This frame could influence positively in particular the social and behavioral aspects of men with KS. It could be interesting analyze if also at an andrological level exists a significant difference between typical SCAs and other KS, that could condition their phenotype. At the end, in relation to the assumption of therapy, a limit of the present study could be the lack of consideration of its possible influence in particular on the BSRI scores.

CONCLUSION

Klinefelter's Syndrome (KS), is a genetic non-inherited pathology which influences sexual development and physical appearance, cognitive functions, motor and language development, and social skills. A high variability of cognitive and behavior features is a characteristic of different SCAs disorders. The present study confirmed that other SCAs demonstrate more impairment and more developmental problems than typical KS. As the number of X chromosomes increases, the phenotypic severity increases as well. The other SCAs show a lower IQ than typical KS. The same picture emerged for the other ability and problems taken into consideration: other SCAs present an highlighted language, motor and social delay during the development, and a greater possibility to succumb to episodes of seizures and to psychotic disorders. These data are consistent with literature. Taking into consideration the variability of cognitive behavioral phenotype and considering the DSM-IV diagnosis of each patient, the incidence of diagnosis of intellectual disability, of ADHD and of learning disabilities, is higher for other SCAs, as well as for autistic traits.

It appears obvious that the variability in cognitive behavioral phenotypes in KS is wide and future studies could continue to investigate the various problems. In fact, it is important to note that an early identification of the cognitive and behavioral phenotypes in all patients with KS, typical and atypical, may enhance the clinical treatment, anticipatory guidance, and care throughout the lifespan.

REFERENCES

- [1] Geschwind, K. B., Swerdloff, R. S. (2000). Neurobehavioral phenotype of Klinefelter syndrome. *Ment. Retard. Dev., D R* (VI): 107-116.

- [2] Van Rijn, S., Aleman, A., Swaab, H., Kahn, R. (2007). What it is said versus how it is said: comprehension of affective prosody in men with Klinefelter (47, XXY) syndrome? *JINS*. 13 (VI): 1065-70.
- [3] Boks, P. M., De Vette, M. H., Sommer, I. E., Van Rijn, S., Giltay, J. C., et al. (2007). Psychiatric morbidity and X-chromosomal origin in Klinefelter sample. *Schizophr. Res.* (XCIII): 399-402.
- [4] Bojesen, A., Juul, S., Gravholt, C. H. (2003). Prenatal and Postnatal Prevalence of Klinefelter Syndrome: A National Registry Study. *J. Clin. Endocr. Metab.* 88 (II): 622-626.
- [5] Bender, B. G., Harmon, R. J., Linden, M. G., Robinson, A. (1995). Psychosocial adaptation of 39 adolescents with sex chromosome abnormalities. *Pediatrics* (XCVI): 302-308.
- [6] Schiavi, R. C., Theilgaard, A., Owen, D. R., White, D. (1988). Sex chromosome anomalies, hormones, and sexuality. *Arch. Gen. Psychiat.* (VL): 19-24.
- [7] Visootsak, J. J., Rosner, B., Dykens, E., Tartaglia, N. (2006). Adaptive and Maladaptive Behavior of Males with Sex Chromosome Aneuploidy. *J. Invest. Med.* (LIV): S 280.
- [8] Gotz, M. J., Johnstone, E. C., Ratcliffe, S. G. (1999). *Criminality and antisocial behavior in XXIX*: 953-962.
- [9] Rovet, J., Netley, C., Keenan, M., Bailey, J., Stewart, D. (1996). The psycho educational profile of boys with Klinefelter syndrome. *J. Learn Disabil.* 29 (II):180-96.
- [10] Sørensen, K. (1987). Klinefelter's Syndrome in Childhood, Adolescence & Youth: A genetic, clinical, developmental, psychiatric & psychological study of unselected men with sex chromosomal abnormalities. *Psychol. Med. Mayo Clin. Proc.* Elsevier.
- [11] Lanfranco, F., Kamischke, A., Zitzmann, M., Nieschlag, E. (2004). Klinefelter's syndrome. *Lancet* 364 :273-83.
- [12] Tartaglia, N., Ayari, N., Howell, S., D' Epagnier, C., Zeitler P (2011). 48, XXYY, 48,XXXY and 49, XXXXY syndromes: not just variants of Klinefelter syndrome. *Acta Paediatr.* 100 (VI): 851-860.
- [13] Wattendorf, D. J., Muenke, M. (2005). Klinefelter Syndrome. *Am. Fam. Physician.* 72 (XI): 2259-2262.
- [14] Tartaglia, N., Davis, S., Nimishakavi, S., Beauregard, R., Reynolds, A., et al. (2008). A new look at XXYY syndrome: medical and psychological features. *Am. J. Med. Gen.* 146 (XII): 1509-1522.
- [15] Wechsler, D. (1987). *Scala d'intelligenza Wechsler a livello prescolare e di scuola elementare*, Italian version. Organizzazioni Speciali, Firenze. [*Wechsler intelligence scale at preschool and elementary school level*, Italian version. Special Organizations, Florence].
- [16] Wechsler, D. (1997). *Scala d'Intelligenza Wechsler per Adulti-Riveduta (WAIS-R)* Italian version. Organizzazioni Speciali, Firenze. [*Wechsler Intelligence Scale for Adults-Revised (WAIS-R)* Italian version. Special Organizations, Florence].
- [17] Raven, J., Raven, J. C., Court, J. H. (2003, updated 2004). *Manual for Raven's Progressive Matrices and Vocabulary Scales*. San Antonio, TX.

- [18] De Renzi, E., Vignolo L. A., (1962). The token test: A sensitive test to detect receptive disturbances in aphasics. *Brain* 85 (IV): 665-678.
- [19] Sparrow, S. S., Balla, D. A., Cicchetti, D. V., Harrison, P. L., Doll, E. A. (1984). Vineland adaptive behavior scales (VABS). American Guidance. *Service Circle Pines*, MN.
- [20] Derogatis, L. R. (1994). *The Symptom Checklist administration, scoring, and procedures manual* (SCL-90-R). NCS Pearson, Minneapolis.
- [21] Rutter, M., Bailey, A., Catherine, L. (2003). The Social Behaviour Questionnaire (SCQ). *Western Psychological Services*, Los Angeles, CA.
- [22] Bem, S. L. (1971). Bem sex role inventory (BSRI) Early Career Award. *APA* (XXXII) 88-89.
- [23] Demily, C., Babinet, M.N., Rigard, C., Peyroux, É., Dragomir AR, et al., (2016) Social cognition disorder in Klinefelter Syndrome: a specific phenotype?. *GénoPsy*, Centre de dépistage et de prises en charge des troubles psychiatriques d'origine génétique.
- [24] Ross, J. (2007). *Klinefelter syndrome. Neurogenetic Developmental Disorders: Variation of Manifestation in Childhood*. MIT Press, Boston.
- [25] Snoeijs-Schouwenaars, F. M., Van Ool, J. S., Tan, I. Y., Schelhaas, H. J., Majole, M. H. (2017). Evaluation of perampanel in patients with intellectual disability and epilepsy. *Epilepsy and Behavior* 66: 64-67.
- [26] Tatum, W.O. 4th, Passaro, E. A., Elia, M., Guerrini, R., Gieron, M. et al., (1998). Seizures in Klinefelter's syndrome. *Pediatr Neurol* 19 (IV): 275-8.
- [27] Grosso, S., Farnetani, M. A., Di Bartolo, R. M., Berardi, R., Pucci, L. et al., (2004). Electroencephalographic and epileptic patterns in X chromosome anomalies. *J. Clin. Neurophysiol.* 21 (IV): 249-53.
- [28] Verri, A. P., Galimberti, C. A., Perucca, P., Cremante, A., Vernice, M. et al. (2008). Psychotic disorder and focal epilepsy in a left-handed patient with chromosome XYY abnormality. *Genet couns* 19 (IV): 373-9.
- [29] Samango-Sprouse, C. A., Rogol, A. (2002). XXY: the hidden disability and a prototype for an infantile presentation of developmental dyspraxia (IDD). *Infant Young Child* (XV):11-18.
- [30] Verri, A., Cremante, A., Clerici, F., Destefani, V., Radicioni, A. (2010). Klinefelter's syndrome and psychoneurologic function. *Mol. Hum. Reprod.* 16 (VI): 425-433.
- [31] Graham, J., Bashir, A. S., Stark, R. E., Silbert, A. (1988). Oral and written language abilities of XXY boys: Implications for anticipatory guidance. *Pediatrics* 81: 795-806.
- [32] Walzer, S., Bashir, A., Silbert, A. (1990). Cognitive and behavioral factors in the learning disabilities of XXY and XYY boys. *Birth Defects-Orig.* 26: 45-58.
- [33] Bender, B. G., Linden, M. G., Robinson, A. (1993). Neuropsychological impairment in 42 adolescents with sex chromosome abnormalities. *Am. J. Med. Genet.* 48 (III):169-73.
- [34] Vernice, M., Cremante, A., Clerici, F., Verri, A. P. (2014). Referential choice in the narratives of Italian speakers with Klinefelter Syndrome. *Int. j. speech lang. pathol. Audiol.* 2 (II), 81-85.

- [35] Gropman, A., Samango-Sprouse, C. A. (2013). Neurocognitive variance and neurological underpinnings of the X and Y chromosomal underpinnings of the X and Y chromosomal variations. *Am. J. Med. Genet.* (Part C) 163 C: 35-43.
- [36] Crow, T. J. (2004). Directional asymmetry is the key to the Origin of Modern Homo sapiens (the Broca-Annett axiom). Response to Lesley Rogers. *Laterality* 9: 233-242.
- [37] DeLisi, L. E., Maurizio, A. M., Svetina, C. (2005). Klinefelter's syndrome (XXY) as a genetic model for psychotic disorders. *Am. J. Med. Genet. part B* 135 (I):15-23.
- [38] Van Rijn, S., Aleman, A., Swaab, H., Vink, M., Sommer, I. (2008). Effects of an extra X chromosome on language lateralization: An fMRI study with Klinefelter men (47, XXY). *Schizophr Res* 101: 17-25.
- [39] DeLisi, L. E., Maurizio, A. M., Svetina, C. (2004). Schizophrenia and sex chromosome anomalies. *Schizophrenia Bull.* 20: 495-505.
- [40] Bruining, H., Swaab, H., de Sonnevile, L. M., Van Rijn, S., Van Engeland, H. (2011). *In search for significant cognitive features in Klinefelter syndrome through cross-species comparison of a supernumerary X chromosome.* G2B 10 (VI): 658-662.
- [41] Zitzmann, W. M., Werler, J., Steffi, H. N. (2014). New approaches to the Klinefelter Syndrome. *Annales d'endocrinologie* 75: 88-97.
- [42] Temple, C. M., Sanfilippo, P. M. (2003). Executive skills in Klinefelter's syndrome. *Neuropsychologia* 41(XI): 1547-59.
- [43] Steinman, K., Ross, J., Lai, S., Reiss, A., Hoeft, F. (2009). Structural and functional neuroimaging in Klinefelter (47, XXY) syndrome: a review of the literature and preliminary results from a functional magnetic resonance imaging study of language. *Ment Retard Dev D R* 15 (IV): 295-308.
- [44] Visootsak, J., Graham, J. M. (2006). Klinefelter syndrome and other sex chromosomal aneuploidies. *Orphanet Journal of Rare Diseases* 1: 42.
- [45] Weiten, W. (1997). *Psychology themes and variations.* Thomson/ Wadsworth.
- [46] Van Rijn, S., Swaab, H., Aleman, A., Kahn, R. S. (2006). X chromosomal effects on social cognitive processing and emotion regulation: A study with Klinefelter men (47, XXY). *Schizophr. Res.* 84: 1984-203.
- [47] Van Rijn, S., Stockman, L., Van Buggenhout, G., Swaab, H. (2014). Social cognition and underlying cognitive mechanisms in children with an extra X chromosome: a comparison with autism spectrum disorder. *Genes, Brain and Behavior* 459-467.
- [48] Girardin, C. M., Lemyere, E., Alos, N., Deal, C., Huot, C., et al., (2009). Comparison of Adolescents with Klinefelter Syndrome according to the Circumstances of diagnosis: amniocentesis versus Clinical Sings. *Horm. Res.* 72: 98-105.
- [49] Bonomi, M., Rochira, V., Pasquali, D., Balercia, G., Jannini, E.A. et al., (2017). Klinefelter Syndrome (KS): genetics, clinical phenotype and hpogonadism. *J. Endocrinol. Invest.* 40 (II): 123-134.
- [50] Jannini, E. A., Lenzi, A., Wagner, G. (2006). Behavioral therapy and counseling. *Andrology for the clinician.* Springer Berlin Heidelberg 598-607.

Chapter 79

TRANSCRANIAL DIRECT CURRENT STIMULATION (tDCS) AND COGNITIVE EMPOWERMENT FOR THE FUNCTIONAL RECOVERY OF DISEASES WITH CHRONIC IMPAIRMENT AND GENETIC ETIOPATHOGENESIS

***Antonio Gangemi¹, Tindara Capri¹,
Rosa Angela Fabio^{1,*}, Paola Puggioni²,
Alessandra Maria Falzone¹ and Gabriella Martino¹***

¹Cognitive Empowerment Laboratory, Department of Cognitive Science,
University of Messina, Messina, Italy

²Rehabilitation Department, Tor Vergata University of Rome, Rome

ABSTRACT

Transcranial direct current stimulation (tDCS) is a non-invasive, painless brain stimulation treatment that uses direct electrical currents of low intensity to stimulate specific parts of the brain. tDCS could both facilitate (anodic stimulation) and inhibit (catodic stimulation) specific areas of the brain (Ardolino, Bossi, Barbieri, & Priori, 2005), as many neurological and psychiatric disorders are linked to a hypofunction or hyperfunction of specific areas of the nervous system. Such phenomenon is based on two processes: rearrangement of functional neural circuits, and their reconstruction (Kaas, & Garraghty, 1989; Kandel, Schwartz, & Jessell, 2000).

In light of the studies mentioned above, it is assumed that tDCS can represent a useful tool to facilitate the process of neuroplasticity in subjects affected by chronic neurological diseases and genetic etiopathogenesis, such as Rett Syndrome (RTT). The aim of the present study is to examine the neurophysiological and cognitive effects of cognitive empowerment combined with tDCS in young girls and women with RTT, with chronic language impairments. Despite results in cognitive rehabilitation showing a positive trend,

* Corresponding Author's Email: rafabio@unime.it (Tel: +39090344831).

the efficacy of specific intervention on articulated speech is less consolidated. Lack of current research on successful outcomes in language production prompted the current study, which focuses more on intervention of articulated speech and cognitive functions.

In this chapter, we propose an integrated intervention: tDCS and cognitive empowerment applied to language in order to boost speech production (new functional sounds and new words). Given that maximal gains are usually achieved when tDCS is coupled with behavioural training, we applied tDCS stimulation on Broca's area together with linguistic training. Fourteen young girls and women with RTT were randomly allocated into two subgroups: AtDCS ($n = 7$) or placebo tDCS ($n = 7$). tDCS was applied over Broca's area for a 20-minute session for ten consecutive days. During tDCS stimulation, speech rehabilitation was divided into two sessions: production of vowels and word associations, and discrimination between corresponding words and images. Neurophysiological and cognitive parameters were measured at baseline, post-training and one month after intervention.

Results show a general enhancement in language, motor coordination and neurophysiological parameters in the AtDCS group compared to the placebo group. The present study provides evidence that tDCS combined with cognitive empowerment can improve language abilities and motor coordination and foster brain plasticity in young girls and women with RTT. Hence, this study supports the role of tDCS stimulation as a new methodology in the rehabilitation of diseases with chronic impairment and genetic etiopathogenesis.

Keywords: tDCS stimulation, cognitive empowerment, neuroplasticity, Rett Syndrome, rare genetic disorders

1. INTRODUCTION

The human genome contains an estimated total of 20,000-25,000 genes that serve as blueprints for building all of our proteins (International Human Genome Sequencing Consortium, 2004). In diseases caused by a single gene, a mutation in just one of these genes may be responsible for a disease. Single-gene diseases run in families and can be dominant or recessive, and autosomal or sex-linked (Chial, 2008).

Rett Syndrome (RTT) is a rare childhood developmental disorder, characterized by a primary disturbance in neuronal development. Females are primarily affected (1/10,000, Chahrour & Zoghbi, 2007), although a few cases of males have been reported in literature (Leonard et al., 2001; Cohen et al., 2002). Its aetiology involves the genetic mutation of gene MECP2 on the X-chromosome (Amir et al., 1999; Guy, Hendrich, Holmes, Martin, & Bird, 2011). Neurological abnormalities in RTT appear in several behavioural and cognitive impairments such as stereotypies, loss of speech and hand skills, gait apraxia, irregular breathing with hyperventilation while awake, and frequent seizures (Fabio, Billeci et al., 2016; Fabio, Cardile et al., 2017; Fabio, Castelli, Antonietti, & Marchetti, 2009; 2013).

However, the core of phenotype symptoms includes severe linguistic and motor impairments. With reference to the recovery of cognitive functions in RTT, recent investigations in which patients with RTT received intensive cognitive rehabilitation showed that they can go beyond the stage of pre-intentional level of development. Moreover, several studies using cognitive empowerment have provided evidence for positive neuroplasticity, as well as for transfer of benefits to untrained cognitive abilities (Anderson et al., 2013; Anguera,

White-Schwoch, Parbery-Clark, & Kraus 2013). In addition, results in cognitive rehabilitation generally showing a positive trend (Fabio, Capri, et al., 2017; Fabio, Colombo, et al., 2014; Fabio, Giannatiempo, Antonietti, Budden, 2009; Fabio, Giannatiempo, Oliva, Murdaca, 2011).

Within the landscape of studies on the recovery of the cerebral functions, after neurological lesions or disorders, neurostimulation techniques have recently been adopted. These techniques can influence specific parts of the brain by activation or inhibition, and their functionality. Transcranial direct current stimulation (tDCS) is a non-invasive, painless brain stimulation treatment that uses direct electrical currents of low intensity to stimulate specific parts of the brain. tDCS can both facilitate (anodic stimulation) and inhibit (cathodic stimulation) specific areas of the brain (Ardolino, Bossi, Barbieri, & Priori, 2005), as many neurological and psychiatric disorders are linked to a hypofunction or hyperfunction of specific areas of the nervous system. Such phenomenon is based on two processes: rearrangement of functional neural circuits, and their reconstruction (Kaas, & Garraghty, 1989; Kandel, Schwartz, & Jessell, 2000).

In a study aimed at exploring how tDCS influences language networks, Fertonani, Rosani, Cotelli, et al., (2010) found that anodal stimulation on the left dorsolateral prefrontal cortex improves naming performance and speeds up verbal reaction times, whereas cathodal stimulation had actually no effect. By administering an attentive task, they excluded non-specific effects due to a general increase in arousal. The authors concluded that left dorsolateral stimulation of the prefrontal cortex belonged to the cerebral network dedicated to lexical retrieval/selection processing in naming.

In a combined EEG–tDCS study (Wirth et al., 2017), the authors traced the effects of tDCS over the left dorsal prefrontal cortex, testing electrophysiological and behavioural variables during evident picture naming. The authors used the semantic blocking paradigm in which lexical-semantic competition increases when subjects have to name pictures of objects displayed in a semantically homogeneous context (i.e., cherries among grapes, pears and oranges) and decreases when the target object appears among semantically unrelated objects (heterogeneous blocks containing for example, cherries among flies, a cocktail and a bed). Anodal tDCS induced modulations of behavioural and electrophysiological data. The authors concluded that electrophysiological variables could help to understand how prefrontal anodal tDCS influences language production.

Based on these insights, we assume that tDCS can represent a useful tool to facilitate the process of neuroplasticity in subjects affected by chronic neurological diseases and genetic etiopathogenesis, such as RTT. In this chapter, we describe one of our studies in which we used an integrated intervention: tDCS and cognitive empowerment applied to language in order to boost speech production (new functional sounds and new words). Given that maximal gains are usually achieved when tDCS is coupled with behavioural training (Reis, Schambra, Cohen, Buch, Fritsch, Zarah, et al., 2009; Vannorsdall, Schretlen, Andrejczuk, Ledoux, Bosley, Weaver, et al., 2012), we applied tDCS stimulation on Broca's area together with cognitive training. We examined the cognitive and neurophysiological effects of tDCS comparing two groups: one in which the participants with RTT received anodal stimulation and another in which they received placebo stimulation. We hypothesized that tDCS combined with cognitive empowerment can induce neurophysiological and cognitive effects. We predicted that tDCS combined with cognitive training would positively influence language skills and the power beta and alpha band activity in the participants receiving anodal stimulation compared to the placebo group.

2. CURRENT STUDY

2.1. Methods

2.1.1. Participants

Fourteen young girls and women with RTT were identified through specific inclusion and exclusion criteria and were invited to participate in this study. The participants aged from 10 to 40 years old (mean = 18.34 years, SD = 5.36) and were in a chronic phase with stable language, but poor speech production and no changes after the third stage. Their families had been contacted by the Italian Rett Association and Italian Union Rett Onlus. Written informed consent was obtained from parents prior to the experiments. The experiment was approved by the ethics committee of the Department of Cognitive Science, University of Messina.

2.2. Procedure

The design of this study was ABAA: pre-test assessment, AtDCS or placebo tDCS and cognitive empowerment (speech rehabilitation), post-test assessment (ten days after the experiment) and follow-up assessment (one month after the experiment). In the pre-test phase, all participants underwent a neurological (EEG measure) and neuropsychological assessment to evaluate behavioural measures and linguistic pre-requisites prior to beginning the experiment. The scores obtained in the pre-test phase were compared with those observed in the post-treatment assessment phase and follow-up, to evaluate the effects of integrated intervention. In the neuropsychological assessment, the Vineland Adaptive Behavior Scales-Interview second edition (VABS) (Sparrow, Balla & Cicchetti, 1984), the Rett Assessment Rating Scales (RARS) (Fabio, Martinazzoli, & Antonietti, 2005), Fanzago's test (1983) and Raven's Progressive Matrices were used. VABS is a standardized semi-structured interview instrument that assesses day-to-day adaptive functioning and is administered to primary caregivers by fully trained research assistants in Vineland interview and scoring procedures. It consists of four domains: Communication (Receptive, Expressive, Written); Daily Living (Personal, Domestic, Community); Socialization (Interpersonal Relationships, Play and Leisure Time, Coping Skills); and Motor Skills (Gross, Fine). RARS is used to evaluate symptom severity in girls with Rett syndrome (Fabio, Martinazzoli, & Antonietti, 2005). The items in RARS were formed following the diagnostic criteria for RTT proposed by DSM-IV-TR (DSM-IV-TR; APA, 2000) and recent research. The Italian Fanzago phonetic articulation test (1983) evaluates the status of vocal sound articulation and objective production of articulated voice functional to communication. This instrument is based on spontaneous/repetition elicited from a denomination of 114 Figures grouped in 22 tables.

For cognitive measures, Modified Raven's Coloured Progressive Matrices were used (Antonietti et al., 2003). This test is made up of four series (A, B, C, D) of increasing complexity, with each series including 12 items (incomplete figures). The individual must complete an abstract figure choosing among six alternatives.

EEG data were acquired using a gold-standard digital EEG amplifier (Cardinal Medical System) of 21 electrodes placed on the scalp of the participant according to the expected parameters of the international measuring system 10/20. Quantitative analysis was performed

by tailor-made algorithms developed in Matlab code. The power spectral density (PSD) was evaluated by transforming the signal from the time domain to the frequency domain using the Welch method (Welch, 1967). A spectral analysis of EEG rhythms and the diffusion of the effect of the tDCS on all channels of registration were assessed. Neurophysiological and neuropsychological assessment was assessed at baseline (pre-test), ten days after the intervention and at one month after treatment.

2.3. tDCS Stimulation Protocol

Participants were randomly allocated into two subgroups: AtDCS ($n = 7$) or placebo tDCS ($n = 7$). In the AtDCS group, the participants received the integrated intervention anodal tDCS plus cognitive empowerment during the treatment phase. In the placebo tDCS group, participants received placebo stimulation plus cognitive empowerment. tDCS was administered at an intensity of two mA, using a stimulator connected to a pair of electrodes connected to a constant current stimulator (HDC type Company Stimulation Omicron T) isolated from the power supply as it has its own power supply (low voltage battery). The active electrode was placed on Broca's area, while the reference electrode was applied in the contralateral cephalic site. Speech rehabilitation was studied in two sessions: production of vowels and word associations, and discrimination between corresponding words and images.

2.4. Results

2.4.1. Comparative Analyses

The results indicate a steadily increasing trend in the production of new vowel/consonant sounds and words in the AtDCS group, as shown in Figures 1, 2, 3 and 4.

With reference to the motor coordination parameters, Figure 4 also shows a significant improvement of performance in the discrimination between corresponding words and images session. This suggests that tDCS combined with cognitive empowerment produces multiple effects: the plasticity of cortical structures has probably determined the cytoarchitecture reorganization of sub-cortical structures, such as the basal ganglia. It is well known that the basal ganglia form the extrapyramidal system which is involved in the control of complex motor actions, such as hand movement and phonetic production.

2.4.2. Statistical Analyses

Data were analysed using SPSS Version 14.0 for Windows. Descriptive statistics of the dependent variables were tabulated and examined. Alpha level was set to 0.05 for all statistical tests. In the case of significant effects, the effect size of the test was reported. ANOVA 2 (groups: atDCS vs placebo tDCS) $\times$ 4 (phases: pre-test, training, post-test and follow-up) was carried out.

With reference to groups, we found no significant effect, $F(3.25) = 4.84$; $p < .08$. There was significant groups $\times$ phases interaction, $F(3.27) = 19.21$; $p < .001$. This result indicates that the two groups show a different trend, as shown in Figure 1. In particular, the AtDCS group

shows a clear improvement in performance, whereas the placebo tDCS group presents a decrease in performance.

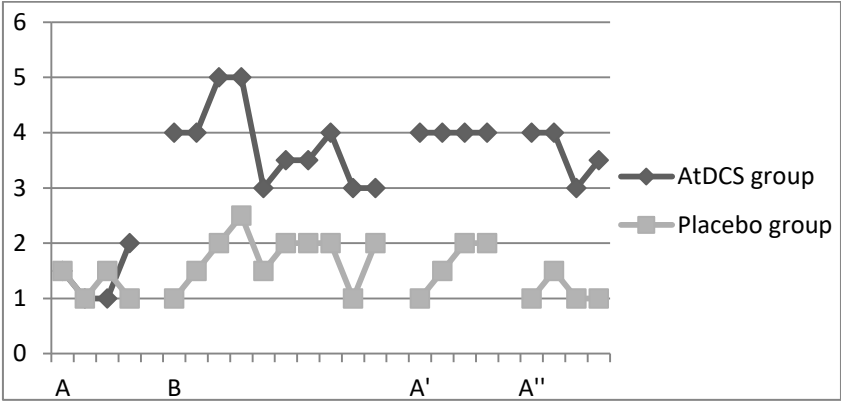


Figure 1. Number of vowels with elicited denomination for the two groups in the pre-test (A), training (B), post-test (A') and follow-up (A'') phases.

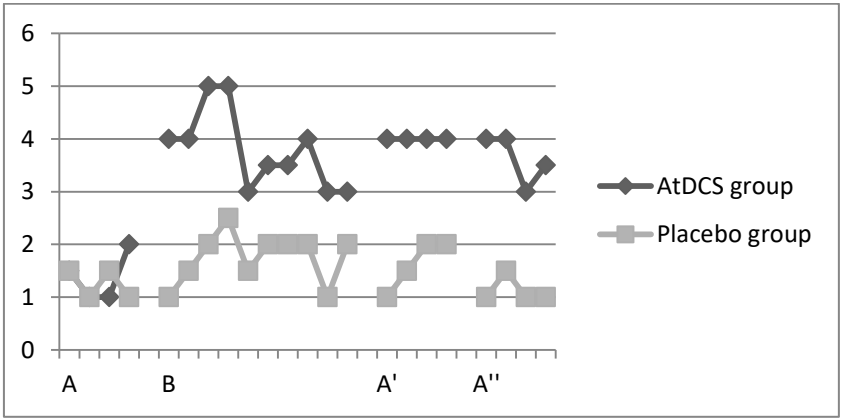


Figure 2. Number of consonants with elicited denomination for the two groups in the pre-test (A), training (B), post-test (A') and follow-up (A'') phases.

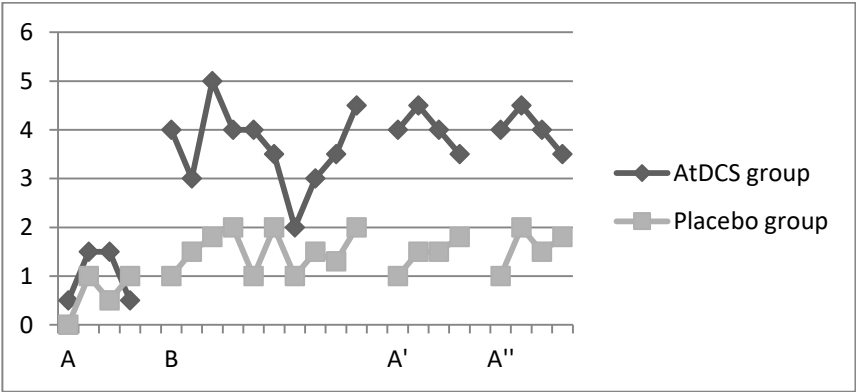


Figure 3. Number of vowels with elicited denomination for the two groups in the pre-test (A), training (B), post-test (A') and follow-up (A'') phases.

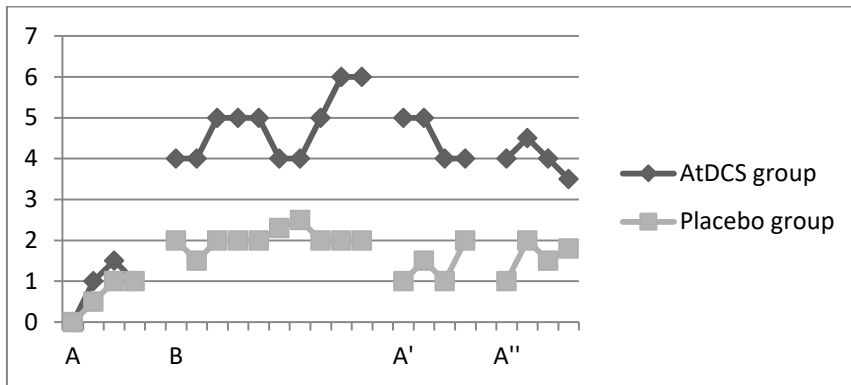


Figure 4. Trend of two groups related to the motor abilities in the pre-test (A), post-test (A') and follow-up (A'') phases.

2.4.3. Production of Vowels

Table 1 shows the means and standard deviations of the number of vowels with elicited denomination. The groups x phases interaction showed significant effects, $F(3,22) = 19.18$; $p < .001$. As shown in Figure 5, this result indicates that the AtDCS group shows a higher number of vowels than the placebo tDCS group. This trend remains stable in the post-test and follow-up phases.

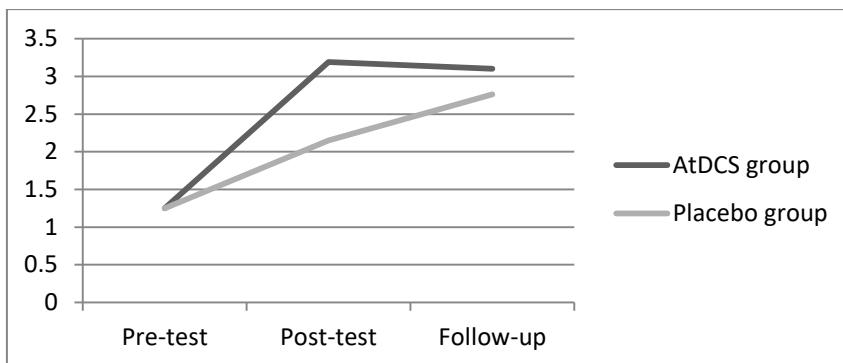


Figure 5. Trend of two groups related to the production of vowels in the pre-test, post-test and follow-up phases.

Table 1. Means and standard deviation (SD) related to the production of vowels in the three phases

Groups	Pre-test	Post-test	Follow-up
AtDCS	1.25 (0,33)	3.19 (0,88)	3.1 (0,99)
Placebo tDCS	1.25 (0)	2.15 (0)	2.75 (0)

2.4.4. Production of Consonants

Table 2 shows the means and standard deviations of the number of consonants with elicited denomination. The groups x phases interaction shows significant effects, $F(3,27) = 19.21$; p

<.001. As shown in Figure 6, this result indicates that the AtDCS group shows a higher number of consonants compared to the placebo tDCS group. This trend remains stable in the post-test and follow-up phases.

Table 2. Means and standard deviation (SD) related to the production of consonants in the three phases

Groups	Pre-test	Post-test	Follow-up
AtDCS	1.5 (1,01)	3.15 (1,24)	3.4 (1.64)
Placebo tDCS	1.58 (0)	2.7 (0)	2.5 (0)

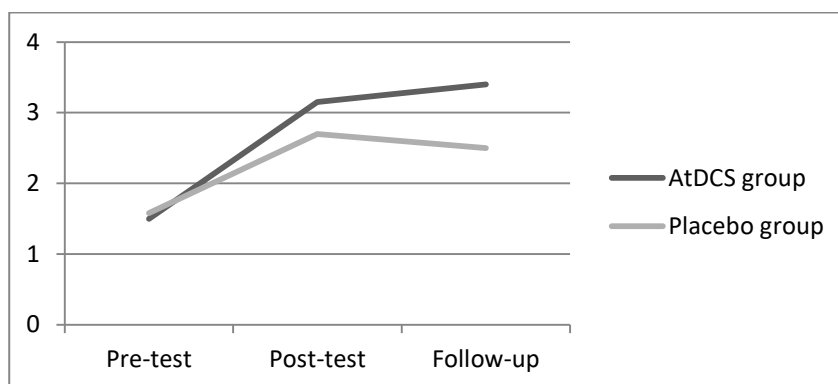


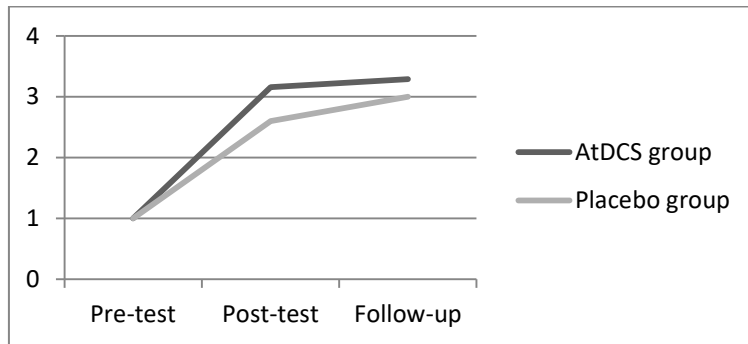
Figure 6. Trend of two groups related to the production of consonants in the pre-test, post-test and follow-up phases.

2.4.5. Production of Words

Table 3 shows the means and standard deviations of the number of words with elicited denomination. The groups x phases interaction shows significant effects, $F(3,27) = 2.95$; $p < .05$. As shown in Figure 7, this result indicates that the AtDCS group shows a higher number of words than the placebo tDCS group. This trend remains stable in post-test and follow-up phase.

Table 3. Means and standard deviation (SD) related to the production of words in the three phases

Groups	Pre-test	Post-test	Follow-up
AtDCS	1 (0.46)	3.16 (0.85)	3.29(1.16)
Placebo tDCS	1 (0)	2.6 (0)	3 (0)



Figures 7. Trend of two groups related to the production of words in the pre-test, post-test and follow-up phases.

2.4.6. Quantitative EEG Analysis

Quantitative analysis (QA) was performed by tailor made algorithms developed in Matlab code. The power spectral density (PSD) was evaluated by transforming the signal from the time domain to the frequency domain using the Welch method (Welch, 1967). PSDs were calculated for each epoch and were averaged. To begin with, the absolute total power of the signal and the absolute power of each band was calculated for each electrode. The considered bands were: theta (4 – 7 Hz), alpha (8 – 13 Hz) and beta (14 – 29 Hz).

As shown in Table 4, QA presents a significant rise in the frequency and power of alpha and beta bands and a decrease in pathological bands, such as theta.

Table 4. Quantitative analysis of rhythmic theta, beta and alpha related to pre-test, post-test and follow-up phases for the two groups

	Frequency bands	Pre-test (SD)	Post-test (SD)	Follow-up (SD)
AtDCS group	Theta (4 – 7 Hz)	6.1 (0.90)	7.9 (0.79)	7.2 (0.79)
	Beta (14 – 29 Hz)	14.2 (0.60)	21 (0.30)	19.8 (0.38)
	Alpha (8 – 13 Hz)	8.1 (0.75)	9 (0.73)	8.8 (0.73)
Placebo tDCS group	Theta (4 – 7 Hz)	6.3 (0.90)	6.5 (0.80) 14.1	6.2 (0.80)
	Beta (14 – 29 Hz)	13.9 (0.61)	(0.50)	15.3 (0.35)
	Alpha (8 – 13 Hz)	8.3 (0.75)	8 (0.73)	8 (0.73)

CONCLUSION

The aim of the present chapter was to examine the neurophysiological and cognitive effects of cognitive empowerment combined with tDCS in young girls and women with RTT with chronic language impairments. Here, we report our experimental study in which we adopted tDCS combined with cognitive empowerment in participants with RTT. The results of this study show a general enhancement in training abilities and neurophysiological parameters in the AtDCS group compared to the placebo group. Therefore, we can assume that cognitive empowerment combined with tDCS stimulation promotes neuroplasticity and facilitates the recovery of impaired functions in subjects with this rare genetic disease.

The ability of tDCS stimulation to induce changes in cortical function has been examined in a wide range of patients. Promising results have been reported with this approach in the treatment of neurological disorders, such as aphasia, Parkinson's, MCI, Alzheimer's disease, and subjects with focal brain injury, such as stroke (Martin et al., 2004; Falzone, Gangemi, & Fabio, 2015; Fabio, Gangemi, Caprì, Budden, & Falzone, in press; Fabio, Caprì, T., Lotan, Towey, & Martino, 2017). Hence, tDCS used alone or in combination with other approaches, like cognitive training, induces or modulates neuroplastic responses, demonstrates causal relations between brain and behaviour, translates into behavioural modification, and even leads to new therapeutic interventions (Bartrés-Faz & Vidal-Piñeiro, 2016). The results of the present study are in line with the research mentioned above. We propose that the efficacy of tDCS is due to the excitatory effect of the impulse on Broca's area which induces long-term enhancement (LPT) favouring the mechanisms of synaptic plasticity.

In conclusion, the results of the current study suggest that tDCS combined with cognitive training is a promising way to intervene and improve behaviours and brain activation in patients with RTT, and suggests the application of this kind of treatment in clinical practice.

REFERENCES

- Amir, R. E., Van den Veyver, I. B., Wan, M., Tran, C. Q., Francke, U., Zoghbi, H. Y. (1999). Rett syndrome is caused by mutations in x-linked MECP2, encoding methyl-CpG-binding protein 2. *Nat Gen*, 23, 185–188.
- Anderson, S., White-Schwoch, T., Parbery-Clark, A., & Kraus, N. (2013). Reversal of age related neural timing delays with training. *Proc Natl Acad Sci USA*, 110, 4357–4362.
- Anguera, J. A., Boccanfuso, J., Rintoul, J. L., Al-Hashimi, O., Faraji, F., Janowich, J., et al. (2013). Video game training enhances cognitive control in older adults. *Nature*, 501, 97–101.
- Antonietti, A., Castelli, I., Fabio, R. A., & Marchetti, A. (2003). *La sindrome di Rett. Prospettive e strumenti per l'intervento*. Roma: Carocci.
- Ardolino, G., Bossi, B., Barbieri, S., & Priori, A., (2005). Non-synaptic mechanisms underlie the after-effects of cathodal transcutaneous direct current stimulation of the human brain. *J. Physiol*. 568, 653–663.
- Baker, J. M., Rorden, C., & Fridriksson, J. (2010). Using transcranial direct-current stimulation to treat stroke patients with aphasia. *Stroke*, 41, 1229–1236.
- Bartrés-Faz, D., Vidal-Piñeiro, D. (2016). Noninvasive brain stimulation for the study of memory enhancement in aging. *Eur Psychol*, 21, 41–54.
- Buschkuhl, M., Jaeggi, S. M., Hutchison, S., et al., (2008). Impact of working memory training on memory performance in old-old adults. *Psychol Aging*, 23, 743–753.
- Castelli, I., Antonietti, A., Fabio, R. A., Lucchini, B., Marchetti, A. (2013). Rett syndrome girls possess Theory of Mind. *Life Span and Disability XVI*, 2, 157–168.
- Chahrour, M., & Zoghbi, H. Y. (2007). The story of Rett syndrome: from clinic to neurobiology. *Neuron*, 56, 422–437.
- Chial, H. (2008) Rare Genetic Disorders: Learning About Genetic Disease Through Gene Mapping, SNPs, and Microarray Data. *Nature Education*, 1(1), 192

- Fabio, R. A., Martinazzoli, C., & Antonietti, A. (2005). Costruzione e standardizzazione dello strumento "R.A.R.S." (Rett Assessment Rating Scale). *Ciclo Evoltivo & Disabilità*, 8, 257-381.
- Fabio, R. A., Cardile, S., Troise, E., Polimeni, S., Germanò E., Di Rosa, G., Siracusano, R., Nicotera, A., Gagliano, A., Tortorella, G. (2017). Cognitive empowerment with new technologies improves neuropsychological and Neurophysiological parameters in Rett syndrome. In *Horizons in Neuroscience Research*. New York: Nova Science Publisher.
- Fabio, R. A., Castelli, I., Antonietti, A., Marchetti, A. (2013). Training communication abilities in Rett Syndrome through reading and writing. *Frontiers in Psychology for Clinical Settings*, 4, 911, 1-9. doi: 10.3389/fpsyg.2013.00911.
- Fabio, R. A., Antonietti, A., Marchetti, A., Castelli, I. (2009). Attention and communication in Rett syndrome. *Research in Autism Spectrum Disorders*, 3, 329–335.
- Fabio, R. A., Billeci, L., Crifaci, G., Troise, E., Tortorella, G., Pioggia, G. (2016). Cognitive training modifies frequency EEG bands and neuropsychological measures in Rett syndrome, *Research in Developmental Disabilities*, 54, 73-85.
- Fabio, R. A., Capri, T., Lotan, M., Towey, G. E., Martino, G. (2017). Motor abilities are related to the specific genotype in Rett Syndrome. In: *Advances in Genetic Research*, vol. 14, chapter id 32719. New York: Nova Science Publisher.
- Fabio, R. A., Colombo, B., Russo, S., Cogliati, S. F, Masciandri M., Antonietti, A., Tavian, D. (2014), Recent Insights into Genotype-Phenotype Relationships in Patients with Rett Syndrome Using a Fine Grain Scale, *Research in Developmental Disabilities*, 35(11): 2976–2986.
- Fabio, R. A., Gangemi, A., Capri, T., S. Budden, Falzone, A. M. (in press). Neurophysiological and cognitive effects of Transcranial Direct Current Stimulation in a girl with Rett Syndrome with chronic language impairments. *Research in Intellectual and Developmental Disabilities*.
- Fabio, R. A., Giannatiempo, S., Antonietti, A., Budden, S., (2009). The role of stereotypies in over selectivity processes in Rett Syndrome. *Research in Developmental Disabilities*, 30, 136–145.
- Fabio, R. A., Giannatiempo, S., Oliva, P., Murdaca, A. M. (2011). The increase of attention in Rett Syndrome. A pretest-post-test research design. *Journal of Developmental and Physical Disability*, 23, 99-111.
- Falzone, A. M., Gangemi, A., Fabio, R. A. (2015). Genotype-phenotype Relationships in Language Processes in Rett Syndrome. In *Advances in Genetics Research*, 14, Nova Publisher.
- Fanzago, F. (1983). Test di valutazione dell'articolazione. [Joint assessment test]. *Quaderni di ActaPhoniatria Latina*, 2, 80-85.
- Fox, K., Henley, J. & Isaac, J. (1999). Experience-dependent development of NMDA receptor transmission. *Nature Neuroscience*, 2, 297-9.
- Gangemi, A., Fabio, R. A., Capri, T., & Falzone, A. M. (2017). Neuroplasticity and Cognitive Empowerment: Relationships and New Methodic Directions. In *Horizons in Neuroscience Research*. New York: Nova Science Publisher.
- Guy, J., Cheval, H., Selfridge, J., Bird, A. (2011). The role of MeCP2 in the brain. *Annu Rev Cell Dev Biol*, 10, 631–652.
- Heugtena, C. M., Ponds, R. W. H. M., & Kessels, R. P. C. (2016). Brain training: hype or hope? *Neuropsychological Rehabilitation*, 26, 639-644.

- Kandel, E. R., Schwartz, J. H., & Jessell, T. M. (2000). *Principles of neural science* (4th ed.). New York: McGraw-Hill, Health Professions Division.
- Kleim, J., Kleim, E., & Cramer, S. (2007). Systematic assessment of training-induced changes in corticospinal output to hand using frameless stereotaxic transcranial magnetic stimulation. *Nature Protocols*, 2, 1675–84.
- Leonard, H., Silberstein, J., Falk, R., Houwink-Manville, I., Ellaway, C., Raffaele, L. S., et al., (2001). Occurrence of Rett syndrome in boys. *J Child Neurol*, 5, 333–338.
- Makin, S. (2016). Brain training: Memory games. *Nature*, 531, S10–S11.
- Martin P, Naeser M, Theoret H, Tormos J, Nicholas M, Kurland J, et al., (2004). Transcranial magnetic stimulation as a complementary treatment for aphasia. *Semin Speech Lang*, 25, 181–91.
- Merzenich, M. M, Kaas, J. H., Wall, J., Nelson, R. J., Sur, M., & Felleman, D. (1983). Topographic reorganization of somatosensory cortical areas 3b and 1 in adult monkeys following restricted deafferentation. *Neuroscience*, 8, 33–55.
- Reijnders, J., Van Heugten, C., & van Boxtel, M. (2013). Cognitive interventions in healthy older adults and people with mild cognitive impairment: A systematic review. *Ageing Research Reviews*, 12, 263–275.
- Reis, J., Schambra, H. M., Cohen, L. G., Buch, E. R., Fritsch, B., Zarahn, E., et al., (2009). Noninvasive cortical stimulation enhances motor skill acquisition over multiple days through an effect on consolidation. *Proceedings of the National Academy of Sciences of the United States*, 106, 1590–1595.
- Reus, G. Z., Stringari, R. B., Ribeiro, K. F., Ferraro, A. K., Vitto, M. F., Cesconetto, P., et al., (2011). Ketamine plus imipramine treatment induces antidepressant-like behavior and increases CREB and BDNF protein levels and PKA and PKC phosphorylation in rat brain. *Behav Brain Res*, 221, 166–171.
- Sparrow, S., Balla, D., & Cicchetti, D. V. (1984). *The Vineland Adaptive Behavior Scales* (Survey Form). Circle Pines, MN: American Guidance Service.
- Vignoli, A., Fabio, R. A., La Briola, F., Giannatiempo, S., Antonietti, A, Maggiolini, S., Canevini, M. P. (2010). Correlations between neurophysiological, behavioral, and cognitive function in Rett syndrome. *Epilepsy & Behavior*, 17, 489–496.
- Vines, B. W., Cerruti, C., et al., (2000). Dual-hemisphere tDCS facilitates greater improvements for healthy subjects' non-dominant hand compared to uni-hemisphere stimulation. *BMC*.
- Wirth J. H., Bernstein M. J., Wesselmann E. D., LeRoy A. S. (2017). Social cues establish expectation of rejection and affect the response to being rejected. *Group Process. Intergroup Relat.* 20, 32–51.

Chapter 80

CONTROL OF FORCE AND TIMING DURING UNIMANUAL AND BIMANUAL TAPPING MOVEMENTS OF ADOLESCENTS WITH DOWN SYNDROME

Nobuyuki Inui* and Junya Masumoto

Naruto University of Education, Japan

ABSTRACT

The present study examined the control of force and timing during finger tapping sequences of adolescents with Down syndrome. Data were obtained from two groups. An experimental group was composed of nine male adolescents with Down syndrome (15–17 years old). Two participants were moderate in intelligence level while other seven were severely intellectually disabled. A comparison group consisted of nine male high school students (16–17 years old). Participants performed both unimanual and bimanual tapping tasks with one self-paced test trial after three audible-synchronized practice trials with concurrent feedback of force output. All tasks consisted of a target force of 2N and a target intertap interval of 500 ms. Adolescents with Down syndrome exhibited a greater magnitude of positive constant error and variable error for peak force than typical adolescents. They also exhibited a greater magnitude of negative constant error and variable error for intertap interval than typical adolescents. Although normally developing comparison adolescents exhibited a linear relationship between peak force and press duration or time-to-peak force, the relationship was not familiar to adolescents with Down's syndrome. This may suggest differences in the manner of motor unit recruitment between the group with Down's syndrome and comparison adolescents. On the other hand, there was no difference between unimanual and bimanual tasks for variable error of intertap interval in adolescents with Down's syndrome. Because people with Down syndrome have exhibited a thinner corpus callosum than typical people, they may be unable to combine the output of two separate timing systems.

* Corresponding Author's Email: inui@naruto-u.ac.jp (Professor N. Inui, Laboratory of Human Motor Control, Naruto University of Education, Takashima, Naruto-cho, Naruto-shi, 772-8502, Japan; Tel: (+81)88-687-6517; Fax: (+81)88-687-6028).

INTRODUCTION

A large body of research has found that people with Down syndrome move slower, less coordinated, and less accurately than the typical population (Anson and Mawston, 2000; Inui et al. 1995; Latash, 2000; Robertson et al. 2002). In particular, people with Down syndrome take longer reaction time to initiate a response to a stimulus (for review, Anson, 1992, Inui, 2007) and longer movement time to complete a motor task (for review, Latash, 1992).

The origin of perceptual-motor dysfunction in Down syndrome has been attributed to both structural and functional differences in the central nervous system. Structurally, individuals with Down syndrome have reduced volume in the cerebellum, hippocampus, and cerebral gray matter and white matter of the frontal cortex compared with age-matched non-Down syndrome peers (Teipel et al. 2004). Cerebellar atrophy has been suggested to be the most prominent structural difference underlying neuromotor impairment in Down syndrome (Latash, 2000).

In a simple reaction time setup (Anson and Mawston, 2000), whereas the electromyogram patterns for anterior deltoid and extensor indicis of the typical participants are “steplike,” those of participants with Down syndrome typically show a “ramplike” change in activation from baseline. In addition, whereas reaction times in typical participants are most often synchronized with the initial burst of muscle activity, those in individuals with Down syndrome often occurred after the initial burst of muscle activation. These findings have suggested differences between typical and Down’s syndrome persons for neuro-muscular system, being thought to be the reason that Down syndrome produces slower simple reaction times than in typical individuals.

Henderson et al. (1981) found that while children with Down syndrome had no difficulty with the spatial components of tracking and drawing tasks, they were impaired on the temporal aspects of tracking. Planinsek (1996) also observed that children with Down syndrome were slow in timing the grasp of the ball in simple catching tasks. Thus, individuals with Down syndrome may be unable to process and utilize predictable information and control timing of the onset and offset of muscular force. In particular, because individuals with Down syndrome are more dependent on response-produced feedback, they appear to be unable to acquire feedforward strategies with practice (Elliott, 1990; Elliott et al. 2010).

Charlton et al. (1996) and Vimercati et al. (2013) also reported that children or adults with Down syndrome relied on feedback control during a reaching or tapping task. Charlton et al. (1996) examined the kinematic characteristics of reaching to grasp in children with Down syndrome (mean age: 9 years) compared with both chronological age-matched and mental age-matched groups. Movement time, peak velocity of the wrist and number of movement units were analyzed. Number of movement units was derived from velocity profiles, and each movement unit consisted of a period of acceleration and deceleration. The results showed that children with Down syndrome moved slower and with reduced peak velocity, as found in previous studies (Latash, 2000). In addition, trajectories of children with Down syndrome exhibited greater irregularities and a greater number of movement units. The preprogramming part of the movement is spatially inaccurate for children with Down syndrome, causing the need for successive corrective movements and greater reliance on feedback. Charlton et al. (1996) thus pointed out that because feedback-based corrections take time, larger number of movement units was associated with longer durations of deceleration phase for children with Down syndrome. Vimercati et al. (2013) further examined movement strategies of adult participants

with Down syndrome and of age-matched controls during an arm tapping task. Participants with Down syndrome relied on feedback control.

On the other hand, many studies reported that persons with Down syndrome performed unimanual discrete movements (e.g., flip a switch) more accurately in visual instructions than in verbal instructions (Hartley, 1982; Elliot et al. 1987), proposing a model of atypical cerebral specialization (Elliot and Weeks, 1993). Using bimanual tapping movements, Elliott et al. (1986) reported that participants with Down syndrome were less lateralized for sequential processing although they performed more slowly than did comparison participants. However, some studies found that adults with Down syndrome performed bimanual continuous movements more accurately in auditory-motor, than verbal-motor (Robertson et al. 2002), or visual-motor (Ringenbach et al. 2002) situations. On a bimanual tapping task of a 3:2 polyrhythm, Inui and Asama (2003a, 2003b) found that because strong negative correlations between-hand taps were observed for adolescents with mental retardation, and their slow-hand movements were further subordinate to the movements of the fast hand, they adopted a hierarchical integrated organization. Although adolescents with autism or Down syndrome also participated in this or preliminary experiment, they were unable to meet the criteria in the test (recall) trials, and thus their data were not acquired.

There may be differences between unimanual and bimanual movements for movement organization in persons with Down syndrome. In addition, it can be anticipated that there are differences between typical and Down's syndrome persons for press duration and time-peak force in finger tapping sequences from the results of Anson and Mawston (2000). However, previous studies did not focus on the control of force and timing in tapping movements in the population with Down syndrome. Therefore, the present study examined the control of force and timing in both unimanual and bimanual tapping tasks to possibly reveal a different strategy for force control in adolescents with Down syndrome compared with typical adolescents (Masumoto et al. 2012).

METHOD

Participants

Data were obtained from two groups. An experimental group was composed of nine male adolescents with Down syndrome (15–17 years old) recruited from a high school for disabled children. Although their intelligence quotients could not be measured accurately, two participants were moderate in intelligence level while other seven were severe. A comparison group consisted of nine male high school students (16–17 years old). Handedness was tested using the Edinburgh Handedness Inventory (Oldfield, 1971). The laterality quotients of right-handed participants were +100 overall. Informed consent for participation in the experiment was obtained by parents of every participant. The procedures were approved by the ethics committee at Naruto University of Education, and the study was conducted according to the Declaration of Helsinki.

Apparatus and Measurements

Figure 1A showed that the outputs of the two load cells (Model LUB-5KB, Kyowa Electronic Instruments, Co., Tokyo, Japan; rated load 5 kg) used for finger tapping were amplified by a strain amplifier (Kyowa Model MCC-8A) and displayed on an oscilloscope (Model MD625BM-12, Leader Electronic Corp., Yokohama, Japan). The force output was also recorded by a personal computer (Apple PowerBook G4) and monitored on a screen (832 x 624 pixel resolution) after the amplified signal was converted from analogue to digital (PowerLab/8sp, AD Instruments, Dunedin, NZ). Data were sampled at a frequency of 1000 Hz by a 16 bit A/D converter with a low-pass filter of 100 Hz. In the task of finger tapping (Figure 1B), the peak force and intertap interval during each trial were measured using software for analysis of peak force, interval, press duration and time-to-peak force (Emile Soft Co., Ltd., Tokushima, Japan). The peak force of each tap was defined as the peak output voltage from the load cell. The intertap interval was defined as the onset-to-onset times of the tap. Press duration was defined as the time that a participant's finger was in contact with the load cell. Time-to-peak force was defined as the time to reach peak force. Press duration and time-to-peak force have been regarded as representative parameters of the time spent to accumulate force in a tapping movement (Piek et al. 1993).

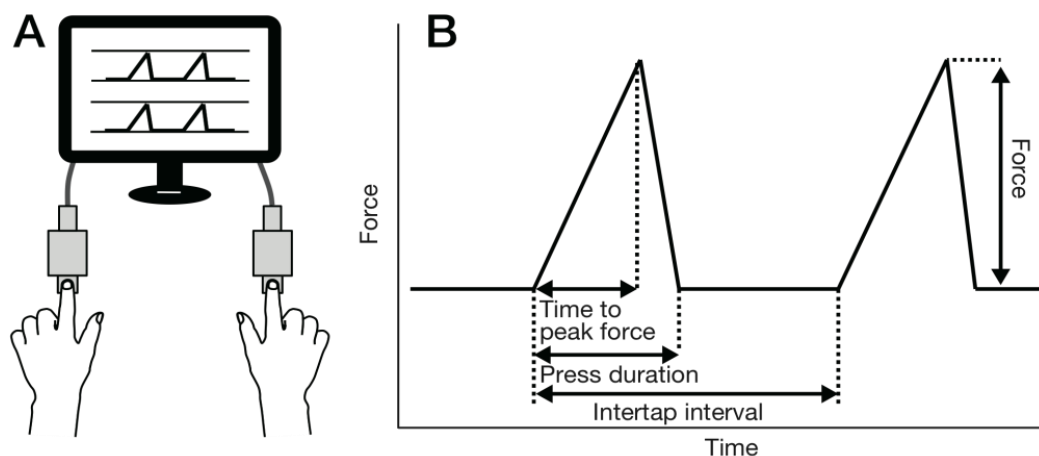


Figure 1. Experimental setup in the task of bimanual tapping movement (A), and the definition and measurement of dependent variables (B).

Procedure

Participants were seated facing the two load cells, their palms resting on a support surface 6 cm in height from a table. In this posture, participants could make tapping movements by means of an extension-flexion pulse of the index finger at the metacarpophalangeal joint. All participants performed both unimanual and bimanual finger-tapping tasks, which consisted of producing a target force of 2N at a prescribed intertap interval of 500 ms. In a preliminary experiment of our previous studies (Inui and Asama, 2003a, 2003b), because a bimanual tapping task of a 3:2 polyrhythm was too difficult for adolescents with Down syndrome, they were unable to perform of the task. The present study thus adopted the task of bimanual finger-

tapping sequences. Half of the participants performed first the unimanual task whereas the other half performed first the bimanual task. Similarly, half of the participants performed first the unimanual task with the right hand whereas the remainder performed first the task with the left hand. In the unimanual tasks, they were instructed to match the target force at a prescribed intertap interval with either the right or the left hand, whereas in the bimanual task they had to match the target force and prescribed intertap interval with both hands simultaneously. Participants practiced each task separately, with the corresponding test trial immediately following three practice trials of 30 s each. During practice trials, the tapping rate was prescribed by means of an audible metronome (Model SQ100-88, Seiko Holdings Corp., Tokyo, Japan). The participants were instructed to synchronize finger taps on the one or two load cells with the metronome. The output of the load cell was displayed on an oscilloscope so that the participant could see the difference between the peak force produced and the target force, which was indicated on the oscilloscope by one or two horizontal lines. In the test trial immediately after the practice trials, participants tapped for 30 s only one time. They were instructed to produce the force and intertap interval acquired during practice by means of self-paced movement without feedback.

Statistical Analysis

In the analyses of the test trial, the dependent measures were constant error, variable error, press duration and time-to-peak force corresponding to the separate intertap interval and peak force produced. The constant error retained the sign of each error (the difference between the target and realized force or interval) when the average was calculated to produce an arithmetic mean error. The variable error was calculated as the standard deviation around the mean constant error for each participant. These values were calculated from 60 measures produced by each participant in each trial. In the analyses of the practice trials, on the other hand, data from the final trial were analyzed. A 2 (experimental vs. comparison group) $\times$ 2 (unimanual vs. bimanual task) $\times$ 2 (practice vs. test trial) $\times$ 2 (right vs. left hand) analysis of variance (ANOVA) was performed to examine the main effect of group, task, trial, and hand on the dependent measures. When significant overall condition effects were found for a dependent measure, post-hoc multiple comparisons were corrected using Tukey's honestly significant difference. Statistical significance was defined at the $p < 0.05$ level. To examine a linear relationship between peak force and press duration or time-to-peak force, correlations between peak force and press duration or time-to-peak force were calculated by means of all trials of both tasks performed by the typical or experimental group.

RESULTS

A main result of the present study is that adolescents with Down syndrome exhibited a greater magnitude of positive constant and variable errors for peak force than typical adolescents. Although there was no difference between groups for press duration, adolescents with Down syndrome exhibited a shorter time-to-peak force than typical adolescents. In

addition, there was no difference between unimanual and bimanual tasks for both constant and variable errors of peak force and intertap interval in adolescents with Down syndrome.

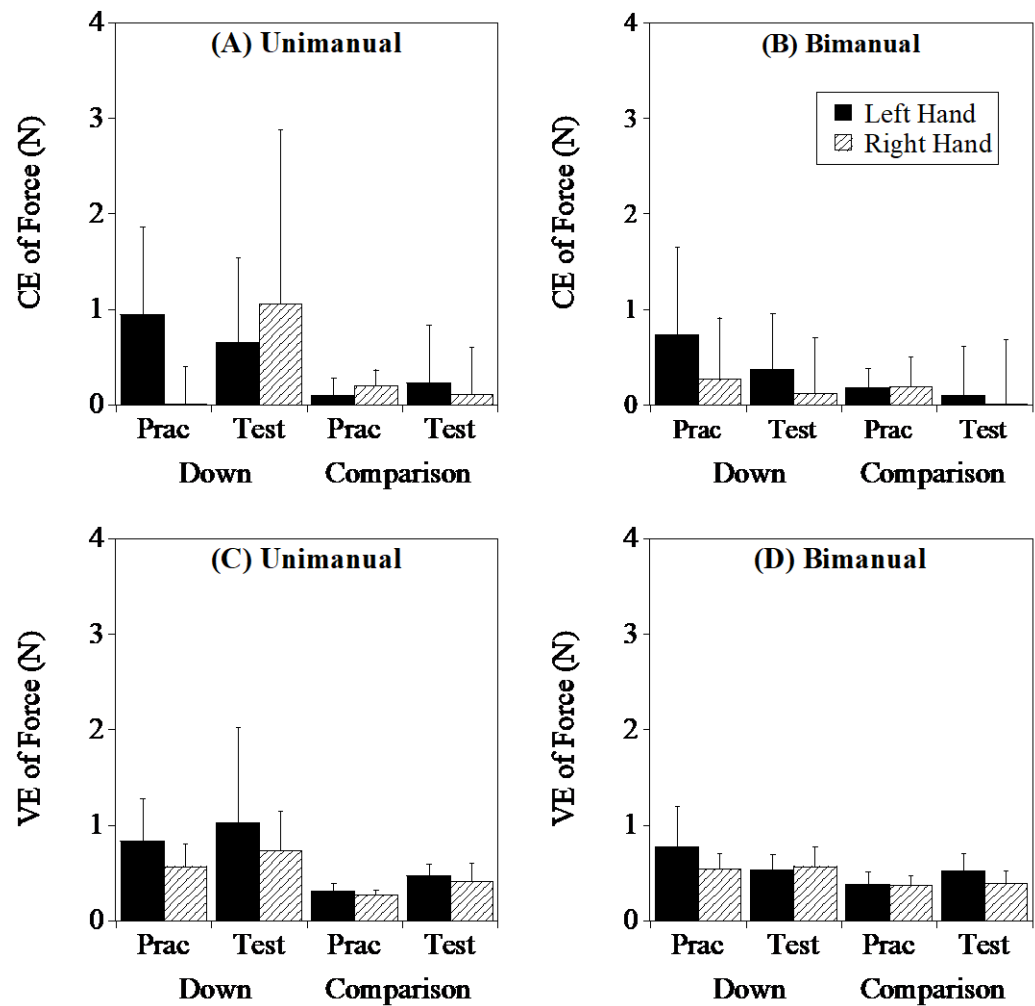


Figure 2. Constant errors (A and B), variable errors (C and D), and their standard deviations of peak force in both the unimanual and bimanual tasks (Masumoto et al., 2012). Abbreviations. CE: constant error, VE: variable error, Prac: practice trial, Test: test trial, Down: group with Down syndrome, Comparison: comparison group.

Figure 2 shows both constant error (A and B) and variable error (C and D) of peak force in both the unimanual and bimanual tasks. The constant error of peak force differed across group ($F(1, 128) = 9.72, p < 0.005$) but did not differ across task, trial, and hand. There was no interaction. Post hoc tests indicated that the experimental group exhibited a larger magnitude of positive constant error than the comparison group. The variable error of peak force differed across group ($F(1, 128) = 29.90, p < 0.0001$) and hand ($F(1, 128) = 4.96, p < 0.05$) but did not differ across task and trial. Post hoc tests indicated that the experimental group exhibited a larger magnitude of variable error than the comparison group and the left hand exhibited a larger magnitude of variable error than the right hand. The interaction of group and task was

further significant ($F(1, 128) = 4.28, p < 0.05$). Separate analyses on variable error showed that, although the experimental group exhibited a larger magnitude of variable error than the comparison group in the unimanual task, there was no difference between groups for variable error in the bimanual task.

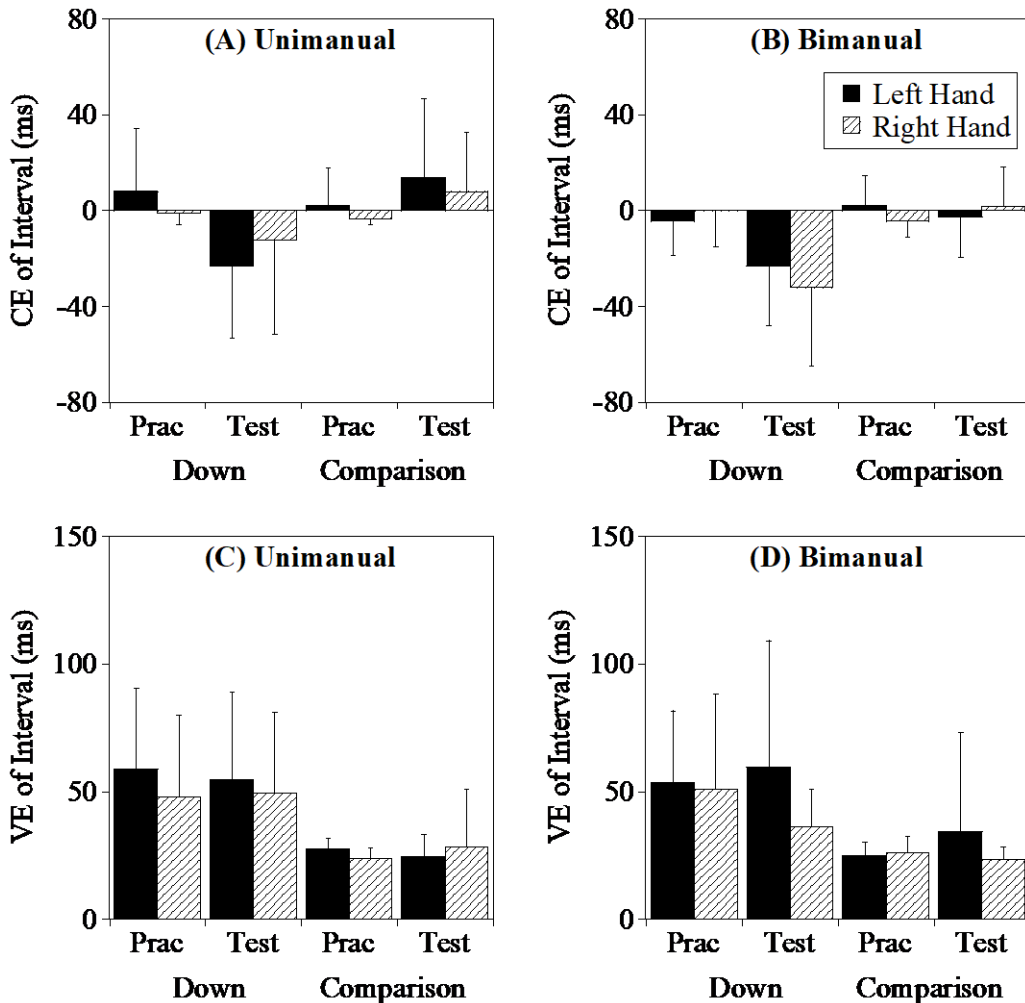


Figure 3. Constant errors (A and B), variable errors (C and D), and their standard deviations of intertap interval in both the unimanual and bimanual tasks (Masumoto et al., 2012). Conventions are as shown in Figure 2.

To examine timing control for the target intertap interval, Figure 3 shows both constant error (A and B) and variable error (C and D) of intertap interval in both tasks. The constant error of intertap interval differed across group ($F(1, 128) = 12.36, p < 0.005$) and trial ($F(1, 128) = 5.27, p < 0.05$) but did not differ across task and hand. Post hoc tests indicated that the experimental group exhibited a larger magnitude of negative error than the comparison group and the test trial exhibited a larger magnitude of negative error than the practice trial. The interaction of group and trial was further significant ($F(1, 128) = 7.38, p < 0.01$). Separate analyses on constant error showed, although there was no difference between groups in the

practice trial, the experimental group exhibited negative errors in the test trial whereas the comparison group had positive errors. The variable error of intertap interval differed across group ($F(1, 128) = 31.23, p < 0.0001$) but did not differ task, trial, and hand. There was no interaction. Post hoc tests indicated that the experimental group exhibited a greater magnitude of variable error than the comparison group did.

To examine interactions of force and timing, Figures 4A and 4B show the mean and SD of press duration in both the tasks. The ANOVA on mean showed no significant main effects or interaction. Figures 4C and 4D show the mean and SD of time-to-peak force in both the tasks. The mean of time-to-peak force differed across group ($F(1, 128) = 60.13, p < 0.0001$) but did not differ task, trial, and hand. Post-hoc tests indicated that the comparison group exhibited a longer time than the experimental group. The interaction of group and hand was further significant ($F(1, 128) = 8.11, p < 0.005$). Separate analyses on mean showed that, although the left hand of the comparison group had a longer time than the right hand of the group, there was no left-right difference for the time of the experimental group.

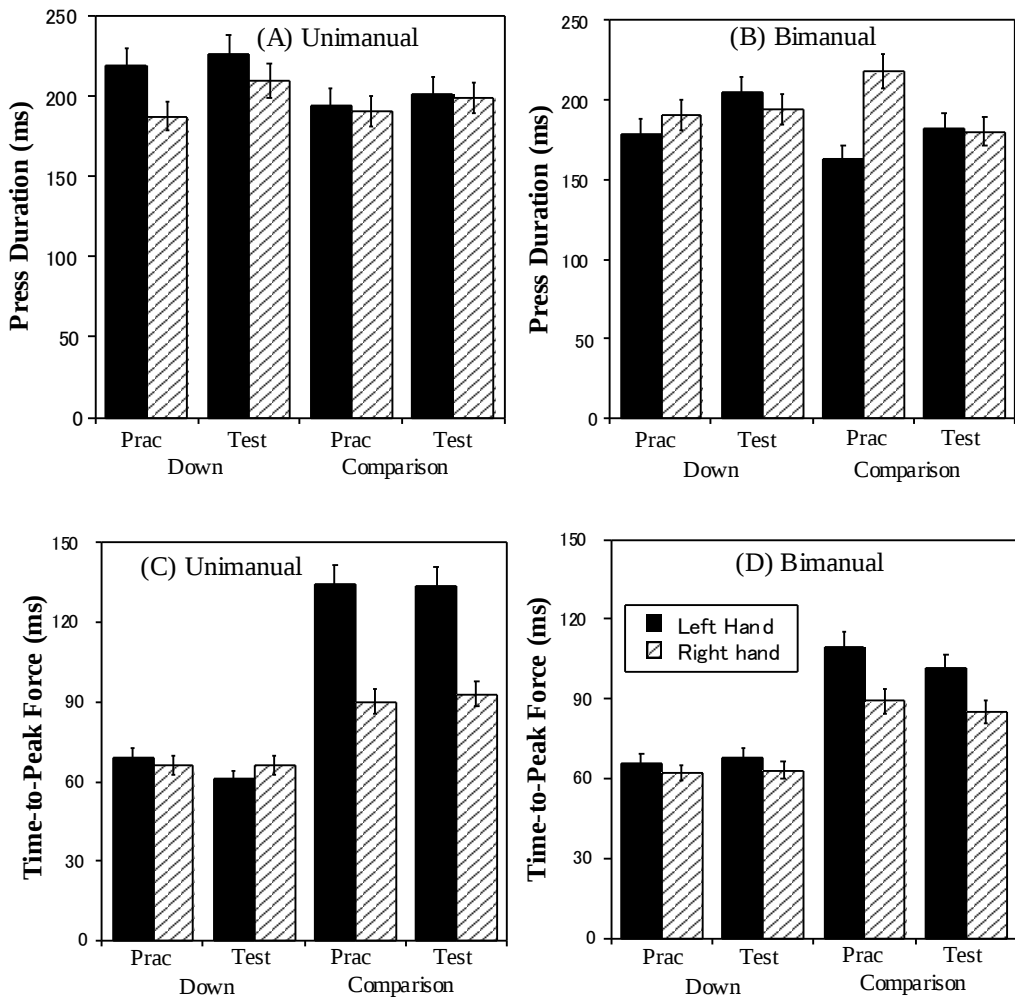


Figure 4. Means and standard deviations of press duration (A and B) and time-to-peak force (C and D) in both the unimanual and bimanual tasks. Conventions are as shown in Figure 2.

To examine a linear relationship between peak force and press duration or time-to-peak force, correlations between peak force and press duration or time-to-peak force were calculated as follows: peak force vs press duration (comparison, $r(70) = 0.32$, $p < 0.01$, experimental, $r(70) = 0.19$, *ns*), peak force vs time-to-peak force (comparison, $r(70) = 0.29$, $p < 0.05$, experimental, $r(70) = 0.21$, *ns*). Although the comparison group exhibited a linear relationship between peak force and press duration or time-to-peak force, the relationship was not familiar to the experimental group.

To examine the lead-lag relationship between the left and right hands in the bimanual task, the mean and SD of the left-right difference for the tapping onset were measured. Although the left hand proceeded to the right hand for the comparison group in the mean of the test trial, the experimental group showed the opposite result for both trials. However the ANOVA on mean and SD showed no significant main effects or interactions.

DISCUSSION

A new finding of the present study is that adolescents with Down syndrome exhibited a greater magnitude of positive constant and variable errors for peak force than typical adolescents. Although there was no difference between groups for press duration, adolescents with Down syndrome exhibited a shorter time-to-peak force than typical adolescents. Whereas typical adolescents exhibited a linear relationship between peak force and press duration or time-to-peak force (also see Ivry, 1986), the relationship was not familiar to Down's syndrome adolescents. Piek et al. (1993) pointed out that because press duration was affected by the direction of the force change, the duration could be attributed to a mechanical aspect of the force change, namely recruitment of motor units or increased motor unit firing rates. Hence, the results of the present study may suggest differences in the manner of motor unit recruitment between typical and Down's syndrome adolescents.

For Down's syndrome persons compared to typical persons, whereas the present study found greater magnitude for peak force in the finger tapping movement, the lower maximal force production was reported by single finger tasks (Latash et al. 2002) and a handgrip task (Heffernan et al. 2009). The discrepancy between the present and previous studies may be influenced by the difference between force tasks. While the previous studies performed the maximal force production, the present study asked the participants to produce the target force with a prescribed interval. Participants did not need to control peak force and timing in the previous studies, but did those in the present study. Because the movement speed probably constrained Down's syndrome participants' performance for achieving the goal of the motor task in the current study, the participants appeared to overshoot the target force.

Heffernan et al. (2009) examined the structure of force variability for persons with Down syndrome in an isometric handgrip task at a constant force using a visual target. Although the experimental group exhibited lower mean force than the comparison group, the experimental group had more variable force than the comparison group. The experimental group further had a greater proportion of spectral power within the 0-4 Hz bandwidth than the comparison group. Using functional magnetic resonance imaging, Vaillancourt et al. (2006) found that increased force variability was associated with reduced cerebellar activity in an isometric force production task performed by typical adults. Because the low-frequency proportion of power

is related to sensorimotor feedback control (Vaillancourt and Newell, 2003), the increase in the low-frequency proportion reported by Heffernan et al. (2009) supports the hypothesis that deficits in cerebellar feed-forward processing are at least partially responsible for elevated force variability (Vaillancourt et al. 2006). In the present study, the adolescents with Down syndrome may have attenuated ability to anticipate the effects of a voluntary change in force. This attenuation may result in greater reliance on feedback processing as indicated by the elevated variable error and positive constant error of force.

On timing, the present study showed that adolescents with Down syndrome had greater magnitude of negative constant and variable errors of intertap interval than typical adolescents. The bimanual control of timing depends on interhemispheric or subcortical information processing. Bimanual coordination of continuous drawing of a circle depends on interhemispheric information processing across the corpus callosum (Spencer et al. 2003), whereas bimanual discrete tapping is controlled by the cerebellum (Kennerley et al. 2002), relating to intertap interval variability for adolescents with Down syndrome. On the other hand, Helmuth and Ivry (1996), asked participants to tap simultaneously with both hands and found less variability in timing with two hands, suggesting that the effect was due to combining the output of two separate timing systems. Masumoto and Inui (2012, 2013) show the same result as Helmuth and Ivry (1996) in bimanual isometric force production of normal male college students. In the present study, however, there was no difference between unimanual and bimanual tasks for variable error of intertap interval. The discrepancy between the previous and present studies may be influenced by the difference between typical and Down's syndrome individuals for the thickness of the corpus callosum. Because people with Down syndrome have exhibited a thinner corpus callosum than typical people (Wang et al. 1992), the problem appears to arise for Down's syndrome people in the transmission of information between cerebral hemispheres. Thus, Down's syndrome people may be unable to combine the output of two separate timing systems.

CONCLUSION

The purpose of this study is to examine the control of force and timing during unimanual and bimanual tapping movements. Adolescents with Down syndrome exhibited greater magnitude for peak force and a systematic delay on the onset of the movement. Although these results suggested differences in motor unit recruitment between Down's syndrome and typical adolescents, the physiological demonstration of this suggestion depends on the research hereafter. Contrary to our prediction, there was no difference for adolescents with Down's syndrome between unimanual and bimanual tasks for both constant and variable errors of peak force and intertap interval. Because people with Down syndrome have exhibited a thinner corpus callosum than typical people, they may be unable to combine the output of two separate timing systems.

REFERENCES

- Anson, J. G. (1992) Neuromotor control and Down syndrome. In J. J. Summers (Ed.), *Approaches to the study of motor control and learning*. Amsterdam: North Holland, pp. 387-412.
- Anson, J. G., and Mawston, G. A. (2000) Patterns of Muscle Activation in Simple Reaction-Time Tasks. In D. J. Weeks, R. Chua, & D. Elliott (Eds), *Perceptual-Motor Behavior in Down Syndrome*. Champaign, IL: Human Kinetics. pp. 3-25.
- Charlton, J. L., Ihsen, E., and Oxley, J. (1996) Kinematic characteristics of reaching in children with Down syndrome. *Human Movement Science*, 15, 727-743.
- Elliott, D. (1990) Movement control in Down syndrome: a neuropsychological approach. In G. Reid (Ed), *Problems in Movement Control*. Amsterdam: North-Holland. pp. 201-216.
- Elliott, D., Hansen, S., Grierson, L. E., Lyons, J., Bennett, S. J., and Hayes, S. J. (2010) Goal-directed aiming: two components but multiple processes. *Psychological Bulletin*, 136, 1023-1044.
- Elliott, D., and Weeks, D. J. (1993) Cerebral specialization and the control of oral and limb movements for individuals with Down's syndrome. *Cortex*, 29, 103-113.
- Elliott, D., Weeks, D. J., and Elliott, C. L. (1987) Cerebral specialization individuals with Down syndrome. *American Journal of Mental Retardation*, 92, 263-271.
- Elliott, D., Weeks, D. J., and Jones, R. (1986) Lateral asymmetries in finger-tapping by adolescents and young adults with Down syndrome. *American Journal of Mental Deficiency*, 90, 472-475.
- Henderson, S. E., Morris, J., and Frith, U. (1981) The motor deficit in Down's syndrome children: A problem of timing? *Journal of Child Psychology and Psychiatry*, 22, 233-245.
- Hartley, X. Y. (1982) Receptive language processing of Down's syndrome children. *Journal of Mental Deficiency Research*, 26, 263-269.
- Helmuth, L. L., and Ivry, R. B. (1996) When two hands are better than one: reduced timing variability during bimanual movements. *Journal of Experimental Psychology: Human Perception and Performance*, 22, 278-293.
- Heffernan, K. S., Sosnoff, J. J., Ofori, E., Jae, S. Y., Baynard, T., Collier, S. R., Gouloupoulou, S., Figueroa, A., Woods, J. A., Pitetti, K. H., and Fernhall, B. (2009) Complexity of force output during static exercise in individuals with Down syndrome. *Journal of Applied Physiology*, 106, 1227-1233.
- Henderson, S. E., Morris, J., and Frith, U. (1981) The motor deficit in Down syndrome children: A problem of timing? *Journal of Child Psychology and Psychiatry*, 22, 223-245.
- Inui, N. (2007) Simple reaction times and timing of unimanual and bimanual serial actions in adolescents with mental retardation, autism, and Down syndrome. In E. Heinz (Ed.) *Mental Retardation Research Advances*, Hauppauge: NY, Nova Science Publishers, Inc., pp. 1-33.
- Inui, N., and Asama, K. (2003a) Timing of bimaual alternating finger-tapping sequences in adolescents with mental retardation: a pilot study. *Perceptual and Motor Skills*, 97, 398-400.
- Inui, N., and Asama, K. (2003b) Timing of bimaual rhythmic finger tapping in adolescents with mental retardation or autism. *Journal of Human Movement Studies*, 45, 59-80.

- Inui, N., Yamanishi, M., and Tada, S. (1995) Simple reaction times and timing of serial reactions of adolescents with mental retardation, autism, and Down syndrome. *Perceptual and Motor Skills*, 81, 739-745.
- Kennerley, S. W., Diedrichsen, J., Hazeltine, E., Semjen, A., and Ivry, R. B. (2002) Callosotomy patients exhibit temporal uncoupling during continuous bimanual movements. *Nature Neuoscience*, 5, 367-381.
- Ivry, R. B. (1986) Force and timing components of the motor program. *Journal of Motor Behavior*, 18, 449-474.
- Latash, M. L. (1992) Motor control in Down syndrome: the role of adaptation and practice. *Journal of Developmental and Physical Disabilities*, 4, 227-161.
- Latash, M. L. (2000) Motor coordination in Down syndrome: the role of adaptive changes. In D. J. Weeks, R. Chua & D. Elliot (Eds.) *Perceptual-Motor Behavior in Down Syndrome*. Champaign, IL: Human Kinetics. pp. 225-248.
- Latash, M. L., Kang, K., and Patterson, D. (2002) Finger coordination in persons with Down syndrome: atypical patterns of coordination and the effects of practice. *Experimental Brain Research*, 146, 345-355.
- Masumoto, J., Abe, T., and Inui, N. (2012) Adolescents with Down syndrome exhibit greater force and delay in onset of tapping movements. *Perceptual and Motor Skills*, 114, 826-836.
- Masumoto, J., and Inui, N. (2012) Effects of force levels on error compensation in periodic bimanual isometric force control. *Journal of Motor Behavior*, 44, 261-266.
- Masumoto, J., and Inui, N. (2013) Effects of movement duration on error compensation in periodic bimanual isometric force production. *Experimental Brain Research*, 227, 447-455.
- Oldfield, R. C. (1971) The assessment and analysis of handedness: the Edinburgh inventory. *Neuropsychologia*, 9, 97-113.
- Piek, J. P., Glencross, D. J., Barrett, N. C., and Love, G. L. (1993) The effect of temporal and force changes on the patterning of sequential movements. *Psychological research*, 55, 116-123.
- Planinsek, T. (1996) *The control of timing among children with Down Syndrome*. Unpublished master's thesis, Catholic University, Leuven, Belgium.
- Ringenbach, S. D., Chua, J. B., Maraj, B. K. V., Kao, J. C., and Weeks, D. J. (2002) Bimanual coordination dynamics in adults with Down syndrome. *Motor control*, 6, 390-409.
- Robertson, S. D., Van Gemmert, A. W. A. and Maraj, B. K. V. (2002) Auditory information is beneficial for adults with Down syndrome in a continuous bimanual task. *Acta Psychologica*, 110, 213-229.
- Spencer, R. M. C., Zelaznik, H. N., Diedrichsen, J., and Ivry, R. B. (2003) Disrupted timing of discontinuous but not continuous movements by cerebellar lesions. *Science*, 300, 1437-1439.
- Teipel, S. J., Alexander, G. E., Schapiro, M. B., Moller, H. J., Rapoport, S. I., and Hampel, H. (2004) Age-related cortical grey matter reductions in non-demented Down's syndrome adults determined by MRI with voxel-based morphometry. *Brain* 127, 811-824.
- Vaillancourt, D. E., Mayka, M. A., and Corcos, D. M. (2006) Intermittent visuomotor processing in the human cerebellum, parietal cortex, and premotor cortex. *Journal of Neurophysiology*, 95, 922-931.

- Vaillancourt, D. E., and Newell, K. M. (2003) Aging and the time and frequency structure of force variability. *Journal of Applied Physiology*, 97, 454-455.
- Vimercati, S. L., Galli, M., Rigoldi, C., Ancillao, A., and Albertini, G. (2013) Motor strategies and motor programs during an arm tapping task in adults with Down syndrome. *Experimental Brain Research*, 225, 333-338.
- Wang, P. P., Doherty, S., Hesselink, J. R., and Bellugi, U. (1992) Callosal morphology concurs with neurobehavioral and neuropsychological findings in two neurodevelopmental disorders. *Archives of Neurology*, 49, 407-411.

Chapter 81

MANAGEMENT OF EXECUTIVE FUNCTION FOLLOWING ASSISTED CYCLING THERAPY IN ADOLESCENTS WITH DOWN SYNDROME

***Shannon D. R. Ringenbach¹, Simon D. Holzapfel¹,
Madeline Richter¹ and Jay L. Alberts²***

¹Exercise Science and Health Promotion,
Arizona State University, Phoenix, AZ, US

²Biomedical Engineering, Cleveland Clinic, Cleveland, OH, US

ABSTRACT

We have shown promising results of Assisted Cycling Therapy (ACT) for improving executive functioning in adolescents with Down syndrome (DS). The current study examines the one month retention of executive function benefits gained by adolescents with DS. Fifteen participants were randomly assigned to voluntary cycling (VC; i.e., self-selected cadence) or Assisted Cycling Therapy (ACT; i.e., 65% faster than self-selected cadence accomplished by a motor). Both cycling groups rode a stationary bicycle, for 30 minutes, three times a week, for eight weeks. At the beginning (i.e., pre-test) and end (post-test) of the 8-week session, and at a one month retention (follow-up), three executive functions including set-switching, inhibition, and cognitive planning, were measured. The results showed improved cognitive planning and set-switching for the ACT group after 8 weeks of intervention and these improvements were maintained for one month after the intervention. However, no significant differences were found between the cycling groups for our measure of inhibition. Thus, our results suggest that, especially in regards to cognitive planning and set switching, ACT may lead to relatively permanent changes in the brain.

Keywords: Intellectual disability, treatment, cognitive planning, prefrontal cortex

INTRODUCTION

Down syndrome (DS), also known as Trisomy 21, is a chromosomal disorder in which a third copy of the 21st chromosome affects both structural and behavioral development of one in every 800 live births in the United States (Chapman & Hesketh, 2000). Intellectual disability is a hallmark in persons with DS. Executive function is a term used to describe a set of higher order cognitive processes beyond instinct or automatic responses that include inhibitory control, attentional control, and cognitive planning. These processes are crucial for selecting and monitoring behaviors to achieve goals of activities of daily living (e.g., brushing teeth, crossing the street, etc.). Research has shown that adolescents with DS display pervasive deficits in all domains of executive function (Lanfranchi, Jerman, Dal Pont, Alberty, & Vianello, 2010). Furthermore, these cognitive disabilities are a challenge and seem to limit their opportunities to physical activity participation (Cowley et al., 2010; Rhiitman et al., 2010).

Exercise may be an effective treatment for cognitive impairments because the positive influence of exercise on cognition has been demonstrated in other populations (e.g., elderly (Colcombe & Kramer, 2003; van Uffelen, Chin, Hopman-Rock, van Mechelen, 2008); typical children (Hillman, Erickson, & Kramer, 2003; Hillman, Snook, & Jerome, 2003)) and mice models (Ts65Dn) of DS (Llorens-Martin et al., 2010). There is an emerging body of literature in healthy older adults and individuals with Alzheimer's disease indicating that exercise results in structural and functional changes in the brain (Petzinger et al., 2006; Petzinger et al., 2007). These alterations in brain structure and function suggest that CNS function can be altered via voluntary exercise in individuals with relatively normal patterns of activation within the motor cortex. However, because persons with DS have limited motor output due to physiological and psychosocial factors, their ability to induce changes in CNS function may be compromised when engaging in voluntary exercise performed at their preferred (i.e., low) rates. For example, individuals with DS also display reduced exercise capacity which can hinder their ability to perform certain functional tasks or activities (Mendonca, Pereira, & Fernhall, 2010). Other physical characteristics can be particularly detrimental to the individual's participation in exercise and aerobic activities. For example, hypotonia and ligamentous laxity, cause shorter step length, increased knee flexion when walking, and decreased single-limb support (Lewis & Fragala-Pinkham, 2005). These physiological issues associated with DS make physical exertion more difficult. Whether caused by these physical limitations or their reduced cognitive function, individuals with DS are inherently prone to a more sedentary lifestyle (Jobling & Cuskelly, 2006) which creates an increased risk for obesity, diabetes, and cardiovascular disease (Lewis et al., 2005).

It is therefore necessary to employ assisted-exercise techniques and machinery in order to increase the rate of movement and subsequent potential cognitive benefits in this population. Though several studies have examined the effects of aerobic exercise on improving physical fitness, counteracting potential chronic disease risks, and improving executive function in typical populations (Smith et al., 2010; Warburton, Nicol, & Bredin, 2006), there is little-to-no research focused on the relationship between exercise and executive function in individuals with DS.

Animal studies have indicated that assisted-exercise, requiring an animal to exercise at a rate elevated from that at which it would exercise on its own, improves motor function and displays neuroprotective properties for Parkinson's disease (PD; Fisher, Petzinger, & Nixon,

2004). Similarly, researchers have found assisted aerobic exercise to improve cognitive function in human adults with Parkinson's disease (Ridgel, Kim, & Fickes, 2010). Recent research has demonstrated acute improvements in manual motor functioning and information processing in adolescents with DS after a single session of assisted cycling in comparison to voluntary cycling or no cycling (Ringenbach, Chen, Albert, Lichtsinn, & Alberts, 2013; Ringenbach, Albert, Chen, & Alberts, 2014). Benefits of long term assisted cycling on executive function in adolescents with DS have also been reported. Eight weeks of assisted cycling was associated with improvements in planning ability, inhibitory control, and reaction times (Holzapfel, Ringenbach, Mulvey, Sandoval-Menendez, Cook, Ganger, & Bennett, 2015; Ringenbach, Holzapfel, Mulvey, Jimenez, Benson, & Richter, 2016) in adolescents with DS. Voluntary cycling appeared to benefit set-switching and both assisted and voluntary cycling led to improvements in verbal fluency (Ringenbach et al., 2016). However, it has not been investigated whether these benefits are retained after the cessation of the cycling intervention.

Improvements in global motor function, that is, improvements in upper limb control after exercise in the lower limbs, and improvements in overall executive function following Assisted Cycling Therapy but not Voluntary Cycling suggest structural changes are occurring at the cortical level. In addition, the persistence of enhanced motor, cognitive, and clinical measures after the exercise intervention is stopped would indicate more permanent changes such as changes in the neural structure of the brain. Thus, we predict that, similar to patients with Parkinson's disease (Alberts, Linder, Penko, Lowe, & Phillips, 2011; Ridgel, Vitek, & Alberts, 2009), adolescents with DS will show improvements in executive function (i.e., cognitive planning, inhibition, and set-switching) following assisted, but not voluntary cycling after eight weeks, and that these changes will be maintained after one month of no cycling.

METHODS

Participants: Sixteen individuals with DS between the ages of 9 and 26, with a mean age of 18.6 years, completed this intervention (see Table 1). Though "adolescent" commonly refers to the teenage years, 13-19, it has been noted that physical, psychological, and cultural maturation can occur prior to or after this period (Coleman & Roker 1998). Due to the common delay of development within the DS population, this study refers to all participants between the ages of 9 and 26 as adolescents. Participants for this exercise intervention were recruited through fliers, email announcements, and by word of mouth from a variety of local organizations for persons with DS. For health and safety reasons all participants were screened for exercise preparedness prior to entering the intervention. Guardians were asked to complete a seven-question "Physical Activity Readiness Questionnaire" on behalf of the participant. Participants were considered cleared for exercise if all seven questions were answered as "no," or if specifically cleared for this study by their healthcare practitioner. The mental age of participants was assessed with the Peabody Picture Vocabulary Test 4 (PPVT-4; Dunn & Dunn, 2007). All protocols were approved by the Human Subjects Institutional Review Board of Arizona State University.

Table 1. Participant Characteristics

	ACT (n = 8)		VC (n = 7)	
	M	/F	M	/F
Gender	5	/3	3	/4
	Mean	± SD	Mean	± SD
Chronological age (years)	19.71	± 4.14	18.75	± 3.76
Mental age (years)	5.33	± 2.95	5.27	± 2.68
BMI (kg/m ²)	27.76	± 8.43	28.98	± 6.03
Cadence (rpm)	80.61	± 8.83	48.80	± 7.54
Heart rate (bpm)	97.64	± 8.07	103.95	± 9.07

ACT = Assisted Cycling Therapy, BMI = Body mass index, VC = voluntary cycling.

Intervention: This study included two distinct and randomly assigned interventions: Assisted Cycling Therapy (ACT) and an active control intervention termed voluntary cycling (VC). The ACT group consisted of eight participants and the VC group consisted of seven participants. Both exercise groups completed three 30-minute cycling sessions per week for eight weeks. Participants were kept as close to this schedule as possible with an average span of participation of 9.34 weeks.

In the VC intervention, the mechanical motor of the stationary bicycle was not engaged. Participants pedaled at their own self-selected and self-regulated rate for 30 minutes. The ACT intervention utilized the mechanical motor of the bicycle to increase the cycling cadence. First, participants in the ACT group were asked to cycle at a self-selected rate for 5 minutes in order to warm up. This initial rate was then used to determine an accelerated rate at which the bicycle motor would be set; which was, on average, a 65% increase compared to the VC cadence (see Table 1). During the first ACT session, the motor was programmed to a cadence which was 35% faster than the voluntary warm-up cadence. This rate was generally increased in increments of 2-5 rpm as the 8-week intervention progressed to achieve a maximal cadence that the participants were comfortable maintaining.

Because this study focused on the benefits of aerobic exercise at varying rates of movement, it was important to ensure that the exercise intensity remained in the light to moderate aerobic intensity range. Heart rate was continuously monitored during the cycling sessions and the participants were instructed to slow down or the motor-assisted cadence was reduced if the heart rate exceeded 80% of the age-predicted maximal heart rate based on this formula: Age-predicted maximal heart rate = $210 - 0.56 \times \text{age} - 31$ (Fernhall et al., 2001).

Cognitive Testing: Participants completed three cognitive tests before (pre) and after (post) the 8-week session, as well as after their one month retention period (follow-up). These tests were designed to evaluate their cognitive planning, inhibition, and set-switching capabilities.

I. Cognitive Planning as measured by the “Tower of London” test: This test utilized a wooden platform with three pegs of graduating height and three wooden balls (blue, red, and yellow). The three pegs could accommodate a maximum of one, two, or three ball, respectively. The goal of each trial was to move the balls from their starting position to the final position, as depicted by a printed image, in the given number of moves and within the 45-second time limit. Participants were told to move only one ball at a time and that each move must end with the ball on one of the three pegs which still had available space. After one practice trial, participants moved through 17 different trials of increasing complexity and difficulty. Each trial was

considered a success if the participant achieved the goal arrangement in under 45 seconds and with the provided number of moves. The test was ended if the participant failed to complete four trials in a row, either due to time limitations, failure to follow the rules, or completing the trial in too many moves. The number of trials that were solved correctly within 45 seconds was used as the outcome measure.

II. Inhibition as measured by the “Knock-Tap” test: In the first trial, participants were shown how to knock (i.e., using their knuckles) and tap (i.e., using the palm of their hand) on the table at which they were seated across from the researcher. They were instructed to tap on the table each time the researcher knocked, and knock on the table each time the researcher tapped. A trial of 4 actions, two knocks and two taps, were performed with researcher prompting before the 15-action trial was performed. One point was awarded for each correct action for a maximum of 15 points. The second trial involved the actions of responding with a knock, fist, or no response; as was demonstrated by the researcher for the participant to mimic. Participants were asked to knock when the researcher made a fist, to make a fist when the researcher knocked, and to remain still (i.e., no active response) when the researcher tapped. There was a 6-action practice prior to the 15-action trial. A total of 30 points was possible after the combination of the two trials. The total number of correct responses/actions was used as the outcome measure.

III. Set-Switching as measured by the “Card Sorting” test: For the following two trials, the cards were always presented in a pre-determined order and the instructions were read from a pre-written script in order to minimize variation between researchers. Participants were reminded of the rules and shown an example of proper sorting for one card at the beginning of each trial. Each participant was presented with two trays; one held a card with a blue rabbit on the left-hand side and the other had a red boat on the right-hand side of the participant. For the first trial the participants were asked to sort six cards, depicting either a red rabbit or a blue boat, one-by-one, into the appropriate tray based on color (i.e., the red rabbit would go into the right-hand tray showing the red boat). Each card was placed face down in the tray to avoid patterned sorting. One point was awarded for each card placed in the proper tray, allowing for a maximum of six points. For the second trial, the participants were asked to sort the same six cards by shape rather than color (i.e., the red rabbit would go into the left-hand tray which showed a blue rabbit). This meant that participants had to cope with the rule switch of sorting by color to sorting by shape. Again, the cards were placed face down before the next card was shown, and a maximum of six points was awarded for correct placement. The number of correctly sorted cards during the second trial (sorting by shape) was used as the outcome measure of set-switching ability.

Statistical Procedures: All data were analyzed using a within-between subjects general linear model analysis. Time (pre, post, and follow-up) was the within-subjects factor and group (ACT and VC) was the between subjects factor. Next, the VC and ACT groups were analyzed independently using a one-way ANOVA with time as a fixed factor to examine differences between time points within each group. Tukey HSD post-hoc test were used for multiple comparisons between time points. Data was analyzed with the Statistical Package for the Social Sciences v. 23 (SPSS; IBM Corporation, Armonk, NY) and the null hypothesis was rejected at $p \leq 0.005$.

RESULTS

Cognitive Planning (Figure 1): The main effect of time for the number of correct trials approached conventional levels of significance ($F(2,26) = 2.90$, $p = 0.073$). There was a significant interaction between time and group ($F(2,26) = 5.01$, $p = 0.014$). The main effect of time was significant for the ACT group ($F(2,21) = 6.86$, $p = 0.005$) and post-hoc comparisons indicated a pre- to post-test improvement ($\bar{x}_{\text{post}} - \bar{x}_{\text{pre}} = 2.75$, $SE = 0.94$, $p = 0.022$), a pre- to follow-up improvement ($\bar{x}_{\text{follow-up}} - \bar{x}_{\text{pre}} = 3.25$, $SE = 0.94$, $p = 0.007$), and no change from post-testing to follow-up ($\bar{x}_{\text{follow-up}} - \bar{x}_{\text{post}} = 0.50$, $SE = 0.94$, $p = 0.858$) for the number of correct trials in the ACT group. The main effect of time was not significant for the VC group ($F(2,21) = 0.71$, $p = 0.504$) and post-hoc comparisons indicated no pre- to post-test change ($\bar{x}_{\text{post}} - \bar{x}_{\text{pre}} = -1.29$, $SE = 1.24$, $p = 0.566$), no pre- to follow-up change ($\bar{x}_{\text{follow-up}} - \bar{x}_{\text{pre}} = 0.00$, $SE = 1.24$, $p = 0.007$), and no change from post-testing to follow-up ($\bar{x}_{\text{follow-up}} - \bar{x}_{\text{post}} = 1.29$, $SE = 1.24$, $p = 0.858$) for the number of correct trials in the VC group.

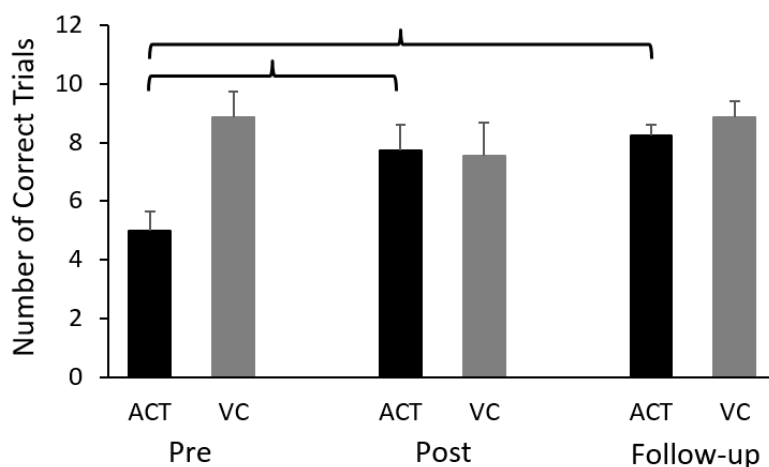


Figure 1. Means and standard errors of successful “Tower of London” (e.g., cognitive planning) trials as a function of group and time. Brackets indicate significant differences based on post-hoc tests between the means they connect. ACT = Assisted Cycling Therapy, VC = voluntary cycling.

Inhibition (Figure 2): The main effect of time for the number of correct responses was significant ($F(2,26) = 3.82$, $p = 0.035$). There was no interaction between time and group ($F(2,26) = 0.17$, $p = 0.849$). The main effect of time was not significant for the ACT group ($F(2,21) = 0.18$, $p = 0.839$) and post-hoc comparisons indicated no pre- to post-test change ($\bar{x}_{\text{post}} - \bar{x}_{\text{pre}} = 2.50$, $SE = 5.11$, $p = 0.877$), no pre- to follow-up change ($\bar{x}_{\text{follow-up}} - \bar{x}_{\text{pre}} = 2.75$, $SE = 5.11$, $p = 0.853$), and no change from post-testing to follow-up ($\bar{x}_{\text{follow-up}} - \bar{x}_{\text{post}} = 0.25$, $SE = 5.11$, $p = 0.999$) for the number of correct responses in the ACT group. The main effect of time was not significant for the VC group ($F(2,21) = 0.20$, $p = 0.824$) and post-hoc comparisons indicated no change from pre- to post-test ($\bar{x}_{\text{post}} - \bar{x}_{\text{pre}} = 1.86$, $SE = 5.48$, $p = 0.939$), no change from pre-test to follow-up ($\bar{x}_{\text{follow-up}} - \bar{x}_{\text{pre}} = 3.43$, $SE = 5.48$, $p = 0.808$), and no change from post-testing to follow-up ($\bar{x}_{\text{follow-up}} - \bar{x}_{\text{post}} = 1.57$, $SE = 5.48$, $p = 0.956$) for the number of correct responses in the VC group. Although there was a main effect of time, the mean improvements in either group were not significant due to substantial error variance.

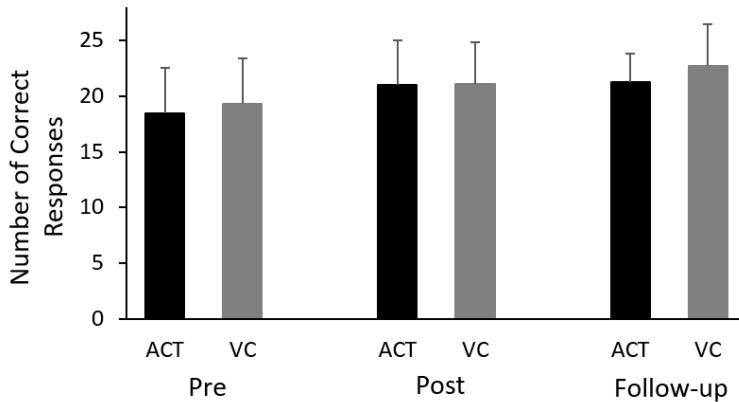


Figure 2. Means and standard errors of correct “Knock-Tap” (e.g., inhibitory control) responses as a function of group and time. ACT = Assisted Cycling Therapy, VC = voluntary cycling.

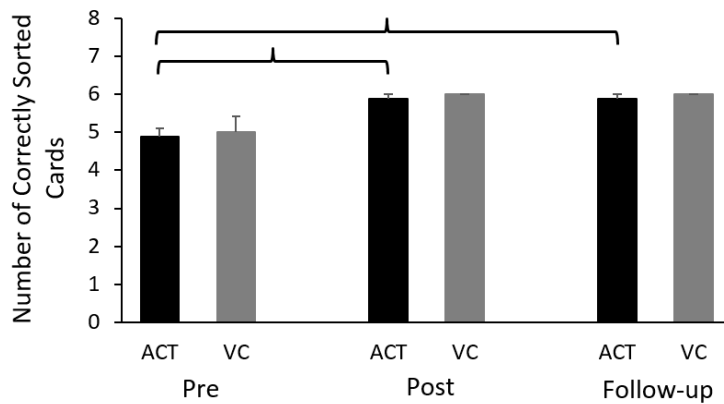


Figure 3. Means and standard errors of correctly sorted cards during the second trial (i.e., sort by shape) of the “Card Sorting Test” (i.e., set-switching) as a function of group and time. Brackets indicate significant differences based on post-hoc tests between the means they connect. ACT = Assisted Cycling Therapy, VC = voluntary cycling.

Set-Switching (Figure 3): The main effect of time for the number of correctly sorted cards was significant ($F(2,26) = 16.18, p < 0.001$). There was no interaction between time and group ($F(2,26) = 0.00, p = 1.000$). The main effect of time was significant for the ACT group ($F(2,21) = 12.11, p < 0.001$) and post-hoc comparisons indicated a pre- to post-test improvement ($\bar{x}_{\text{post}} - \bar{x}_{\text{pre}} = 1.00, SE = 0.23, p < 0.001$), a pre- to follow-up improvement ($\bar{x}_{\text{follow-up}} - \bar{x}_{\text{pre}} = 1.00, SE = 0.23, p < 0.001$), and no change from post-testing to follow-up ($\bar{x}_{\text{follow-up}} - \bar{x}_{\text{post}} = 0.00, SE = 0.23, p = 1.000$) for the number of correctly sorted cards in the ACT group. The main effect of time was not significant for the VC group ($F(2,21) = 2.33, p = 0.126$) and post-hoc comparisons indicated no change from pre- to post-test ($\bar{x}_{\text{post}} - \bar{x}_{\text{pre}} = 1.00, SE = 0.53, p = 0.176$), no change from pre-test to follow-up ($\bar{x}_{\text{follow-up}} - \bar{x}_{\text{pre}} = 1.00, SE = 0.53, p = 0.176$), and no change from post-testing to follow-up ($\bar{x}_{\text{follow-up}} - \bar{x}_{\text{post}} = 0.00, SE = 0.53, p = 1.000$) for the number of correctly sorted cards in the VC group. The mean improvement was the same for both groups, but the improvements in the ACT group reached statistical significance due to lower error variance.

DISCUSSION

This is the first study, to our knowledge, that has utilized a chronic (i.e., 8 week) between group intervention using Assisted Cycling Therapy (ACT) in adolescents with DS and measured changes in executive functioning. We believe it is also the first study reporting on the retention of improvements in executive function following an aerobic exercise intervention. Enhancing executive functioning is critical to improving activities of daily living (e.g., dressing, preparing food, self-care, etc.), fostering independence, and improving quality of life for persons with DS (Alptekin et al., 2005; Borella et al., 2010; Diamond, 2013; Evans & Gray, 2000; Holzapfel et al., 2015; Jefferson et al., 2006; Myers & Pueschel, 1991; Shaw et al., 2006; Wegener et al., 2005; Whitmer & Banich 2007).

Given that the brain of an individual with DS contains a reduced number of neurons and abnormal neuron morphology, especially in the cerebral cortex, it is not surprising that many persons with DS suffer some reduced executive functions (Reeves & Irving, 1995). This study aimed to improve executive function in three specific areas: planning, inhibition, and set-switching, and to observe the retention of improvements over a one month period. Though improvements were only observed in cognitive planning ability and set-switching ability, retention of post-intervention levels was observed in all three measures. The improvements were most evident following ACT which has been suggested to promote cognitive improvements in adolescents with DS (Holzapfel et al., 2015; Ringenbach et al., 2016). The data from this study are consistent with this finding that improvements may be more permanent (e.g., last at least 4 weeks), which may indicate changes at the neuronal level.

Our results show that cognitive planning and set-switching ability improved only in the ACT group (see Figures 1 and 3). The improvements from pre- to post-test were significant and the improvements were retained for 4 weeks following the post-test. No changes in cognitive planning or set-switching were observed in the VC group. There was a trend for improvements in inhibitory control across both groups (see Figure 2). The main effect of time was significant but there was no interaction between time and group. Within group improvements in inhibition did not reach significance.

Thus, improvements in executive function were greatest after ACT and improvements were generally maintained. The relative preservation of executive function improvements indicates neuroplastic changes in the prefrontal cortex (Alberts et al., 2011; Holzapfel et al., 2015; Ringenbach et al., 2016). Our results are consistent with many studies reporting the maintenance of executive function following cognitive intervention programs (Valenzuela & Sachdev, 2009; Wexler, 2007). However, exercise interventions do not always result in the retention of gains in executive function (Quaney et al., 2009) unless the exercise regimen is continued (Emery, Shermer, Hauck, Hsiao, & MacIntyre, 2003).

Aerobic exercise has been shown to increase cortical excitability and cerebral blood flow which are associated with the upregulation of dopamine, brain-derived neurotrophic factor, glial-derived neurotrophic factor, insulin-like growth factor, and vascular-endothelial derived growth factor (Alberts et al., 2011; Christensen et al., 2000; Cotman et al., 2007; Cotman & Engesser-Cesar, 2002; Nobrega et al., 1994; Piepmeier & Etnier, 2015; Tajiri et al., 2009; Tillerson, Caudle, Reveron, & Miller, 2003). These neurotrophic and growth factors regulate the growth and plasticity of neurons and growth and health of blood vessels which are the mechanisms that may explain the link between exercise and improvements in executive

function (Audiffren & André, 2015; Cotman et al., 2007; Piepmeier & Etnier, 2015). These relationships and mechanisms are illustrated in Figure 4 (Alberts et al., 2011).

The question that remains is why ACT seems to benefit executive function more than VC. The exercise intensities as measured by heart rate were similar between the ACT and VC group. The only difference was in the cadence and the nature of the movement production. The ACT cadence was on average 65% faster than the VC cadence (see Table 1) and the movement output was assisted during ACT. Greater frequencies of mechanical stimulation have been shown to elicit greater corticospinal excitability compared to lower frequencies (Christova et al., 2011). Similarly, the increased cycling cadence during ACT compared to VC may have enhanced the flexion and extension moments of force that are associated with cycling and thereby increased neural activity in the involved musculature (Ericson, 1985). The increased movement rate also resulted in more rapid muscle shortening and lengthening which is thought to stimulate velocity dependent muscle spindle fibers (Corbett et al., 2013). The increased proprioceptive input from these fibers may have enhanced corticospinal excitability (Alberts et al., 2011; Corbett et al., 2013) and thereby upregulated the activity of brain-derived neurotrophic factor, glial-derived neurotrophic factor, and vascular-endothelial derived neurotrophic factor (Alberts et al., 2011; Cotman et al., 2007; Piepmeier & Etnier, 2015; Tillerson, Caudle, Revereon, & Miller, 2003). Based on our results, these biochemical processes and resultant neuroplastic changes appear to be robust and temporally stable up to four weeks post-exercise in adolescents with DS.

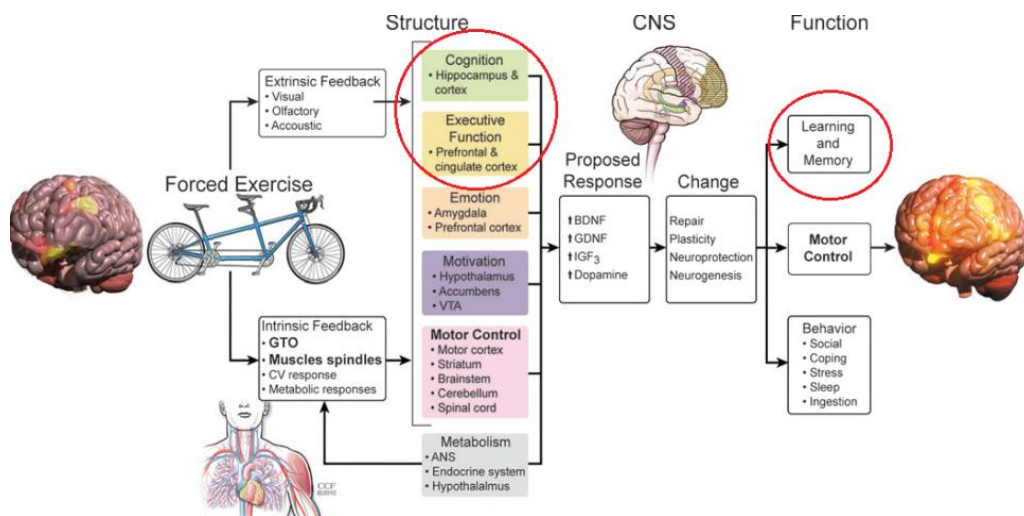


Figure 4. Model of mechanisms involved in Assisted Cycling Therapy.

It may be that neuroplastic and perhaps vascular changes last longer in individuals with DS compared to other populations as we found retention of executive benefits whereas other studies did not, following an exercise intervention (Emery et al., 2003; Quaney et al., 2009). However, our follow-up period of four weeks is shorter compared to other studies and we do not know how much time after the exercise intervention would need to pass before executive functions return to baseline levels.

STRENGTHS AND LIMITATIONS

To the best of our knowledge, this is the first study which examined the effects of aerobic exercise on executive function and the retention of executive function benefits in persons with DS. Participants were randomized to two different types of cycling exercise, whereas VC served as an active rather than passive control group. However, a passive control group could have further informed the results by allowing for the quantification of a learning effect that may have taken place across the 12 weeks. The significant results, in spite of the limited sample size, indicate that the intervention effects were quite substantial. In fact, the improvements in measures of executive function ranged from 0.17 to 1.54 (Cohen's *d*) in the ACT group.

CONCLUSION

The present study examined an innovative exercise intervention that may be able to overcome low exercise capacities commonly seen in persons with DS by augmenting the movement rates during cycling exercise. The assisted nature of ACT may maximize neurological benefits of exercise, by compensating for the decreased motivation and slow movement rates in adolescents with DS. We found that ACT may improve cognitive function through potentially long-term changes to the brain's structure. The results of this study indicate that cognitive planning and set-switching may display long-term improvement after an 8-week ACT intervention. However, longer intervention periods may be needed to produce improvements in inhibitory control. It is unknown at this point whether the observed improvements would last longer than 4 weeks. This area of research could be further expanded to test dose-response relationships between cadence and executive functions and to develop an assisted exercise plan modeled directly towards managing executive function deficits in people with DS.

REFERENCES

- Alberts J. L., Linder, S. M., Penko, A. L., Lowe M. J., & Phillips M. (2011). It is not about the bike, it is about the pedaling: Forced exercise and Parkinson's disease. *Exercise and Sport Sciences Review*, 39(4), 177-186.
- Alptekin, K., Akvardar, Y., Akdede, B. B. K., Dumlu, K., Işık, D., Pirinçci R., Kitiş, A. (2005). Is quality of life associated with cognitive impairment in schizophrenia?. *Progress in Neuro-Psychopharmacology and Biological Psychiatry*, 29(2), 239-244.
- Audiffren, M. & André N. (2015). The strength model of self-control revisited: Linking acute and chronic effects of exercise on executive functions. *Journal of Sport and Health Science*, 4, 30-46.
- Borella E., Carretti B., & Pelegrina S. (2010). The specific role of inhibition in reading comprehension in good and poor comprehenders. *Journal of Learning Disabilities*, 43(6), 541-552.

- Chapman, R. S., & Hesketh, L. J. (2000). Behavioral phenotype of individuals with Down syndrome. *Mental Retardation and Developmental Disabilities Research Reviews*, 6(2), 84-95.
- Christensen, L. O., Johannsen, P., Sinkjær, T., Petersen, N., Pyndt, H. S., & Nielsen, J. B. (2000). Cerebral activation during bicycle movements in man. *Experimental Brain Research*, 135(1), 66-72.
- Christova, M., Rafolt, D., Golaszewski, S., & Gallasch, E. (2011). Outlasting corticomotor excitability changes induced by 25 Hz whole-hand mechanical stimulation. *European Journal of Applied Physiology*, 111(12), 3051-3059.
- Colcombe, S., & Kramer, A. F. (2003). Fitness effects on the cognitive function of older adults: a meta-analytic study. *Psychological Science*, 14(2), 125-130.
- Coleman, J., & Roker, D. (1998). Adolescence. *Psychologist*, 11(12), 593-596.
- Corbett, D. B., Peer, K. S., & Ridgel, A. L. (2013). Biomechanical muscle stimulation and active-assisted cycling improves active range of motion in individuals with Parkinson's disease. *NeuroRehabilitation*, 33(2), 313-322.
- Cotman, C. W., Berchtold, N. C., & Christie, L. A. (2007). Exercise builds brain health: Key roles of growth factor cascades and inflammation. *Trends in Neurosciences*, 30, 464-72.
- Cotman, C. W., & Engesser-Cesar, C. (2002). Exercise enhances and protects brain function. *Exercise and Sport Sciences Reviews*, 30(2), 75-79.
- Cowley, P. M., Ploutz-Snyder L. L., Baynard T., Heffernan K., Jae S. Y., Hsu S., Fernhall B. (2010). Physical fitness predicts functional tasks in individuals with Down syndrome. *Medicine and Science in Sports and Exercise*, 42(2), 388-393.
- Diamond, A. (2013). Executive functions. *Annual Review of Psychology*, 64(1), 135-168.
- Dunn, D. M., & Dunn, L. M. (2007). *Peabody Picture Vocabulary Test* (4th ed). San Antonio, TX: Pearson.
- Emery, C. F., Shermer, R. L., Hauck, E. R., Hsiao, E. T., & MacIntyre, N. R. (2003). Cognitive and psychological outcomes of exercise in a 1-year follow-up study of patients with chronic obstructive pulmonary disease. *Health Psychology*, 22(6), 598.
- Ericson, M. (1985). On the biomechanics of cycling. A study of joint and muscle load during exercise on the bicycle ergometer. *Scandinavian Journal of Rehabilitation Medicine. Supplement*, 16, 1-43.
- Evans, D. W., & Gray, F. L. (2000). Compulsive-like behavior in individuals with Down syndrome: Its relation to mental age level, adaptive and maladaptive behavior. *Child Development*, 71(2), 288-300.
- Fernhall, B. O., McCubbin, J. A., Pitetti, K. H., Rintala, P., Rimmer, J. H., Millar, A. L., & De Silva, A. (2001). Prediction of maximal heart rate in individuals with mental retardation. *Medicine and Science in Sports and Exercise*, 33(10), 1655-1660.
- Fisher, B. E., Petzinger, G. M., Nixon, K. Hogg, E., Bremner, S., Meshul, C. K., & Jakowec, M. W. (2004). Exercise-induced behavioral recovery and neuroplasticity in the 1-methyl-4-phenyl-1,2,3,6-tetrahydropyridine-lesioned mouse basal ganglia. *Journal of Neuroscience Research*, 77 (3), 378-390.
- Hillman, C. H., Erickson, K. I., & Kramer, A. F. (2008). Be smart, exercise your heart: Exercise effects on brain and cognition. *Nature Reviews Neuroscience*, 9(1), 58-65.
- Hillman, C. H., Snook E. M., & Jerome G. J. (2003). Acute cardiovascular exercise and executive control function. *International Journal of Psychophysiology*, 48(3), 307-314.

- Holzapfel, S. D., Ringenbach, S. D., Mulvey, G. M., Sandoval-Menendez, A. M., Cook, M. R., Ganger, R. O., & Bennett, K. (2015). Improvements in manual dexterity relate to improvements in cognitive planning after assisted cycling therapy (ACT) in adolescents with down syndrome. *Research in Developmental Disabilities*, 45, 261-270.
- Jefferson, A. L., Paul, R. H., Ozonoff, A., & Cohen, R. A. (2006). Evaluating elements of executive functioning as predictors of instrumental activities of daily living (IADLs). *Archives of Clinical Neuropsychology*, 21(4), 311-320.
- Jobling, A., & Cuskelly, M. (2006). Young people with Down syndrome: A preliminary investigation of health knowledge and associated behaviours. *Journal of Intellectual and Developmental Disability*, 31(4), 210-218.
- Lanfranchi, S., Jerman, O., Dal Pont, E., Alberti, A., & Vianello, R. (2010). Executive function in adolescents with Down syndrome. *Journal of Intellectual Disability Research*, 54(4), 308-319.
- Lewis, C. L., & Fragala-Pinkham, M. A. (2005). Effects of aerobic conditioning and strength training on a child with Down syndrome: A case study. *Pediatric Physical Therapy*, 17(1), 30-36.
- Llorens-Martin, M. V., Rueda, N., Tejeda, G. S., Florez, J., Trejo, J. L., & Martínez-Cué, C. (2010). Effects of voluntary physical exercise on adult hippocampal neurogenesis and behavior of Ts65Dn mice, a model of Down syndrome. *Neuroscience*, 171(4), 1228-1240.
- Mendonca, G. V., Pereira, F. D., & Fernhall, B. (2010). Reduced exercise capacity in persons with down syndrome: Cause, effect, and management. *Therapeutics and Clinical Risk Management*, 6, 601-610.
- Myers B. A., & Pueschel S. M., (1991). Psychiatric disorders in persons with Down syndrome. *The Journal of Nervous and Mental Disease*, 179(10), 609-613.
- Nobrega, A. C., Williamson, J. W., Friedman, D. B., Araujo, C. G., & Mitchell, J. H. (1994). Cardiovascular responses to active and passive cycling movements. *Medicine and Science in Sports and Exercise*, 26(6), 709-714.
- Petzinger, G. M., Fisher, B., Hogg, E., Abernathy, A., Arevalo, P., Nixon, K., & Jakowec, M. W. (2006). Behavioral motor recovery in the 1-methyl-4-phenyl-1, 2, 3, 6-tetrahydropyridine-lesioned squirrel monkey (*Saimiri sciureus*): Changes in striatal dopamine and expression of tyrosine hydroxylase and dopamine transporter proteins. *Journal of Neuroscience Research*, 83(2), 332-347.
- Petzinger, G. M., Walsh, J. P., Akopian, G., Hogg, E., Abernathy, A., Arevalo, P. Jakowec, M. W. (2007). Effects of treadmill exercise on dopaminergic transmission in the 1-methyl-4-phenyl-1, 2, 3, 6-tetrahydropyridine-lesioned mouse model of basal ganglia injury. *The Journal of Neuroscience*, 27(20), 5291-5300.
- Piepmeyer, A. T. & Etnier, J. L. (2015). Brain-derived neurotrophic factor (BDNF) as a potential mechanism of the effects of acute exercise on cognitive performance. *Journal of Sport and Health Science*, 4, 14-23.
- Quaney, B. M., Boyd, L. A., McDowd, J. M., Zahner, L. H., He, J., Mayo, M. S., & Macko, R. F. (2009). Aerobic exercise improves cognition and motor function poststroke. *Neurorehabilitation and Neural Repair*, 23(9), 879-885.
- Reeves, R. H., Irving, N. G., Moran, T. H., Wohn, A., Kitt, C., Sisodia, S. S., & Davidsson, M. T. (1995). A mouse model for down syndrome exhibits learning and behaviour deficits. *Nature Genetics*, 11(2), 177-184.

- Ridgel, A. L., Kim, C. H., Fickes, E. J., Muller, M. D., & Alberts, J. L. (2011). Changes in executive function after acute bouts of passive cycling in Parkinson's disease. *Journal of Aging and Physical Activity*, 19(2), 87-98.
- Ridgel, A. L., Vitek, J. L., & Alberts, J. L. (2009). Forced, not voluntary, exercise improves motor function in Parkinson's disease patients. *Neurorehabilitation & Neural Repair*, 23(6), 600-608.
- Rihtman, T., Tekuzener, E., Parush, S., Tenenbaum, A., Bachrach, S. J., & Ornoy, A. (2010). Are the cognitive functions of children with Down syndrome related to their participation? *Developmental Medicine & Child Neurology*, 52(1), 72-78.
- Ringenbach, S. D., Albert, A. R., Chen, C. C., & Alberts, J. L. (2014). Acute bouts of assisted cycling improves cognitive and upper extremity movement functions in adolescents with Down syndrome. *Mental Retardation*, 52(2), 124-135.
- Ringenbach, S. D., Chen, C. C., Albert, A. R., Lichtsinn, K., & Alberts, J. L. (2013). Assisted cycle therapy (ACT): Implications for improvements in motor control. In M. Leitner & M. Fuchs (Eds.), *Motor Behavior and Control: New Research* (103-114). Hauppauge, NY: Nova Science Publishers, Inc.
- Ringenbach, S. D. R., Holzapfel, S. D., Mulvey, G. M., Jimenez, A., Benson, A., & Richter, M. (2016). The effects of assisted cycling therapy (ACT) and voluntary cycling on reaction time and measures of executive function in adolescents with Down syndrome. *Journal of Intellectual Disability Research*. epub ahead of print. doi: 10.1111/jir.12275.
- Shaw, E. G., Rosdhal, R., D'Agostino, R. B., Lovato, J., Naughton, M. J., Robbins, M. E., & Rapp, S. R. (2006). Phase II study of donepezil in irradiated brain tumor patients: effect on cognitive function, mood, and quality of life. *Journal of Clinical Oncology*, 24(9), 1415-1420.
- Smith, P. J., Blumenthal, J. A., Hoffman, B. M., Cooper, H., Strauman, T. A., Welsh-Bohmer, K., Sherwood, A. (2010). Aerobic exercise and neurocognitive performance: a meta-analytic review of randomized controlled trials. *Psychosomatic Medicine*, 72(3), 239.
- Tajiri, N., Yasuhara, T., Shingo, T., Kondo, A., Yuan, W., Kadota, T., ... & Date, I. (2010). Exercise exerts neuroprotective effects on Parkinson's disease model of rats. *Brain Research*, 1310, 200-207.
- Tillerson, J. L., Caudle, W. M., Reveron, M. E., & Miller, G. W. (2003). Exercise induces behavioral recovery and attenuates neurochemical deficits in rodent models of Parkinson's disease. *Neuroscience*, 119, 899-911.
- Valenzuela, M., & Sachdev, P. (2009). Can cognitive exercise prevent the onset of dementia? Systematic review of randomized clinical trials with longitudinal follow-up. *The American Journal of Geriatric Psychiatry*, 17(3), 179-187.
- van Uffelen, J. G., Paw, M. J. C. A., Hopman-Rock, M., & van Mechelen, W. (2008). The effects of exercise on cognition in older adults with and without cognitive decline: A systematic review. *Clinical Journal of Sport Medicine*, 18(6), 486-500.
- Warburton, D. E. R., Nicol, C. W., & Bredin, S. S. D. (2006). Health benefits of physical activity: The evidence. *Canadian Medical Association Journal*, 174(6), 801-809.
- Wegener, S., Redoblado-Hodge, M. A., Wegener, S., Redoblado-Hodge, M. A., Lucas, S., Fitzgerald, D., ... Brennan, J. (2005). Relative contributions of psychiatric symptoms and neuropsychological functioning to quality of life in first-episode psychosis. *Australian and New Zealand Journal of Psychiatry*, 39(6), 487-492.

- Wexler, B. (2007). Neurocognitive enhancement therapy with work therapy in schizophrenia: 6-month follow-up of neuropsychological performance. *Journal of Rehabilitation Research and Development*, 44(5), 761-770.
- Whitmer, A. J., & Banich, M. T. (2007). Inhibition versus switching deficits in different forms of rumination. *Psychological Science*, 18(6), 546-553.

BIOGRAPHICAL SKETCH

Name: Shannon Dora (Robertson) Ringenbach

Address: Sensorimotor Development Research Laboratory

Exercise Science and Health Promotion

School of Nutrition and Health Promotion

E-mail: shannon.ringenbach@asu.edu

Arizona State University

Web: <http://healthpromotion.asu.edu/directory/192246>

EDUCATION

1998, Ph.D., Purdue University, Department of Health, Kinesiology and Leisure Studies

1993, M.Sc., McMaster University, Department of Human Biodynamics

1991, B.Pe., McMaster University, Department of Physical Education

Professional Experience:

- 2013 – present Associate Dean, Barrett, The Honors College for the Downtown Phoenix Campus
- 2013 – present Affiliated Faculty, Cognitive Science, Department of Psychology, ASU
- 2005 – present Associate Professor, Program of Kinesiology, ASU
- 2001 - present Affiliated Faculty, Department of Psychology, ASU
- 2003 - 2005 Researcher, *The localization of Parkinsonian micrographia*, (Stelmach, G. E. S. (PI)), National Institute for Neurological Disorders and Stroke, ASU
- 1999 - 2005 Assistant Professor, Interdisciplinary Exercise Science Ph.D., ASU
- 1998 - 2005 Assistant Professor, Department of Kinesiology, ASU
- 1998 - 1999 Supporting Faculty, *Adaptive and Computational Aspects of Motor Coordination*, Flinn Foundation, Motor Control Laboratory, ASU
- 1996 - 1998 Researcher, *The Development of Bimanual Coordination*, Purdue Research Foundation, Purdue University
- 1993 - 1996 Research Assistant, *Prosodic Elements of American Sign Language* (Anne Smith, PI, and Howard Zelaznik Co-PI, National Institute of Health, NIDCD). Purdue University.
- 1993, Feb. Research Assistant, Quebec-Ontario researcher exchange grant, University of Montreal.
- 1991 - 1993 Teaching Assistant, McMaster University.
- 06 - 08, 1989-1991 Instructor, McMaster Sport Fitness School

Honors and Awards:

2015	Excellence in Research Award	NASPSA
2014	Apple Polisher Award	Devils Advocates ASU
2014	Super Mentor	BUILDing SCHOLARS Grant
2012	Outstanding Professor Award	Order of Omega
2011	Nominated for Professor of the Year	ASU Parents Association
2010	Nominated for teaching award	CLAS, ASU
2009	Special Recognition for Professor of the Year	ASU
2008	Featured as Arizona's Leading Women Scientists	Phoenix Woman Magazine
2006	Research Career Award	World Down Syndrome Congress
2005	Early Career Distinguished Scholar (nomination)	NASPSA
2003 - 2004	ASASU Centennial Professor (see internal grants)	Arizona State University
1997 - 1998	Outstanding Graduate Student in Scholarship/Service	Purdue University
1995 - 1996	Young Scientist Award	SCAPPS
1988 - 1991	Dean's Honor List	McMaster University
1991 - 1992	Centennial Entrance Scholarship	McMaster University
1991 - 1992	McMaster Graduate Scholarship	McMaster University
1989 - 1990	University Scholarship	McMaster University
1987 - 1988	McMaster Entrance Scholarship	McMaster University
1985 - 1986	Public Speaking University Entrance Scholarship	Optimist International

Publications (last three years)

1. Ringenbach, SDR, DHolzapfel SD, Mulvey, GM, UGJimenez, A, UGBenson, A., & Richter, M. (in press). The effects of Assisted Cycling Therapy (ACT) and voluntary cycling on reaction time and measures of executive function in adolescents with Down syndrome. *Journal of Intellectual Disability Research*. (Impact factor: 2.41)
2. ^DChen,C-C., & Ringenbach, S. D. R. (in press). Dose-response relationship between intensity of exercise and cognitive performance in individuals with Down syndrome: A preliminary study. *Journal of Intellectual Disability Research*.
3. Ringenbach, S.D.R., ^DHolzapfel, S.D., Mulvey, G.M., ^{UG}Jimenez, A., ^{UG}Benson, A., & ^DBirchfield, N. (in press). The effects of assisted cycling therapy (ACT) on reaction time, executive function and language fluency in adolescents with Down syndrome. *Research in Developmental Disabilities*.
4. ^DChen, C-C., & Ringenbach, S.D.R. (in press). A Pilot study of Test–Retest reliability of the Purdue pegboard test in adolescents and young adults with Down syndrome. *Journal of Motor Learning and Development*.
5. ^DChen, C.-C.(JJ), ^{UG}Kelsey, A., ^DMulvey, G.M., & Ringenbach, S.D.R. (in press). Examining the Davidson's Model via an exercise variable in individuals with intellectual disabilities, *International Journal of Developmental Disabilities*, 62(1), 70-75.

6. Ringenbach, S. D. R., ^{UG}Lichtsinn, K. C. & ^DHolzapfel, S. D. (2015). Assisted cycle therapy (ACT) improves inhibition in adolescents with autism spectrum disorder. *Journal of Intellectual and Developmental Disability*, 40(4), 376-387.
7. ^DChen, C-C, & Ringenbach S.D.R. (in press). The effect of an acute bout of aerobic exercise on cognitive control in adults with Down syndrome. *Mental Health and Physical Activity*.
8. ^DChen, C.-C., Ringenbach, S. D. R., ^{UG}Biwer, A., ^{UG}Riekena, A. (2015). Cerebral lateralization of the EEG during perceptual-motor integration in young adults with Down syndrome. *Brazilian Journal of Motor Control*, 9(1), 1-7.
9. ^DHolzapfel, S.D., Ringenbach, S.D.R., Mulvey, G.M., ^{UG}Sandoval-Menendez, A.M., ^{UG}Cook, M.R., ^{UG}Ganger, R.O., & ^{UG}Bennett, K. (2015). Improvements in manual dexterity relate to improvements in cognitive planning after assisted cycling therapy (ACT) in adolescents with Down syndrome. *Research in Developmental Disabilities*, 45-46, 261-270.
10. ^DChen, C-C., Crews, D., ^{UG}Mundt, S., Ringenbach, S.D.R. (2015). Effects of equine-assisted activity on frontal EEG asymmetry in young children with autism spectrum disorder: A pilot study. *International Journal of Developmental Disabilities*, 61(1), 56-59.
11. ^DChen, C.-C., Ringenbach, SDR, Crews, D, Kulinna, PH, and Amazeen, EL (2015), The association between a single bout of moderate physical activity and executive function in young adults with Down syndrome: a preliminary study. *J Intellect Disabil Res*, 59, 589–598. doi: 10.1111/jir.12163.
12. Ringenbach, S. D. R., ^{UG}Zimmerman, K., ^DChen, C.C., ^DMulvey, G. M., & ^DHolzapfel, S. D. Weeks, D. J., & Thaut, H. H. (2014). Adults with Down syndrome performed repetitive movements fast with continuous music cues. *Journal of Motor Learning and Development*, 2(3), 47-54.* Excellence in Research Award.
13. ^DChen, C-C., & Ringenbach, S.D.R (2014). The Association between sleep disorders and fine manual dexterity in young adults with Down syndrome. *Journal of Sleep Disorders: Treatment and Care*, 3(2), 1-4. doi: <http://dx.doi.org/10.4172/2325-9639.1000132>.
14. Ringenbach, S. D. R., ^{UG}Albert, A. R., ^DChen,J.J., Alberts, J. L. (2014). Acute bouts of assisted cycling improves cognitive and upper extremity movement functions in adolescents with Down syndrome. *Intellectual and Developmental Disabilities*, 52(5), 124-135.
15. ^DChen, C-C., Ringenbach, S.D.R, & ^{UG}Snow, M. (2014). Treadmill walking effects on grip strength in young men with Down syndrome. *Research in Developmental Disabilities*, 35(2), 288-293.
16. ^DChen CC, Ringenbach SD, ^{UG}Albert AR. (2014). Assisted cycling exercise improves fine manual dexterity in persons with Down syndrome. *Journal of Applied Research in Intellectual Disabilities*. 27(3), 264-272.
17. ^DChen, C-C (JJ), Ringenbach, S.D.R., ^{UG}Albert, A., & ^MSemken, K. (2014). Fine motor control is related to cognitive control in adolescents with Down syndrome. *International Journal of Disability, Development and Education*, 61(1), 6-15.
18. Ringenbach, S. D. R., ^MBonertz, C., & Maraj, B. K. V. (2014). Relatedness of auditory instructions is important for motor performance in persons with Down syndrome. *Journal on Developmental Disabilities*, 20(1), 84-91.

Chapter 82

ENDOCRINE FEATURES AND MANAGEMENT OF ENDOCRINE PROBLEMS IN DOWN SYNDROME

***Stefano Stagi^{1,*}, Elena Sandini¹, Elisabetta Lapi²,
Perla Scalini¹ and Maurizio de Martino¹***

¹Department of Health Sciences, University of Florence,

²Genetics and Molecular Medicine Unit,

Anna Meyer Children's University Hospital, Florence, Italy

ABSTRACT

Down syndrome (DS), well known as trisomy 21, occurs in one in one out of 700-1000 live births in all ethnic groups. DS is a very common cause of mental impairment and children with DS present with a variety of medical issues, including cardiovascular defects, endocrine problems, neurodevelopment disorders, hematological problems, gastrointestinal and sleep dysfunctions, visual and hearing impairment. It is well known that there is an association between DS and autoimmune disorders. Thyroid dysfunction (mostly autoimmune hypothyroidism) is the most typical endocrine disorder in individuals with this syndrome (affecting 0 to 66% of these patients). In these patients, in contrast with the general population, there are not substantial differences of incidence between sexes. It can be either congenital or acquired, and usually presents as a subclinical disorder. Hyperthyroidism as well is more commonly seen in people with DS than in the general population, and the phenotypic metamorphosis from hypothyroidism to hyperthyroidism is more frequently seen. Moreover it is described in the literature an increased frequency of shortness of stature, diabetes mellitus, and nutritional disorders (overweight and obesity) in children with DS. An increased oxidative stress, probably due to the over expression of the gene for Cu/Zn superoxide dismutase (*SOD1*) located in chromosome 21, is observed in DS people and this fact may play a role in the higher prevalence and severity of a number of clinical conditions linked with the syndrome, as well as the accelerated ageing observed in these individuals. An endocrine follow up is required for these individuals, in order to early detect any subclinical abnormalities and prevent as many consequences as possible.

* Corresponding Author's Email: stefano.stagi@yahoo.it (Dr. Stefano Stagi, MD, Anna Meyer Children's University Hospital, viale Pieraccini 24, Florence, Italy; Tel. +39-055-5662405; Fax +39-055-5662400).

INTRODUCTION

Down syndrome (DS; OMIM #190685) is a chronic and complex medical condition affecting multiple systems that exhibits a prevalence of 1 in 700 live births and results from an extra 21 chromosome or a duplication of a critical portion of this chromosome [1].

Subjects with DS show distinct physical and facial features such as round face, slanting palpebral fissures with flat nasal bridge and epicanthic fold, short broad hand, hypotonia, simian palmar crease, low set ears, a wide gap between 1st and 2nd toes, a prominent neck fat pad, and a protruding tongue [2]. However, these subjects have a high prevalence of congenital heart disease, neurodevelopment issues and presenile dementia, hematological problems until to leukemia, gastrointestinal and sleep dysfunctions, dislocation of cervical vertebrae and are at higher risk for developing disorders such as hypothyroidism, hearing and visual impairment, obesity, diabetes, precocious puberty, dyslipidemia and bone impairment [3]. Moreover the association between DS and autoimmune endocrinopathies is well studied. These disorders become increasingly frequent as children grow up, particularly during adolescence, and the onset of one is often followed by the development of others [4, 5].

The purpose of this chapter is to evaluate retrospectively the data of auxological and endocrinological parameters in patients with this genetic disorder. Relevant papers have been identified through systematic searches of the Pubmed, EMBASE and Cochrane databases. All published studies in the English and non English language concerning this disorder have been identified. Keywords in the literature search have been entered in all combinations. Searches included manually reviewing the reference lists of all original articles and all systematic review articles, with each study being evaluated for inclusion.

LENGTH, HEIGHT AND GROWTH HORMONE AXIS

DS is notoriously correlated to postnatal growth deficiency that finally leads to short stature [6]. In particular, growth velocity is most reduced between 6-9 months and 3 years of age but subsequently is almost normal [7]. However, the average length and height in prepubertal DS is around the 2nd centile in growth charts for the general population [8].

In this syndrome, the causes of this growth retardation are not known in most patients [9]. However, the main conditions participating to the poor growth are congenital heart diseases [10, 11], thyroid diseases [5, 12, 13], celiac disease [5, 14], pubertal disorders [15], feeding problems [16], and sleep related upper airway obstruction [17].

Many countries elaborated their own specific growth charts for DS [7, 18-22]. For instance, in Sweden specific growth charts were created and they found that mean birth length was 48 cm in both sexes in DS, and mean birth weight was 3.0 kg for boys and 2.9 kg for girls [20]. According to this study, final height in individuals with DS was precociously reached at 16 and 15 years for males and females, and the mean height was 161.5 cm and 147.5 cm respectively [20]. Head growth was impaired, resulting in -0,5 SDS for head circumference (Swedish standard) at birth, decreasing to -2.0 at 4 years of age [20]. Puberty tends to start early and the charts revealed a decreased pubertal growth rate in DS [20]. Interestingly European DS boys are taller than corresponding Americans, whereas European DS girls, although being lighter, have similar height to corresponding American girls [20].

Similarly Al Husain et al. [19] reported growth charts for DS Saudi children less than 5 years of age. Their data confirmed the impaired growth at most ages, showing a high proportion of underweight children with DS in the first two years of life, with a tendency to gain weight, also excessively, by the age of 3 years; half of the children with DS had a head circumference below -2 SD for age and sex [19].

A Dutch study, retrospectively evaluating the growth data in a DS population, demonstrated that there are three age periods when height differences between children with and without DS are particularly relevant: during pregnancy, during the first three years of life, and during puberty [23]. Based on these references health care professionals can focus on the critical age periods and provide an accurate follow up to optimize growth [23].

Regarding the short stature, many studies were conducted to understand the underlying causes in DS, and in particular if growth hormone (GH) deficiency may play a role in growth retardation in these patients [24].

For instance Castells et al. evaluated GH secretion through levodopa and/or clonidine stimulation tests and GH secretory patterns by the integrated 24 h, founding out that DS children present a reduced serum GH response to GH stimulation tests, or disparity in responses to the stimulatory tests and a low GH concentrations in 24 hours [25]. For this author, the data suggest that GH deficiency (GHD) may be implied in growth failure in patients with DS and that hypothalamic dysfunction may be the primary cause for GHD in this syndrome, as reported by the blunted response to levodopa and clonidine tests and the normal results after growth hormone releasing hormone (GHRH) plus arginine test [26].

However, not all of the studies agree with these results. In fact, Pueschel et al. [27] found that the response of GH secretion varies depending on different stimulation tests. In this study, all children tested were able to secrete normally GH ($> \text{or} = 10 \text{ ng/ml}$) at least during one of the tests performed, affirming that GH secretion was significantly lower if compared with levels in normal children only after clonidine administration [27]. This data may support the hypothesis that children with DS present a decreased GH secretion and, consequently, a reduced linear growth [27].

Anyway, with the strong suggestion that an inappropriate GH secretion may be implicated in growth retardation in DS, it was suggested to evaluate the effect of a recombinant human growth hormone (r-hGH) treatment. However, there is a big concern about safety, being GH therapy contraindicated in people prone to cancer development like children with DS.

For this reason many studies about GH therapy were conducted in the past, but nowadays they have a role especially in the study of mechanisms involved in growth retardation. In these regards, many data suggested that height and growth velocity improved after one year or more of r-hGH treatment. The benefits of GH therapy do not stop here, since r-hGH may result in a significant increase in the mean head circumference standard deviation score (SDS) at 12 months in these patients [28]. Bone age may increase during the first year of treatment, resulting correspondent with chronologic age. No significant side effects were recorded, but the long term risk of cancer should be taken into consideration [28].

Meguri et al. [24] confirmed the growth-promoting effects and safety of GH treatment in children with DS and GHD. In fact, in presence of GHD, short stature benefit from 3 years treatment with r-hGH, as the height SDS increased from -3.5 at start to -2.5 after three years of r-hGH treatment. In these patients, no new safety concerns were observed [24].

Moving on to effects of GH therapy besides linear growth, it was postulate that insulin-like growth factor I (IGF-I) may be involved in brain development, in addition to the well known

effect on stature [29]. On this hypothesis, Annerén et al. [6] aimed to study the long term effects of r-hGH on linear growth and learning disability in young children with DS. Fifteen children with DS aged 6 to 9 months were treated with GH for three years. Results were encouraging, since the mean height of the study group increased from -1.8 to -0.8 SDS (Swedish standard) during treatment, whereas that of an untreated DS control group decreased from -1.7 to -2.2 SDS [6]. Concentrations of serum IGF-I and IGFBP-3 became normal during GH treatment. Growth velocity declined after treatment stopped [6]. However no significant differences in head circumference were recorded during treatment, neither in psychomotor development [6].

In conclusion we can assert that growth in children with DS is notably different from that of healthy children. For that reason, the use of DS specific growth charts to properly evaluate and compare the growth of these children is highly recommended. In DS children suffering from a growth deficiency associated with a GH deficiency, even if GH therapy is not strongly recommended in these patients, this treatment appears to be efficacious and sure.

WEIGHT AND NUTRITIONAL HABITS

In the first years of life many patients with DS may show a failure to thrive [30]. The reasons of this fact are multiple, for exemple DS children present a remarkable rate of gastrointestinal tract abnormalities (noted in until 12%) most commonly consisting of duodenal atresia, Hirschprung disease, trachea-esophageal fistula, pyloric stenosis, annular pancreas and anal/rectal atresia [31, 32]. A dysfunctional gastrointestinal tract can cause an incorrect absorption of nutrients. Furthermore the gastrointestinal defects that are frequently discovered in these patients also bear responsibility for nutritional deficiencies, constipation and slow intestinal peristalsis, that are common findings in DS children and adolescents [16, 33]. Some DS children exhibit feeding difficulties in chewing and swallowing food and they prefer consuming foodstuffs made of a lower intake of fresh fruit and vegetables, and increased intake of simple carbohydrates that are easy to chew and swallow [16]. We need to consider also that inappropriate choices of foodstuffs on the basis only of children's tastes are commonly taken. Also this fact contributes to various nutritional deficiencies and a lack of regulating dietary ingredients as well as low dietary fiber intakes [16].

Finally, as causes of failure to thrive in DS there may be also dental features such as delayed or atypical tooth eruptions, agenesis, malocclusion, tooth decay and periodontal disease, that are frequent findings in these patients [34, 35].

Moreover, as well as other children and adolescents with intellectual impairment, DS patients are at increased risk for metabolic disorders and in particular overweight or obesity [36]. However, the prevalence of overweight and obesity in individuals with DS varies in literature. In a Dutch study of 1596 individuals with DS aged 0–26 years, the prevalence of overweight and obesity was reported as 25.5% and 4.2% for men and 32% and 5.1% for women, respectively [37]. In the USA, Rimmer et al. [38], in a sample of 81 DS adolescents (aged 12–18 years), found that 55% was overweight and 31.2% was obese. However, in another, retrospective, study from the USA, involving 283 adolescents and adults with DS, the authors reported a 56% prevalence of overweight for women and 45% for men, over 36% and 33% respectively in the general population [39].

There are many factors that are known to influence obesity in DS, both metabolic and environmental. They range from thyroid dysfunction and decreased basal metabolic rate, along

with hypercholesterolemia, increased leptin levels and deficiencies of vitamins and minerals to low levels of physical activities [40].

Hyperphagia as well seems to play a role in these children. Hyperphagia is known to have a primary role in Prader Willi Syndrome (PWS), but it has not been well investigated in DS. This is the reason why Foerste et al. [41] analyzed this symptom through questionnaires proposed to 52 young people aged 6–18 years, equally classified in DS (17 patients), PWS (16 patients) and lifestyle-related obesity (LRO: 19 patients). As hypothesized, the PWS group scored high levels of hyperphagia and the LRO group had the lowest, but it was pointed out that DS group presented food-seeking behaviour more pronounced than in the LRO group and the latter spent more hours per week engaged in physical activity [41].

There is a metabolic predisposition to overweight that is objective, and, unfortunately, less easy to control. Interestingly, DS patients show high leptin levels, a hormone that regulates hunger and satiety and appetite [16].

However children and adolescents with DS have been shown to take part in less vigorous physical activity and had higher BMI levels compared with their siblings [42–45]. The basis of this may be medical conditions such as cardiac issues and precocious fatigability but also unawareness or unavailability of adequate gym programmes within the community. Moreover practicing sport and activities out of a familiar contest may constitute another hindrance to physical activity for these fragile individuals, in which unwillingness of practicing sport may play an important role [46–48]. Lack of physical activity may be a major contributor to the development of overweight within the DS population, but what is important is that it is, differently from others, a preventable risk factor.

Another factor that can affect DS subjects is tendency to depression and anxiety that significantly affects the relationship with food, such as making inappropriate food choices, tasting and hyperphagia or compulsive eating, and this aspect may become more evident during adolescence. It would be worthwhile to provide further education for families with a DS child to inform about the risks of over-eating, especially in the contest of primary care, and to be included in a health policy in which local dieticians may participate to sustain exercise programmes and physical activity in patients with intellectual disabilities [16, 41].

Finally, many studies demonstrated reduced dietary intakes of fiber and some vitamins and minerals in DS subjects (in particular vitamins A, B and C along with calcium and zinc) [16]. DS children showed deficiencies in the B vitamins group (i.e., B1, B2, B6, B12 and folic acid), which are implicated in intellectual development, notoriously inadequate in DS children. Vitamin B1 deficiency causes weakness, constipation and decreased mobility, whereas B2 deficiency results in mucocutaneous alterations. Vitamin B6 deficiency gives rise to mental retardation and a lack of concentration. Deficiencies of vitamins B6, B12 and folic acid are linked to abnormal levels of homocysteine in the bloodstream of DS children [16].

We conclude assessing that prevalence of obesity and overweight in patients with DS is increased compared to general population and there we strongly underline the need for a health policy and education about appropriate nutritional behavior for individuals with intellectual disability to prevent sedentary activity and the burden of overweight and obesity.

PUBERTY AND FERTILITY

In patients with DS, puberty starts in a range of age similar to that of the general population. Anyway there are several case reports of a syndrome, called Van Wyk and Grumbach syndrome, that includes precocious puberty in children who were also identified to suffer from hypothyroidism, and many case reports of DS children with this syndrome have been reported in literature, providing evidence for a linkage between these two conditions [15, 49].

Van Wyk and Grumbach syndrome was first reported in 1905 by Kendle, who described the case of a 9-year-old girl with menarche at age 5, fully developed breasts, and the clinical symptoms of primary hypothyroidism, who fully recovered after treatment with thyroid extract (in particular her growth moved on, her menstruation stopped, and her symptoms of cretinism resolved) [50]. However, in 1960 Van Wyk and Grumbach well analyzed this syndrome, that is a condition consisting of primary hypothyroidism, precocious puberty, and solitary or multiple ovarian cysts which can become massive in certain cases [51].

The review of the literature reveals that the majority of these patients have bilateral rather than unilateral masses [15, 49]. Sexual precocity usually is associated with increase in linear height, acceleration of bone maturity and epiphyseal fusion, which finally leads to short stature. On the contrary, a long standing, untreated, primary hypothyroidism has classically been associated with retardation of linear growth and delayed puberty. Differently, in the patients with Van Wyk and Grumbach syndrome, untreated hypothyroidism leads to growth delay, delayed skeletal maturation, and precocious puberty, in presence of a retardation of bone age [15, 49]. This syndrome occurs both in boys and girls. In these patients, there is usually pituitary enlargement, demonstrable on computerized tomography (CT) scan or MRI and this fact could mislead to the case of a pituitary tumor. Pituitary hyperplasia may be identified and is probably due to long standing thyrotrope stimulation in response to the decreased thyroid hormone [15, 49].

Girls present with early thelarche, enlarged labia minora, and estrogenic changes in vaginal smear, with or without appearance of pubic hair [52]. There may be irregular vaginal bleeding and solitary or multiple ovarian cysts. Serum prolactin, follicular stimulating hormone (FSH), levels are elevated and luteinizing hormone (LH) levels are low or normal [53]. A similar precocity is known to occur, even if less frequently, in males as well. Boys present with enlargement of the testes because of increase in size of the seminiferous tubules, hence as a direct effect of hypothyroidism on the prepubertal testes leading to over proliferation of Sertoli cells. Signs of virilization and Leydig cell maturation are absent. Plasma testosterone concentration is prepubertal [54].

An explanation of this syndrome could be that the low levels of T_3 and T_4 hormones caused by primary hypothyroidism elicit a TRH secretion, that cause an increase in TSH levels, prolactin, LH and FSH. However, further explanation are required for sexual precocity in primary hypothyroidism, as in uncomplicated juvenile hypothyroidism puberty development is usually delayed. Interestingly, in these patients the FSH level is elevated, LH level is either low or normal. The increased FSH release and high FSH/LH ratio is thought to cause the increased ovarian estrogen secretion in girls and this is different from normal puberty, where LH raises to higher levels and the LH/FSH ratio is high. However, a more accepted theory is the fact that TSH can act on the FSH receptor. TSH, FSH, LH and hCG, share a common beta subunit. It is believed that large amounts of TSH can activate the FSH receptor due to the presence of this

common subunit. Gordon et al. demonstrated that the elevated estrogen levels seen in this disorder come directly from the ovaries, being the patient's serum estradiol level significantly lower than estradiol directly aspirated from the ovary [55].

Regarding the fertility, it is unclear whether people with DS can be considered infertile. There have been cases of males and females with this syndrome that successfully reproduced. In males, FSH and LH levels are normal during the beginning phase of puberty, but are elevated post puberty. The levels of testosterone appear normal throughout puberty. The size of penile length and the size of the testis in DS people are below the average, suggesting a partial gonadal deficiency leading to infertility. In females normal reproductive organs have been observed and there are not abnormalities in both external genitalia. Females usually have regular menstrual cycles. FSH and LH hormone levels are in normal range potentially leading to successful reproduction. Another explanation for the infertility in people with Trisomy 21 could be the lack of knowledge about sexual intercourse, that may contribute in patients there often do not have objective problems to their reproductive system [56].

THYROID FUNCTION AND DISORDERS

These are just few examples where a genetic syndrome and thyroid disorders show a close links as that in DS. In fact for a long time many researchers thought that DS was due to a kind of thyroid disease (TD). In 1866 John Langdon Haydon Down (1828-1896), superintendent of the Earlswood Asylum for Idiots in Surrey (England), made his contribution by differentiating children with cretinism due to congenital hypothyroidism (CH) from children with DS already [57], however in 1896 Telford Smith noted the resemblances between DS and CH, theorizing that they were two aspects of the same problem (Table 1) [58].

However, the clinical resemblance between the two patient groups may make difficult to detect developing TD in children or adults with DS [59]. For this reason optimal care in DS requires close follow-up and early intervention to minimize disease and dysfunction related to TD [3]. Moreover, in the 1970s, individuals who lived longer than 45 were rare, however, today over half the individuals born with DS live into their 50's, 40% of them into their 60's and 13% to the age of 68 [60].

In fact, hypothyroidism can affect a child with DS at any age [61] and is sporadically detected by routine neonatal thyroid screening [61]. Thus, the annual preventive medical protocols for individuals with DS of all ages always should include a blood test for TD [59].

Clinical forms of hypothyroidism found in people with DS include transient and primary hypothyroidism, pituitary-hypothalamic hypothyroidism, chronic lymphocytic thyroiditis (Hashimoto's thyroiditis, HT), and autoimmune hyperthyroidism (Graves' disease, GD) [62].

Many data suggest that the percentage of individuals with DS at risk for TD gradually increases with age. At birth 0.7% of the infants with DS have persistent primary CH [63] whereas the rate of adults with hypothyroidism rises to at least 12% [64].

Table 1. Comparison of physical characteristics of subjects with Down syndrome and hypothyroidism

Characteristic	Down syndrome	Hypothyroidism
Appearance	Dull, chubby	Dull, chubby
Head	Microcephalic	Normal
Tongue	Large	Large
Nasal Bridge	Underdeveloped	Underdeveloped
Eyes	Slanted	Not slanted
Neck	Short	Short
Heart	Murmur (AV canal)	Murmur (thick valve and septum)
Abdomen	Protuberant umbilical hernia	Protuberant umbilical hernia
Neuromuscular	Hypotonia	Hypotonia
Skin	Dry	Dry
Extremities	Short, transverse palmar crease	Short, no transverse

Adapted by [58].

Congenital Hypothyroidism (CH)

CH due to athyreosis has been occasionally described in infants with DS [65, 66]. The incidence of persistent primary CH in infants with DS is 1:141 (28 times more frequent) compared to 1:3800 in the general population [63]. Cutler et al. studied 49 DS children less than three years of age and found that 6% had CH [67]. He noted also that 27% of them already had mildly increased TSH levels [67]. In many patients, in fact, the neonatal hypothyroidism is transient and follow-up is always needed [63, 68]. In fact, DS patients are more likely to have iodine exposure in the neonatal period as a result of cardiac surgery and gastrointestinal contrast studies, potentially causing transient TSH elevation (Wolff–Chaikoff effect) [69].

The exact reason of the high incidence of hypothyroidism in neonates with DS is not fully understood. In the past several data suggested that there is an increased incidence of TD in mothers of DS newborns [70]. However, other studies did not find a consistent increase in the prevalence of thyroid antibodies in their mothers [71].

Recent data seem to suggest that in DS children we observe an idiopathic mild plasma TSH elevation (4–10 mIU/l), up to 100%, during their first 6 months of life, that decreases with age [72], and that DS patients have T₄ levels substantially decreased compared to the general newborn population [72]. The clinical significance of this dysfunction is not well known. The existence of a mild hypothyroid state in DS newborns supports the hypothesis of the presence of a DS-specific thyroid function's regulation [72]. It is important to note that, without appropriate treatment, hypothyroidism leads to relevant developmental delay [60]. However van Trotsenburg et al. [73] has demonstrated that L-T₄ treatment in DS newborn may improve growth and neurological development in the first 2 years of age.

Finally, it is interesting that, as in CH patients, both DS with CH and subclinical hypothyroidism showed a significantly higher presence of gastrointestinal malformations [74], focusing on the importance of organizing a particular follow-up for DS presenting gastrointestinal anomalies and CH.

Hypothyroidism

This thyroid dysfunction, represented commonly by HT, is the most typical endocrine abnormality in patients with DS. The estimated lifetime prevalence rate of thyroid dysfunction in DS varied widely in different studies, depending on variations in population size, age, laboratory assays and definitions of thyroid dysfunction used. This is why the rate of thyroid dysfunction ranged from 30% to 93% of cases of hypothyroidism [75]. During childhood, the problem of diagnosis of hypothyroidism by clinical criteria alone remains a difficult one [59]. The importance of annual preventative medical blood testing cannot be overstated [59]. Moreover, thyroid function assessment must be carried out periodically, annually after the first year of age [3, 76]. By the time that the prominent features of severe hypothyroidism (growth deviation from a previous channel of growth, plateauing of intellectual growth, increased lethargy, constipation and eventually the development of myxedema) are seen in the patient, DS children may already have major adverse effects of the disease process [58], such as myxedema coma [77], vaginal bleeding [49], obstructive sleep apnoea [17], or pericardial effusion [78].

Goitre, which can be seen with hypothyroidism, hyperthyroidism or even euthyroidism is one of the more obvious clinical clues [79].

The first case of DS with HT was reported by Benda [80]. In 1985, Pueschel and Pezzullo reported the thyroid function of 151 children with DS and their sibling controls [81]. In this series, 27% had an abnormality of TSH, T₄ or both. They noted that there were higher TSH concentrations in adolescents and decreasing T₄ levels with the advancing age of the patients [81]. These authors also raised the question of whether the decline of intelligence quotients described in the literature over time in people with DS might be, in part, due to inadequate thyroid function [81]. However, another study has found no differences in intelligence quotients between children with and without elevated TSH [82].

Loudon et al. published a series of 116 home-based children, discovering three hypothyroid, one hyperthyroid and 29% with thyroid antibodies [83]. The authors noted that transient increases in TSH levels seemed common in these children, particularly during periods of intercurrent illness [83].

In 1991, Pueschel updated his series to 181 patients and found 6% of children had both high TSH and low T₄ [84]. In 1993, a five year longitudinal study was published of 101 children with a mean age of five years and three months [82].

Thyroid auto antibodies are also frequently positive, and Hashimoto's thyroiditis is the most common cause of hypothyroidism in DS. The risk of progression appears to be similar to non-DS patients [69]. DS patients may be incline to manifest over time a phenotypic metamorphosis from HT to GD and to subsequently fluctuate from hyperthyroidism to hypothyroidism [13].

The frequency of antithyroid antibodies (ATA) among patients with DS and hypothyroidism increases with age. ATA associated with hypothyroidism are significantly less common in DS children under 8–10 years of age. However, some studies reported ATA-positive DS children under 3 years of age [85, 86]. Several studies observed a much higher prevalence of thyroid disorders in children with DS and Type 1 Diabetes (T1DM) than that usually described in DS without diabetes [87, 88]. This fact therefore suggests that the subgroup of trisomic patients with T1DM carries an even higher risk of aggressive autoimmune disease.

Table 2. Some of more significant studies of prevalence of thyroid disorders in Down syndrome

Author (ref)	Year	Subjects	Age range (yrs.)	Male/Female	Goiter (%)	Hypothyroidism (overt/subclinical)		Hyperthyroidism		Thyroid autoimmunity	
						No	%	No	%	No	%
Hollingsworth et al. [93]	1974	60	9-65	39/21	13	11	18	2	3.3	51	85
Murdoch et al. [94]	1977	82	19-65	44/38	-	13	16	1	1.2	11	13.4
Sare et al. [95]	1978	121	13-48	81/40	18.1	21	17	3	2.5	40	33
Korsager et al. [64]	1978	24	41-60	8/16	0	3	12.5	0	0	8	33
Quinn [96]	1980	49	8-59	-		3	6				
Lobo et al. [97]	1980	101	5-47	64/37	-	7/9	6.9/8.9	-	-	30	29.7
Samuel et al. [98]	1981	54	9-12 days	20/34		10	18				
Hughes et al. [99]	1982	38	16-65	27/11		8	21				
Vladutiu et al. [100]	1984	42	18-64	22/20		23	55				38
Coleman & Abbassi [61]	1984	206	<18	-		16	8				
Loudon et al. [83]	1985	116	0.75-19.83	67/49	-	3	2.5	1	0.8	28/65	43
Pueschel & Pezzullo [81]	1985	151	3-21	92/59	1.9	31	20.5	-	-	47	31.1
Cutler et al. [67]	1986	49	0.4-3.0	24/25	-	16*	32	1	6.6	1	6.6
Kinnell et al. [101]	1987	111	22-72	56/55	0	10	9	2	1.8	33	29
Mani [102]	1988	55	24-67	32/33	-	12	21.8	0	0	11	19
Friedman et al. [103]	1989	138	2-59	72/66	1.4	11	8	2	1.4	26/66	29.3
Dinani & Carpenter [104]	1990	106	20-67	61/45	-	36	33.9	1	0.9	61/106	34
Zori et al. [105]	1990	61	0.4-48	34/27	-	38	57.6	2	3.2	17	27.8

Author (ref)	Year	Subjects	Age range (yrs.)	Male/Female	Goiter (%)	Hypothyroidism (overt/subclinical)		Hyperthyroidism		Thyroid autoimmunity	
						No	%	No	%	No	%
Pozzan et al. [106]	1990	108	0.25-38	55/53	-	38	35.2	2	1.8	13	12
Selikowitz [82]	1993	101	0-11.33	49/52	-	11**	10.1	-	0	3	2.9
Rubello et al. [107]	1995	344	1-53	182/162	1.1	3/112	0.9/32.5*	2	0.6	62	18
Toledo et al. [108]	1997	105	0.25-20	50/55	-	54	51.4	0	0	9	8.5
Ivarsson et al. [85]	1997	70	1-19	32/38	-	17	24	1	1.4	27	39
Karlsson et al. [86]	1998	85	1-25	42/43	-	28	33	2	2.3	18	21
Tüysüz & Beker [109]	2001	320	0-12	165/155	0	90 [§]	28.1	0	0	-	-
Ali et al. [110]	2002	58	11-42	40/18	-	29	50	1	1.7	29/56	51.7
Wong et al. [111]	2004	78	0.2-18.5	52/26	-	25	32	1	1.3	3	3.8
Gibson et al. [112]	2005	122	6-14	-	-	24	19.6	-	-	8/101	7.9
Mak et al. [113]	2006	351	0-19	200-151	-	85	24.2	8	2.3	25/67	37.3
Unachak et al. [114]	2008	140	0-13.75	68/72	-	53	37.9	3	2.1	2	1.4
Rashid et al. [115]	2009	80	0.5-18	52/28	-	28	35	-	-	6	7.5 [#]

* Three have congenital hypothyroidism; one has CH

° Longitudinal study

§ Six patients have diagnosed congenital hypothyroidism and two transient hyperthyrotropinemia

* Patients with clinical and subclinical hypothyroidism

Only anti-TPO was carried out

Zinc deficiency, which appears to be more common in DS children, can affect endocrine and immunological function. Normalisation of serum zinc and TSH was reported in DS patients following 4 months of oral zinc in a non-randomised study and this was later replicated [89]. Conversely, a subsequent study found no difference in thyroid function between DS patients who were zinc deficient and those who were depleted [90].

Approximately 40% of children with DS have congenital heart disease. Some studies of adults with subclinical hypothyroidism (SH) showed abnormalities of myocardial function, reversible with L-thyroxine treatment [91]. A small study of 16 DS children with prolonged SH, showed no difference in myocardial function compared to age and body-surface area matched euthyroid DS controls [92].

In Table 2 we reported some of the more significative studies regarding thyroid function in DS.

During the course of the study, in addition to the three children who entered the study with abnormal thyroid tests, eight more children developed elevated TSH, in all cases with normal T₄ and T₃ tests [82]. Only one of these eight then further developed uncompensated hypothyroidism; this patient had thyroglobulin and microsomal thyroid antibodies [82].

Mildly elevated plasma thyrotropin concentrations in the absence of signs of thyroid autoimmunity are also common [116]. There is little knowledge about the reason of this phenomenon and it is a common therapeutic dilemma. Konings et al. reported the presence of a normal bioactivity of TSH in DS [116]. Other authors have suggested that reduced thyroid function in patients with DS may be linked in some way to the low serum levels of selenium found in these patients [117].

Moreover, autoimmune thyroid disease (ATD) is very uncommon in young children, but has been recognised in association with DS [118]. The pathogenesis of autoimmune thyroid disease is complex and, to date, there is no single unifying hypothesis to explain the changes in immune function and the increased incidence of ATD in patients with DS [119]. There is an interest in the possible effects of over expression of proteins whose genes are encoded on chromosome 21 and which participate in the regulation of the immune response [119]. Increased expression of chromosome 21 gene products may be directly responsible for altered immune function and predisposition to autoimmune disease [118].

The increased frequency of thyroid dysfunction and the mild presentation of hypothyroidism in DS can be responsible of unjustified medically treatment of thyroid diseases in these patients [120]. However, other data seem suggest a low adherence to national guidelines for thyroid screening [121].

The risk of thyroid dysfunction increases with age in patients with DS [94]. In a study conducted in 1990 that involved 106 adults from 20 to 67 years of age (with an average age of 38 years), 40% had blood test that revealed an abnormal thyroid function. However, it should be noted that not all patients with abnormal test results had an active disease process: only seven had active hypothyroidism and one had thyrotoxicosis [104].

Hyperthyroidism. While hypothyroidism is the most frequent alteration in DS patients, we must consider that hyperthyroidism as well does occur in DS [110]. In fact, the incidence of thyrotoxicosis in DS patients exceeds that in the general paediatric population, which varies from around 0.1:100,000 in childhood to 3:100,000 in adolescence [75]. After the first described case of thyrotoxicosis in a 22 yrs. old DS female in 1946 [122], many case reports were published in the literature reporting hyperthyroidism in DS [67, 83]. Before the study of Goday-Arno et al., [123], a total of 46 cases of hyperthyroidism have been previously reported

in DS [123]. The more significant data of prevalence of hyperthyroidism are reported in Table 2.

Hyperthyroidism associated with DS can occur as a result of an autoimmune diffuse enlargement of the thyroid gland (GD) rather than the presence of a toxic adenoma or toxic multinodular goiter. However, exophthalmos appears to be less frequent [110].

In children with DS, clinical recognition of hyperthyroidism can be difficult because symptoms may be hidden [124]. The shortness and chubbiness of the neck makes it more difficult to detect thyromegaly, so a careful palpation is necessary [110, 123].

In DS, suggestive findings of hyperthyroidism also include weight loss, hyperactivity, diarrhoea, nervousness and goitre. The thyroid examination revealed frequently diffuse thyroid enlargement in all patients. All patients had undetectable serum TSH concentrations with elevated serum FT₄ concentrations and elevated serum total T₃ concentrations. Thyroid antibodies were present in 11 of 12 patients (92%). A thyroid scan showed diffuse thyroid uptake in all patients, which was suggestive of GD.

At the present time, there is no clear consensus on the best way to treat hyperthyroidism in children with DS and in most of the cases described details of the therapy were frequently unclear [123]. However, there are three possible treatments of hyperthyroidism. One treatment is aimed at blocking the action of the TH on body tissues. This involves the use of antithyroid drugs, and are often the first treatment used [123]. However, almost all of these drugs can cause significant side effects.

A second treatment is surgery to remove part or the entire thyroid; and then the child or adult should use drugs for thyroid replacement. The third treatment is the use of radioactive iodine, which destroys the thyroid's ability to produce thyroid hormone. The patient then takes replacement thyroid hormone. However, radioactive iodine treatment is not often used in children because of the risk of thyroid cancer.

Thyroid cancer. In literature we found only one benign neoplasm, an oncocytic adenoma [124] and four thyroid cancers: an unspecified cancer, one papillary and one follicular carcinoma [125-127], and one thyroid lymphoma [128].

So, although thyroid disorders such as thyroiditis and goiter, decreased serum selenium, celiac disease, and high body mass index, are all factors potentially favouring an enhanced risk of thyroid cancers, this entity seems to be exceptionally reported in DS, such as confirmed also by the Oxford Record Linkage Study about the data of 1453 individuals with DS [129]. However, thyroid neoplasm appears at a younger age than other carcinomas and life expectancy has greatly improved in people with DS [129]. Finally, it is particularly important to underline that thyroid nodules are rarely reported in people with DS [125].

The review of the literature showed that thyroid neoplasm are under-represented in DS. Moreover a particular distribution of histological types of cancer in DS was observed, where carcinomas are under-represented and lymphomas are more frequent [125, 129].

In conclusion, thyroid diseases (with the exception of thyroid cancer) occur with greater frequency in individuals with DS than in control populations. These patients are at risk from infancy throughout adult life. Therefore, follow-up of growth patterns, searching clinical signs of diabetes, thyroid function monitoring, screening for thyroid auto antibodies and measurement of serum concentrations of antiigliadin and antiendomysium self-antibodies should be started from infancy to prevent further deterioration of growth and mental development. Interpretation of thyroid test results always should be made by a specialist who is aware that elevated TSH levels sometimes are transient in DS.

BONE MINERAL STATUS AND METABOLISM

Among all the disorders described in DS patients, there is also impairment of bone metabolism. DS children are notoriously prone to develop osteoporosis, bone fragility, and related fractures and their vitamin D levels are usually lower than in the general population.

Since life expectancy of DS patients significantly raised, it is important to let them reach a good value of peak bone mass to prevent osteoporosis and bone disorders during the elderly [130]. Several studies have reported that DS patients have a lower level of bone mass than people without DS, but the reason of this disorder remains still unclear. Lifestyle factors like alimentary choices and physical activities may play an important role, but further information need to be gathered whether or not specific effects of chromosome 21 genes are also implied [131, 132].

González-Agüero et al. affirmed that DS subjects have a clear tendency to reach lower bone mass density (BMD) in several regions of their bodies also in pediatric age when compared with age and sex-matched subjects without DS. Furthermore they provided evidence that young females with DS are poorer at acquiring bone mass than young males with DS [133]. 30 patients with DS were examined by McKelvey et al. using dual-energy X-ray absorptiometry (DXA) to measure BMD, revealing that 53.3% showed a BMD ≤ -2 SDS. The authors also demonstrated that both bone formation and reabsorption were suppressed in DS compared with controls, indicating low bone formation and decreased bone turnover as the primary causes of the low bone mass observed in these patients [132]. A similar work was performed by Wu, who analyzed BMD by DXA between preadolescent boys with and without DS and confirmed that patients with DS have lower BMD and bone mineral content (BMC), also localizing in the pelvis the first site to show bone impairment in DS [134]. These results were partially confirmed by peripheral Quantitative Computed Tomography (pQCT), which is another instrument that provides information about volumetric bone mineral density (vBMD). BMD was found to be higher in determined skeletal sites in DS children according to González-Agüero et al., even if they have a higher risk of suffering bone fractures due to a decreased bone resistance to load bending or torsion [135].

Many factors are thought to be related to bone disorders in DS, from hypotonia, that is a well-established characteristic of DS and may limit physical activity, to low amounts of physical activity indeed, or poor calcium and vitamin D intake due to alimentary choices and malabsorption, celiac disease, and hormonal factors [75, 136].

It is believed that the combination of all the risk factors is the cause of bone alterations in DS. In fact neither the increase of physical activity alone, nor the increase of calcium intake alone, demonstrated a significant effect on BMD. Then authors observed patients with DS who received a higher calcium intake and, at the same time, who were subjected to a specific program of physical activity, and found that these subjects showed a benefit in terms of BMD from both the intervention together [137]. Another aspect that needs to be considered are lower levels of vitamin D in the bloodstream reported in DS, particularly in subjects with obesity and autoimmune diseases, as demonstrated by Stagi et al. in a longitudinal study. In the same study parathyroid hormone levels were found to be significantly higher in DS subjects compared to controls and cholecalciferol supplementation improved levels of 25(OH)D levels even if the progress was less remarkable than that observed in controls. This is the reason why it would be worthwhile to give DS patients with obesity and autoimmune diseases higher cholecalciferol

supplementation [138]. Physical activity has a well-known effect on reducing the risk of developing future bone fragility, fractures and osteoporosis, as proved in a study in which adolescents with DS who spent more time practicing sport, had a higher BMD Z-score at the hip, and therefore less risk for osteoporosis [136]. Ferry et al. demonstrated that physical activity produces different effects in some skeletal segments compared to others. In fact one year of training increased bone mineral content at lumbar spine and total hip, but influenced BMD only at lumbar spine when measured with dual X-ray absorptiometry (DXA). Moreover there was no evidence of bone improvement when the bone tissue of the same individuals was analyzed with ultrasound techniques. In Down syndrome bone is able to adapt to mechanical stimulation even if the responsiveness seems to be reduced than that currently reported in the literature for normal individuals [139]. Thus DS are prone to osteoporosis, impaired bone mass and fragile bones. Many factors prevent correct bone turnover and gain in DS, either preventable or not. Therefore it is required a multidisciplinary approach that includes programmes to promote physical activity, an optimal calcium and vitamin D intake and periodic endocrine evaluations of bone mineral density [140].

AUTOIMMUNE DISEASES

DS patients present several immunological disorders, a high rate of infections, cancer and autoimmunity, and there is a well recognized association of DS with autoimmune diseases, in particular, thyroid disorders, type 1 diabetes and celiac disease [141]. Da Rosa Uttyama et al. investigated the prevalence of auto antibodies of various types (such as anti-mitochondrial, smooth-muscle, liver-kidney microsomal, nuclear, gastric parietal cell and neutrophil cytoplasmic antibodies and rheumatoid factor) in 150 Caucasoid DS children and adolescents and 105 healthy children, finding that 28.6% of DS patient revealed positivity to at least one autoantibody in comparison with 8% of the controls. RF was frequently detected (28% of the patients and 6.7% of the controls), even if nobody showed clinical evidence of rheumatic disease, so that the meaning of this finding remains unclear, being hypothesized either an expression of deregulation of immune system or an earlier marker of rheumatic diseases in these patients [142].

Thyroid autoimmunity is a notorious possible finding in DS, as previously seen, but also celiac disease, type 1 diabetes and other autoimmune disorders are common to observe. Autoimmunity in particular is very frequent in DS, especially if compared with other genetic disorders like Williams-Beuren Syndrome [5].

At the root of the tendency to develop autoimmunity in these patients there is an immunological disorder, that some studies attribute to a dysregulation of autoimmune regulator protein (AIRE), a transcription factor located on chromosome 21, that plays a crucial role in autoimmunity by regulating promiscuous gene expression (pGE). In DS individuals in fact AIRE expression is significantly reduced in thymuses compared with controls and decreased expression of AIRE was found to be accompanied by a reduction of pGE [143]. In these subjects the ability to proliferate of DS T cells doesn't seem to be altered significantly, however there is an over expression of T (reg) population that, in spite of this, show a reduced inhibitory potential compared to healthy controls. The defective inhibitory activity may partially explain itself the higher frequency of autoimmune diseases [144]. Another difference in DS children is

that the thymus was found to be smaller and with an abnormal structure since birth in DS subjects. Hypergammaglobulinaemia of immunoglobulin IgG and IgA after the age of 5 years, with high levels of IgG1 and IgG3 and low levels of IgG2 and IgG4, may be present in DS. In DS subjects both B and T lymphocytopenia (that concerns CD4⁺ helper as well as CD8⁺ cytotoxic T lymphocytes) is very frequent in DS, and the physiological expansion during the first year of life is not observed, suggesting an abnormal reaction to antigenic stimulation [145-147].

CONCLUSION

Our study aimed to focus on the major endocrine problems that a DS patient may suffer from. It is well known that DS children are prone to develop thyroid disorders, mainly hypothyroidism, but also pubertal disorders, overweight and obesity, as well as short stature and low bone mineral density are conditions that we frequently observe in DS individuals. The average life expectancy of children with DS considerably increased over the past few years. This is a sign that health care professionals learnt how to deal with their problems and harmonised the required treatment. At the same time this fact implies that we have to take into consideration also long term consequences of medical issues that might come up. Finally we conclude assessing the importance of a regular follow up examination that should include thorough physical exam and blood tests. A strong emphasis is also put on the need to secure a balanced diet for DS people, being overweight and obesity conditions that often affect these patients that are, at least partially, preventable. From this perspective, implementing physical activity is absolutely crucial, and health and education programmes must be targeted at the families with DS children to inform of the benefits of a correct lifestyle. In this regard, not only the endocrine specialist, but also primary health care plays an important part.

REFERENCES

- [1] Weijerman, M. E., van Furth, A. M., Vonk Noordegraaf, A., van Wouwe, J. P., Broers, C. J. & Gemke, R. J. (2008). Prevalence, neonatal characteristics, and first-year mortality of Down syndrome: a national study. *J. Pediatr.*, 152, 15-19.
- [2] Hindley, D. & Medakkar, S. (2002). Diagnosis of Down's syndrome in neonates. *Arch. Dis. Child. Fetal. Neonatal. Ed.*, 87, F220-F221.
- [3] Wexler, I. D., Abu-Libdeh, A., Kastiel, Y., Nimrodi, A., Kerem, E. & Tenenbaum, A. (2009). Optimizing health care for individuals with Down syndrome in Israel. *Isr. Med. Assoc. J.*, 11, 655-659.
- [4] Giménez-Barcons, M., Casteràs, A., Armengol Mdel, P., Porta, E., Correa, P. A., Marín, A., Pujol-Borrell, R. & Colobran, R. (2014). Autoimmune predisposition in Down syndrome may result from a partial central tolerance failure due to insufficient intrathymic expression of AIRE and peripheral antigens. *J. Immunol.*, 193, 3872-3879.
- [5] Stagi, S., Lapi, E., D'Avanzo, M. G., Perferi, G., Romano, S., Giglio, S., Ricci, S., Azzari, C., Chiarelli, F., Seminara, S. & de Martino, M. (2014). Coeliac disease and risk for other

- autoimmune diseases in patients with Williams-Beuren syndrome. *BMC. Med. Genet.*, 15, 61.
- [6] Annerén, G., Tuvemo, T., Carlsson-Skwirut, C., Lönnerholm, T., Bang, P., Sara, V. R. & Gustafsson, J. (1999). Growth hormone treatment in young children with Down's syndrome: effects on growth and psychomotor development. *Arch. Dis. Child.*, 80, 334-338.
- [7] Cronk, C., Crocker, A. C., Pueschel, S. M., Shea, A. M., Zackai, E., Pickens, G. & Reed, R. B. (1988). Growth charts for children with Down syndrome: 1 month to 18 years of age. *Pediatrics.*, 81, 102-110.
- [8] Styles, M. E., Cole, T. J., Dennis, J. & Preece, M. A. (2002). New cross sectional stature, weight, and head circumference references for Down's syndrome in the UK and Republic of Ireland. *Arch. Dis. Child.*, 87, 104-108.
- [9] McCoy, EE. Growth patterns in Down's syndrome. In: Lott IT, McCoy EE, eds. *Down syndrome: advances in medical care*. New York: Wiley-Liss, Inc., 1992.
- [10] Bermudez, B. E., Medeiros, S. L., Bermudez, M. B., Novadzki, I. M. & Magdalena, N. I. (2015). Down syndrome: Prevalence and distribution of congenital heart disease in Brazil. *Sao. Paulo. Med. J.*, 133, 521-524.
- [11] Stoll, C., Dott, B., Alembik, Y. & Roth, M. P. (2015). Associated congenital anomalies among cases with Down syndrome. *Eur. J. Med. Genet.*, 58, 674-680.
- [12] Aversa, T., Lombardo, F., Valenzise, M., Messina, M. F., Sferlazzas, C., Salzano, G., De Luca, F. & Wasniewska, M. (2015). Peculiarities of autoimmune thyroid diseases in children with Turner or Down syndrome: an overview. *Ital. J. Pediatr.*, 41, 39.
- [13] Aversa, T., Valenzise, M., Salerno, M., Corrias, A., Iughetti, L., Radetti, G., De Luca, F. & Wasniewska, M. (2015). Metamorphic thyroid autoimmunity in Down Syndrome: from Hashimoto's thyroiditis to Graves' disease and beyond. *Ital. J. Pediatr.*, 41, 87.
- [14] Szaflarska-Popławska, A., Soroczyńska-Wrzyszcz, A., Barg, E., Józefczuk, J., Korczowski, B., Grzybowska-Chlebowczyk, U., Więcek, S. & Cukrowska, B. (2016). Assessment of coeliac disease prevalence in patients with Down syndrome in Poland - a multi-centre study. *Prz. Gastroenterol.*, 11, 41-46.
- [15] Lim, H. H., Kil, H. R. & Kim, J. Y. (2012). Unusual presentations of a girl with Down syndrome: Van Wyk-Grumbach syndrome. *J. Pediatr. Endocrinol. Metab.*, 25, 1209-1212.
- [16] Mazurek, D. & Wyka, J. (2015). Down syndrome--genetic and nutritional aspects of accompanying disorders. *Rocz. Panstw. Zakl. Hig.*, 66, 189-194.
- [17] Rosen, D. (2011). Management of obstructive sleep apnea associated with Down syndrome and other craniofacial dysmorphologies. *Curr. Opin. Pulm. Med.*, 17, 431-436.
- [18] Aburawi, E. H., Nagelkerke, N., Deeb, A., Abdulla, S. & Abdulrazzaq, Y. M. (2015). National growth charts for United Arab Emirates children with Down syndrome from birth to 15 years of age. *J. Epidemiol.*, 25, 20-29.
- [19] Al Husain, M. (2003). Growth charts for children with Down's syndrome in Saudi Arabia: birth to 5 years. *Int. J. Clin. Pract.*, 57, 170-174.
- [20] Myreliid, A., Gustafsson, J., Ollars, B. & Annerén, G. (2002). Growth charts for Down's syndrome from birth to 18 years of age. *Arch. Dis. Child.*, 87, 97-103.
- [21] Tüysüz, B., Gökmar, N. T. & Oztürk, B. (2012). Growth charts of Turkish children with Down syndrome. *Am. J. Med. Genet. A.*, 158A, 2656-2664.

- [22] Zemel, B. S., Pipan, M., Stallings, V. A., Hall, W., Schadt, K., Freedman, D. S. & Thorpe, P. (2015). Growth Charts for Children with Down Syndrome in the United States. *Pediatrics.*, 136, e1204-e1211.
- [23] Van Gameren-Oosterom, H. B., Van Dommelen, P., Oudesluys-Murphy, A. M., Buitendijk, S. E., Van Buuren, S. & Van Wouwe, J. P. (2012). Healthy growth in children with Down syndrome. *PLoS. One.*, 7, e31079.
- [24] Meguri, K., Inoue, M., Narahara, K., Sato, T., Takata, A., Ohki, N. & Ozono, K. (2013). Therapeutic Efficacy and Safety of GH in Japanese Children with Down Syndrome Short Stature Accompanied by GH Deficiency. *Clin. Pediatr. Endocrinol.*, 22, 65-72.
- [25] Castells, S., Torrado, C., Bastian, W. & Wisniewski, K. E. (1992). Growth hormone deficiency in Down's syndrome children. *J. Intellect. Disabil. Res.*, 36, 29-43.
- [26] Castells, S., Beaulieu, I., Torrado, C., Wisniewski, K. E., Zarny, S. & Gelato, M. C. (1996). Hypothalamic versus pituitary dysfunction in Down's syndrome as cause of growth retardation. *J. Intellect. Disabil. Res.*, 40, 509-517.
- [27] Pueschel, S. M. (1993). Growth hormone response after administration of L-dopa, clonidine, and growth hormone releasing hormone in children with Down syndrome. *Res. Dev. Disabil.*, 14, 291-298.
- [28] Torrado, C., Bastian, W., Wisniewski, K. E. & Castells, S. (1991). Treatment of children with Down syndrome and growth retardation with recombinant human growth hormone. *J. Pediatr.*, 119, 478-483.
- [29] Wit, J. M. & Walenkamp, M. J. (2013). Role of insulin-like growth factors in growth, development and feeding. *World. Rev. Nutr. Diet.*, 106, 60-65.
- [30] Hilhorst, M. I., Brink, M., Wauters, E. A. & Houwen, R. H. (1993). Down syndrome and coeliac disease: five new cases with a review of the literature. *Eur. J. Pediatr.*, 152, 884-887.
- [31] Weijerman, M. E. & de Winter, J. P. (2010). Clinical practice. The care of children with Down syndrome. *Eur. J. Pediatr.*, 169, 1445-1452.
- [32] Freeman, S. B., Torfs, C. P., Romitti, P. A., Royle, M. H., Druschel, C., Hobbs, C. A. & Sherman, S. L. (2009). Congenital gastrointestinal defects in Down syndrome: a report from the Atlanta and National Down Syndrome Projects. *Clin. Genet.*, 75, 180-184.
- [33] Levy, J. (1991). The gastrointestinal tract in Down syndrome. *Prog. Clin. Biol. Res.*, 373, 245-256.
- [34] Sekerci, A. E., Cantekin, K., Aydinbelge, M. & Ucar, F. İ. (2014). Prevalence of dental anomalies in the permanent dentition of children with Down syndrome. *J. Dent. Child. (Chic.)*, 81, 78-83.
- [35] Abanto, J., Ciamponi, A. L., Francischini, E., Murakami, C., de Rezende, N. P. & Gallottini, M. (2011). Medical problems and oral care of patients with Down syndrome: a literature review. *Spec. Care. Dentist.*, 31, 197-203.
- [36] Basil, J. S., Santoro, S. L., Martin, L. J., Healy, K. W., Chini, B. A. & Saal, H. M. (2016). Retrospective Study of Obesity in Children with Down Syndrome. *J. Pediatr.*, pii: S0022-3476(16)00268-7.
- [37] van Gameren-Oosterom, H. B., van Dommelen, P., Schönbeck, Y., Oudesluys-Murphy, A. M., van Wouwe, J. P. & Buitendijk, S. E. (2012). Prevalence of overweight in Dutch children with Down syndrome. *Pediatrics.*, 130, e1520-1526.

- [38] Rimmer, J. H., Yamaki, K., Lowry, B. M., Wang, E. & Vogel, L. C. (2010). Obesity and obesity-related secondary conditions in adolescents with intellectual/developmental disabilities. *J. Intellect. Disabil. Res.*, 54, 787-794.
- [39] Rubin, S. S., Rimmer, J. H., Chicoine, B., Braddock, D. & McGuire, D. E. (1998). Overweight prevalence in persons with Down syndrome. *Ment. Retard.*, 36, 175-181.
- [40] Murray, J. & Ryan-Krause, P. (2010). Obesity in children with Down syndrome: background and recommendations for management. *Pediatr. Nurs.*, 36, 314-319.
- [41] Foerste, T., Sabin, M., Reid, S. & Reddihough, D. (2016). Understanding the causes of obesity in children with trisomy 21: hyperphagia vs physical inactivity. *J. Intellect. Disabil. Res.*, 2016 Mar 3. doi: 10.1111/jir.12259.
- [42] Whitt-Glover, M. C., O'Neill, K. L. & Stettler, N. (2006). Physical activity patterns in children with and without Down syndrome. *Pediatr. Rehabil.*, 9, 158-164.
- [43] Esposito, P. E., MacDonald, M., Hornyak, J. E. & Ulrich, D. A. (2012). Physical activity patterns of youth with Down syndrome. *Intellect. Dev. Disabil.*, 50, 109-119.
- [44] Nordstrøm, M., Hansen, B. H., Paus, B. & Kolset, S. O. (2013). Accelerometer-determined physical activity and walking capacity in persons with Down syndrome, Williams syndrome and Prader-Willi syndrome. *Res. Dev. Disabil.*, 34, 4395-4403.
- [45] Nordstrøm, M., Paus, B., Andersen, L. F. & Kolset, S. O. (2015). Dietary aspects related to health and obesity in Williams syndrome, Down syndrome, and Prader-Willi syndrome. *Food. Nutr. Res.*, 59, 25487.
- [46] Izquierdo-Gomez, R., Martínez-Gómez, D., Fernhall, B., Sanz, A. & Veiga, Ó. L. (2016). The role of fatness on physical fitness in adolescents with and without Down syndrome: The UP&DOWN study. *Int. J. Obes. (Lond.)*, 40, 22-27.
- [47] Izquierdo-Gomez, R., Martínez-Gómez, D., Tejero-Gonzalez, C. M., Cabanas-Sánchez, V., Ruiz Ruiz, J. & Veiga, Ó. L. (2013). Are poor physical fitness and obesity two features of the adolescent with Down syndrome? *Nutr. Hosp.*, 28, 1348-1351.
- [48] Izquierdo-Gomez, R., Martínez-Gómez, D., Villagra, A., Fernhall, B. & Veiga, O. L. (2014). on behalf of the UP&DOWN study group., Associations of physical activity with fatness and fitness in adolescents with Down syndrome: The UP&DOWN study. *Res. Dev. Disabil.*, 36C, 428-436.
- [49] Ozgen, T., Güven, A. & Aydin, M. (2009). Precocious puberty in a girl with Down syndrome due to primary hypothyroidism. *Turk. J. Pediatr.*, 51, 381-383.
- [50] Kendle, F. W. (1905). Case of precocious puberty in a female cretin. *Br. Med. J.*, 1, 246.
- [51] Van Wyk, J. J. & Grumbach, M. M. (1960). Syndrome of precocious menstruation and galactorrhoea in juvenile hypothyroidism: an example of hormonal overlap in pituitary feedback. *J. Paediatrics.*, 57, 416-435.
- [52] Pringle, P. J., Stanhope, R., Hindmarsh, P. & Brook, C. G. (1988). Abnormal pubertal development in primary hypothyroidism. *Clin. Endocrinol. (Oxf)*, 28, 479-486.
- [53] Rakover, Y., Weiner, E., Shalev, E. & Luboshitsky, R. (1993). Vaginal bleeding: presenting symptom of acquired primary hypothyroidism in a seven year old girl. *J. Paediatr. Endocrinol.*, 6, 197-200.
- [54] Castro-Magaña, M., Angulo, M., Cañas, A., Sharp, A. & Fuentes, B. (1988). Hypothalamic-pituitary gonadal axis in boys with primary hypothyroidism and macroorchidism. *J. Pediatr.*, 112, 397-402.

- [55] Gordon, C. M., Austin, D. J., Radovick, S. & Laufer, M. R. (1997). Primary hypothyroidism presenting as severe vaginal bleeding in a prepubertal girl. *J. Pediatr. Adolesc. Gynecol.*, 10, 35-38.
- [56] Pradhan, M., Dalal, A., Khan, F. & Agrawal, S. (2006). Fertility in men with Down syndrome: a case report. *Fertil. Steril.*, 86, 1765.
- [57] Down, J. L. (1866). Observations on an ethnic classification of idiots. *Clinical Lectures and Reports by the Medical and Surgical Staff of the London Hospital*, 3, 259-262.
- [58] Coleman, M. (1994). Thyroid dysfunction in Down syndrome: A review. *Down. Syndr. Res. Pract.*, 2, 112-115.
- [59] Rogers, P. T. & Coleman, M. *Medical Care in Down Syndrome: A Preventative Medicine Approach*. New York: Marcel Dekker, Inc, 1992.
- [60] Chase, D., Osinowo, T. & Pary, R. J. (2002). Medical issues in patients with Down syndrome. *Ment. Health. Aspects. Dev. Disabil.*, 5, 34-43.
- [61] Coleman, M. & Abbassi, V. (1984). Down's syndrome and hypothyroidism: coincidence or consequence? *Lancet.*, 1, 569.
- [62] Prasher, V. P. (1999). Down syndrome and thyroid disorders: a review. *Downs. Syndr. Res. Pract.*, 6, 25-42.
- [63] Fort, P., Lifshitz, F., Bellisario, R., Davis, J., Lanes, R., Pugliese, M., Richman, R., Post, E. M. & David, R. (1984). Abnormalities of thyroid function in infants with Down syndrome. *J. Pediatr.*, 104, 545-549.
- [64] Korsager, S., Chatham, E. M. & Ostergaard Kristensen, H. P. (1978). Thyroid function tests in adults with Down's syndrome. *Acta. Endocrinol. (Copenh.)*, 88, 48-54.
- [65] Verma, I. C. & Ghai, O. P. (1971). Hypothyroidism in children with mongolism. *Indian. J. Pediatr.*, 38, 229-232.
- [66] Zergollern, L., Vukelić, G. & Raos, M. (1974). Athyrosis in Down's syndrome] *Med. Arh.*, 28, 429-432.
- [67] Cutler, A. T., Benezra-Obeiter, R. & Brink, S. J. (1986). Thyroid function in young children with Down syndrome. *Am. J. Dis. Child.*, 140, 479-483.
- [68] Colombo, M. L., Bona, G., Quaglia, P., Zaffaroni, M., Maina, D. & Luotti, D. (1992). [Thyroid function in children with Down's syndrome] *Minerva. Pediatr.*, 44, 11-6.
- [69] O'Grady, M. & Cody, D. (2011). Subclinical hypothyroidism in childhood. *Arch. Dis. Child.*, 96, 280-284.
- [70] Fialkow, P. J., Thuline, H. C., Hecht, F. & Bryant, J. (1971). Familial predisposition to thyroid disease in Down's syndrome: controlled immunoclinical studies. *Am. J. Hum. Genet.*, 23, 67-86.
- [71] Torfs, C. P., van den Berg, B. J., Oechsli, F. W. & Christianson, R. E. (1990). Thyroid antibodies as a risk factor for Down syndrome and other trisomies. *Am. J. Hum. Genet.*, 47, 727-734.
- [72] van Trotsenburg, A. S., Vulsma, T., van Santen, H. M., Cheung, W. & de Vijlder, J. J. (2003). Lower neonatal screening thyroxine concentrations in down syndrome newborns. *J. Clin. Endocrinol. Metab.*, 88, 1512-1515.
- [73] van Trotsenburg, A. S., Vulsma, T., van Rozenburg-Marres, S. L., van Baar, A. L., Ridder, J. C., Heymans, H. S., Tijssen, J. G. & de Vijlder, J. J. (2005). The effect of thyroxine treatment started in the neonatal period on development and growth of two-year-old Down syndrome children: a randomized clinical trial. *J. Clin. Endocrinol. Metab.*, 90, 3304-3311.

- [74] Jaruratanasirikul, S., Patarakijvanich, N. & Patanapisarnsak, C. (1998). The association of congenital hypothyroidism and congenital gastrointestinal anomalies in Down's syndrome infants. *J. Pediatr. Endocrinol. Metab.*, 11, 241-246.
- [75] Hawli, Y., Nasrallah, M. & El-Hajj Fuleihan, G. (2009). Endocrine and musculoskeletal abnormalities in patients with Down syndrome. *Nat. Rev. Endocrinol.*, 5, 327-334.
- [76] American Academy of Pediatrics Committee on Genetics, (1994). Health supervision for children with Down syndrome. *Pediatrics.*, 93, 855-859.
- [77] Bansal, D., Nanda, A., Gupta, E., Croker, M., Williams, M. L., Bacchus, A., Simmons, D. & Erbland, M. (2006). Myxedema coma in a patient with Down's syndrome. *J. Ark. Med. Soc.*, 103, 112-113.
- [78] Dinleyici, E. C., Ucar, B., Kilic, Z., Dogruel, N. & Yazar, C. (2007). Pericardial effusion due to hypothyroidism in Down syndrome: report of four cases. *Neuro. Endocrinol. Lett.*, 28, 141-144.
- [79] Hayles, A. B., Hinrichs, W. L. & Tauxe, W. N. (1965). Thyroid disease among children with Down's syndrome (mongolism). *Pediatrics.*, 36, 608-614.
- [80] Benda, C. E. (1946). *Mongolism and Cretinism*. New York: Grune and Stratton.
- [81] Pueschel, S. M. & Pezzullo, J. C. (1985). Thyroid dysfunction in Down syndrome. *Am. J. Dis. Child.*, 139, 636-639.
- [82] Selikowitz, M. (1993). A five-year longitudinal study of thyroid function in children with Down syndrome. *Dev. Med. Child. Neurol.*, 35, 396-401.
- [83] Loudon, M. M., Day, R. E. & Duke, E. M. (1985). Thyroid dysfunction in Down's syndrome. *Arch. Dis. Child.*, 60, 1149-1151.
- [84] Pueschel, S. M., Jackson, I. M., Giesswein, P., Dean, M. K. & Pezzullo, J. C. (1991). Thyroid function in Down syndrome. *Res. Dev. Disabil.*, 12, 287-296.
- [85] Ivarsson, S. A., Ericsson, U. B., Gustafsson, J., Forslund, M., Vegfors, P. & Annerén, G. (1997). The impact of thyroid autoimmunity in children and adolescents with Down syndrome. *Acta. Paediatr.*, 86, 1065-1067.
- [86] Karlsson, B., Gustafsson, J., Hedov, G., Ivarsson, S. A. & Annerén, G. (1998). Thyroid dysfunction in Down's syndrome: relation to age and thyroid autoimmunity. *Arch. Dis. Child.*, 79, 242-245.
- [87] Shield, J. P. H., Wadsworth, E. J. K., Hassold, T. J., Judis, L. A. & Jacobs, P. A. (1999). Is disomic homozygosity at the APECED locus the cause of increased autoimmunity in Down's syndrome? *Arch. Dis. Child.*, 81, 147-150.
- [88] Griffin, M. E., Fulcher, T., Nikookam, K., Crowley, J., Firth, R. G. & OMeara, N. M. (1997). Down's syndrome, IDDM, and hypothyroidism. *Diabetes. Care.*, 20, 1202-1203.
- [89] Bucci, I., Napolitano, G., Giuliani, C., Lio, S., Minnucci, A., Di Giacomo, F., Calabrese, G., Sabatino, G., Palka, G. & Monaco, F. (1999). Zinc sulfate supplementation improves thyroid function in hypozincemic Down children. *Biol. Trace. Elem. Res.*, 67, 257-268.
- [90] Romano, C., Pettinato, R., Ragusa, L., Barone, C., Alberti, A. & Failla, P. (2002). Is there a relationship between zinc and the peculiar comorbidities of Down syndrome? *Downs. Syndr. Res. Pract.*, 8, 25-28.
- [91] Monzani, F., Di Bello, V., Caraccio, N., Bertini, A., Giorgi, D., Giusti, C. & Ferrannini, E. (2001). Effect of levothyroxine on cardiac function and structure in subclinical hypothyroidism: a double blind, placebo-controlled study. *J. Clin. Endocrinol. Metab.*, 86, 1110-1115.

- [92] Toscano, E., Pacileo, G., Limongelli, G., Verrengia, M., Di Mita, O., Di Maio, S., Salerno, M., Del Giudice, E., Caniello, B., Calabrò, R. & Andria, G. (2003). Subclinical hypothyroidism and Down's syndrome; studies on myocardial structure and function. *Arch. Dis. Child.*, 88, 1005-1008.
- [93] Hollingsworth, D. R., McKean, H. E. & Roeckel, I. (1974). Goiter, immunological observations, and thyroid function tests in Down syndrome. *Am. J. Dis. Child.*, 127, 524-527.
- [94] Murdoch, J. C., Ratcliffe, W. A., McLarty, D. G., Rodger, J. C. & Ratcliffe, J. G. (1977). Thyroid function in adults with Down's syndrome. *J. Clin. Endocrinol. Metab.*, 44, 453-458.
- [95] Sare, Z., Ruvalcaba, R. H. & Kelley, V. C. (1978). Prevalence of thyroid disorder in Down syndrome. *Clin. Genet.*, 14, 154-158.
- [96] Quinn, M. W. (1980). Down's syndrome and hypothyroidism. *Ir. J. Med. Sci.*, 1, 19-22.
- [97] Lobo, E. H., Khan, M. & Tew, J. (1980). Community study of hypothyroidism in Down's syndrome. *Br. Med. J.*, 280, 1253.
- [98] Samuel, A. M., Krishna Murthy, D. S., Kadival, G. V., Patel, Z. M., Mehta, M. N., Ganatra, R. D. & Ambani, L. M. (1981). Thyroid function studies in young Down's syndrome. *Indian. J. Med. Res.*, 73, 223-227.
- [99] Hughes, V. C., Cameron, J. & Goonetilleke, A. S. (1982). The prevalence of thyroid dysfunction in mentally handicapped in-patients. *J. Ment. Defic. Res.*, 26, 115-120.
- [100] Vladutiu, A. O., Chun, T. C., Victor, A., Gienau, L. & Bannerman, R. M. (1984). Down's syndrome and hypothyroidism: a role for thyroid autoimmunity? *Lancet.*, 1, 1416.
- [101] Kinnell, H. G., Gibbs, N., Teale, J. D. & Smith, J. (1987). Thyroid dysfunction in institutionalised Down's syndrome adults. *Psychol. Med.*, 17, 387-392.
- [102] Mani, C. (1988). Hypothyroidism in Down's syndrome. *Br. J. Psychiatry.*, 153, 102-104.
- [103] Friedman, D. L., Kastner, T., Pond, W. S. & O'Brien, D. R. (1989). Thyroid dysfunction in individuals with Down syndrome. *Arch. Intern. Med.*, 149, 1990-1993.
- [104] Dinani, S. & Carpenter, S. (1990). Down's syndrome and thyroid disorder. *J. Ment. Defic. Res.*, 34, 187-193.
- [105] Zori, R. T., Schatz, D. A., Ostrer, H., Williams, C. A., Spillar, R. & Riley, W. J. (1990). Relationship of autoimmunity to thyroid dysfunction in children and adults with Down syndrome. *Am. J. Med. Genet. Suppl.*, 7, 238-241.
- [106] Pozzan, G. B., Rigon, F., Girelli, M. E., Rubello, D., Busnardo, B. & Baccichetti, C. (1990). Thyroid function in patients with Down syndrome: preliminary results from non-institutionalized patients in the Veneto region. *Am. J. Med. Genet. Suppl.*, 7, 57-58.
- [107] Rubello, D., Pozzan, G. B., Casara, D., Girelli, M. E., Boccato, S., Rigon, F., Baccichetti, C., Piccolo, M., Betterle, C. & Busnardo, B. (1995). Natural course of subclinical hypothyroidism in Down's syndrome: prospective study results and therapeutic considerations. *J. Endocrinol. Invest.*, 18, 35-40.
- [108] Toledo, C., Alembik, Y., Aguirre Jaime, A. & Stoll, C. (1999). Growth curves of children with Down syndrome. *Ann. Genet.*, 42, 81-90.
- [109] Tüysüz, B. & Beker, D. B. (2001). Thyroid dysfunction in children with Down's syndrome. *Acta. Paediatr.*, 90, 1389-1393.
- [110] Ali, F. E., al-Busairi, W. A. & al-Mulla, F. A. (1999). Treatment of hyperthyroidism in Down syndrome: case report and review of the literature. *Res. Dev. Disabil.*, 20, 297-303.

- [111] Wong, L. M., Li, K. Y. & TSE, K. (2004). Thyroid Dysfunction in Chinese Children with Down Syndrome. *HK. J. Paediatr.*, 9, 114-117.
- [112] Gibson, P. A., Newton, R. W., Selby, K., Price, D. A., Leyland, K. & Addison, G. M. (2005). Longitudinal study of thyroid function in Down's syndrome in the first two decades. *Arch. Dis. Child.*, 90, 574-578.
- [113] Mak, P. P. Y., But, W. M., Yu, C. M., Chow C. B., Li, K. Y., Lee, L. P., Yam, W. K. L., Tse, P. W. T., Fung, E., Lau, J. & Tse, W. Y. (2006). Thyroid Dysfunction in Chinese Children and Adolescents with Down Syndrome. *HK. J. Paediatr.*, 11, 110-117
- [114] Unachak, K., Tanpaiboon, P., Pongprot, Y., Sittivangkul, R., Silvilairat, S., Dejkharnon, P. & Sudasna, J. (2008). Thyroid functions in children with Down's syndrome. *J. Med. Assoc. Thai.*, 91, 56-61.
- [115] Unachak, K., Tanpaiboon, P., Pongprot, Y., Sittivangkul, R., Silvilairat, S., Dejkharnon, P. & Sudasna, J. (2008). Thyroid functions in children with Down's syndrome. *J. Med. Assoc. Thai.*, 91, 56-61.
- [116] Konings, C. H., van Trotsenburg, A. S., Ris-Stalpers, C., Vulsma, T., Wiedijk, B. M. & de Vijlder, J. J. (2001). Plasma thyrotropin bioactivity in Down's syndrome children with subclinical hypothyroidism. *Eur. J. Endocrinol.*, 144, 1-4.
- [117] Kanavin, O. J., Aaseth, J. & Birketvedt, G. S. (2000). Thyroid hypofunction in Down's syndrome: is it related to oxidative stress? *Biol. Trace. Elem. Res.*, 78, 35-42.
- [118] Söderbergh, A., Gustafsson, J., Ekwall, O., Hallgren, A., Nilsson, T., Kämpe, O., Rorsman, F. & Annerén, G. (2006). Autoantibodies linked to autoimmune polyendocrine syndrome type I are prevalent in Down syndrome. *Acta. Paediatr.*, 95, 1657-1660.
- [119] Kennedy, R. L., Jones, T. H. & Cuckle, H. S. (1992). Down's syndrome and the thyroid. *Clin. Endocrinol. (Oxf.)*, 37, 471-476.
- [120] Carroll, K. N., Arbogast, P. G., Dudley, J. A. & Cooper, W. O. (2008). Increase in incidence of medically treated thyroid disease in children with Down Syndrome after rerelease of American Academy of Pediatrics Health Supervision guidelines. *Pediatrics.*, 122, e493-e498.
- [121] Fergeson, M. A., Mulvihill, J. J., Schaefer, G. B., Dehaai, K. A., Piatt, J., Combs, K., Bright, B. C. & Neas, B. R. (2009). Low adherence to national guidelines for thyroid screening in Down syndrome. *Genet. Med.*, 11, 548-551.
- [122] Gilchrist, L. (1946). Thyrotoxicosis in a mongol. *Br. Med. J.*, 1, 237-238.
- [123] Goday-Arno, A., Cerda-Esteva, M., Flores-Le-Roux, J. A., Chillaron-Jordan, J. J., Corretger, J. M. & Cano-Pérez, J. F. (2009). Hyperthyroidism in a population with Down syndrome (DS). *Clin. Endocrinol. (Oxf.)*, 71, 110-114.
- [124] Zagzag, D., Miller, D., Cangiarella, J., Allen, J. C. & Alba Greco, M. (1992). Brainstem glioma after radiation therapy for acute myeloblastic leukemia in a child with Down syndrome. *Cancer.*, 70, 1188-1193.
- [125] Satgé, D., Ott, G., Sasco, A. J., Mircher, C., Vekemans, M., Ludwig, L. & Grob, J. C. (2004). A low-grade follicular thyroid carcinoma in a woman with Down syndrome. *Tumori.*, 90, 333-336.
- [126] Holland, W. W., Doll, R. & Carter, C. O. (1962). The mortality from leukaemia and other cancers among patients with Down's syndrome (mongols) and among their parents. *Br. J. Cancer.*, 16, 177-186.

- [127] Kissel, P., Schmitt, J., Barrucand, D., Bessot, M., Gilgenkrantz, S. & Jeanpierre, R. (1969). [Trisomy 21 and neoplastic potentialities] *Rev. Otoneuroophthalmol.*, 41, 255-259.
- [128] Yang, C. Y., Juang, S. S., Chou, C. W. & Lin, M. B. (2000). Down's syndrome with mucosa-associated lymphoid tissue, thyroid lymphoma and cerebral infarction. *Chin. Med. J. (Taipei)*, 63, 234-239.
- [129] Goldacre, M. J., Wotton, C. J., Seagroatt, V. & Yeates, D. (2004). Cancers and immune related diseases associated with Down's syndrome: a record linkage study. *Arch. Dis. Child.*, 89, 1014-1017.
- [130] González-Agüero, A., Vicente-Rodríguez, G., Gómez-Cabello, A., Ara, I., Moreno, L. A. & Casajús, J. A. (2012). A 21-week bone deposition promoting exercise programme increases bone mass in young people with Down syndrome. *Dev. Med. Child. Neurol.*, 54, 552-556.
- [131] Fowler, T. W., McKelvey, K. D., Akel, N. S., Vander Schilden, J., Bacon, A. W., Bracey, J. W., Sowder, T., Skinner, R. A., Swain, F. L., Hogue, W. R., Leblanc, D. B., Gaddy, D., Wenger, G. R. & Suva, L. J. (2012). Low bone turnover and low BMD in Down syndrome: effect of intermittent PTH treatment. *PLoS. One.*, 7, e42967.
- [132] McKelvey, K. D., Fowler, T. W., Akel, N. S., Kelsay, J. A., Gaddy, D., Wenger, G. R. & Suva, L. J. (2013). Low bone turnover and low bone density in a cohort of adults with Down syndrome. *Osteoporos. Int.*, 24, 1333-1338.
- [133] González-Agüero, A., Vicente-Rodríguez, G., Moreno, L. A. & Casajús, J. A. (2011). Bone mass in male and female children and adolescents with Down syndrome. *Osteoporos. Int.*, 22, 2151-2157.
- [134] Wu, J. (2013). Bone mass and density in preadolescent boys with and without Down syndrome. *Osteoporos. Int.*, 24, 2847-2854.
- [135] González-Agüero, A., Vicente-Rodríguez, G., Gómez-Cabello, A. & Casajús, J. A. (2013). Cortical and trabecular bone at the radius and tibia in male and female adolescents with Down syndrome: a peripheral quantitative computed tomography (pQCT) study. *Osteoporos. Int.*, 24, 1035-1044.
- [136] Matute-Llorente, A., González-Agüero, A., Gómez-Cabello, A., Vicente-Rodríguez, G. & Casajús, J. A. (2013). Decreased levels of physical activity in adolescents with down syndrome are related with low bone mineral density: a cross-sectional study. *BMC. Endocr. Disord.*, 13, 22.
- [137] Reza, S. M., Rasool, H., Mansour, S. & Abdollah, H. (2013). Effects of calcium and training on the development of bone density in children with Down syndrome. *Res. Dev. Disabil.*, 34, 4304-4309.
- [138] Stagi, S., Lapi, E., Romano, S., Bargiacchi, S., Brambilla, A., Giglio, S., Seminara, S. & de Martino, M. (2015). Determinants of vitamin d levels in children and adolescents with down syndrome. *Int. J. Endocrinol.*, 2015, 896758.
- [139] Ferry, B., Gavris, M., Tifrea, C., Serbanoiu, S., Pop, A. C., Bembea, M. & Courteix, D. (2014). The bone tissue of children and adolescents with Down syndrome is sensitive to mechanical stress in certain skeletal locations: a 1-year physical training program study. *Res. Dev. Disabil.*, 35, 2077-2084.
- [140] Stagi, S., Iurato, C., Lapi, E., Cavalli, L., Brandi, M. L. & de Martino, M. (2015). Bone status in genetic syndromes: a review. *Hormones. (Athens)*, 14, 19-31.

- [141] Marques, I., Silva, A., Castro, S. & Lopes, L. (2015). Down syndrome, insulin-dependent diabetes mellitus and hyperthyroidism: a rare association. *BMJ. Case. Rep.*, 2015.
- [142] da Rosa Utiyama, S. R., Nisihara, R. M., Nass, F. R., Oliveira, N. P., Fiedler, P. T. & de Messias-Reason, I. T. (2008). Autoantibodies in patients with Down syndrome: early senescence of the immune system or precocious markers for immunological diseases? *J. Paediatr. Child. Health.*, 44, 182-186.
- [143] Giménez-Barcons, M., Casteràs, A., Armengol Mdel, P., Porta, E., Correa, P. A., Marín, A., Pujol-Borrell, R. & Colobran, R. (2014). Autoimmune predisposition in Down syndrome may result from a partial central tolerance failure due to insufficient intrathymic expression of AIRE and peripheral antigens. *J. Immunol.*, 193, 3872-3879.
- [144] Pellegrini, F. P., Marinoni, M., Frangione, V., Tedeschi, A., Gandini, V., Ciglia, F., Mortara, L., Accolla, R. S. & Nespoli, L. (2012). Down syndrome, autoimmunity and T regulatory cells. *Clin. Exp. Immunol.*, 169, 238-243.
- [145] Kusters, M. A., Gemen, E. F., Verstegen, R. H., Wever, P. C. & DE Vries, E. (2010). Both normal memory counts and decreased naive cells favor intrinsic defect over early senescence of Down syndrome T lymphocytes. *Pediatr. Res.*, 67, 557-562.
- [146] Verstegen, R. H., Kusters, M. A., Gemen, E. F. & DE Vries, E. (2010). Down syndrome B-lymphocyte subpopulations, intrinsic defect or decreased T-lymphocyte help. *Pediatr. Res.*, 67, 563-569.
- [147] Kusters, M. A., Verstegen, R. H., Gemen, E. F. & de Vries, E. (2009). Intrinsic defect of the immune system in children with Down syndrome: a review. *Clin. Exp. Immunol.*, 156, 189-193.

Chapter 83

**AUTOSOMAL DOMINANT POLYCYSTIC
KIDNEY DISEASE: PATHOPHYSIOLOGY
AND TREATMENT**

***Ashraf M. Mohieldin¹, Viralkumar S. Upadhyay²,
Albert C. M. Ong³ and Surya M. Nauli^{1,2,*}***

¹Department of Medicinal and Biological Chemistry,

²Department of Pharmacology,

The University of Toledo, Toledo, Ohio

³Academic Unit of Nephrology, University of Sheffield Medical School,
Sheffield, UK

ABSTRACT

Autosomal dominant polycystic kidney disease (ADPKD) is an inherited genetic disorder that results in progressive renal cyst formation and ultimately loss of renal function. Mutation in either *PKD1* or *PKD2*, which are the genes coding for polycystin-1 and polycystin-2, respectively, is the main cause of the disease. The mutation in *PKD1* accounts for 85% of all ADPKD cases, whereas only 15% of ADPKD cases result from *PKD2* mutations. ADPKD is a systemic disorder associated with cardiovascular, portal, pancreatic and gastrointestinal systems. ADPKD is a ciliopathy, a disease associated with abnormal primary cilia. Non-motile primary cilia, functioning as mechanosensory organelles, have been an intense research topic in ADPKD. It has been shown that both structural and functional defects in primary cilia result in cystic kidney and vascular hypertension. In particular, polycystin-1 and polycystin-2 are co-localized to primary cilia and are responsible for mechanosensory induced calcium influx in response to fluid-shear stress. Based on the multiple signaling pathways in ADPKD, different molecular targets have been developed for potential therapies.

* Corresponding Author's Email: Surya.Nauli@UToledo.Edu (Surya M. Nauli, PhD, The University of Toledo, Department of Pharmacology; MS 1015, Health Education building; Room 274, 3000 Arlington Ave, Toledo, OH 43614; Tel: 419-383-1910; Fax: 419-383-1909).

1. INTRODUCTION

Polycystic kidney disease (PKD) is a group of renal cyst diseases that are characterized by the formation of fluid-filled cysts. PKD is classified into acquired and hereditary forms. The acquired form of polycystic kidney disease is characterized by long standing chronic renal failure and subsequent dialysis. However, most forms of polycystic kidney disease are hereditary, including nephronophthisis and medullary cystic kidney diseases. The most common hereditary forms of PKD are autosomal dominant PKD (ADPKD) and autosomal recessive PKD (ARPKD). The pathophysiological presentation of these diseases starts from birth in ARPKD and in the adult years in ADPKD. While the key feature of ARPKD is elongated cysts due to collecting duct dilatation, the hallmark of ADPKD is large focal cysts arising from the rapidly dividing tubular epithelial cells. An important difference between the two is that cysts in ADPKD become isolated, while cysts remain in contact with their tubular origin in ARPKD [1].

The estimated prevalence of ADPKD is one in every 500 to 1,000 individuals [2, 3]. ADPKD is caused by mutation in either *PKD1* or *PKD2*, encoding polycystin-1 or polycystin-2, respectively [4-6]. Mutations in *PKD1* are responsible for more than 85% of ADPKD whereas mutations in *PKD2* account for 15%. On the other hand, ARPKD is caused by mutations in *PKHD1* gene, with a prevalence of one in 20,000 live births [7].

In ADPKD, cysts can form not only in the kidney but also in other organs such as the liver, seminal vesicles, pancreas, and arachnoid membrane [8-10]. Clinically, ADPKD is also characterized by vascular abnormalities such as intracranial aneurysms, dilatation of the aortic root, dissection of the aortic thoracic aorta, mitral valve prolapse, and abdominal wall hernias. Imaging studies are primarily used to diagnose ADPKD. Magnetic resonance imaging (MRI) is used to determine kidney volume as well as to exclude intracranial aneurysms, particularly in patients at high risk. Genetic testing is clinically available for both *PKD1* and *PKD2*.

2. PRIMARY CILIA AND CYSTIC KIDNEY

Primary cilia are found on almost all mammalian cell types including renal epithelia where they act as mechanosensory organelles, sensing and responding to urinary flow in the nephron [11, 12]. Previous studies have showed that primary cilia contain polycystin-1 and polycystin-2 [11-13]. Polycystin-1 and polycystin-2 are glycoproteins widely expressed in various tissues, including renal epithelia, vascular endothelia and cardiac myocytes. Polycystin-1, with 11 transmembrane domains, is developmentally regulated [14, 15]. Subcellular localization of polycystin-1 seems to depend on the stage of development and cell polarization of the tubular epithelium [16, 17]. Polycystin-2 is a calcium channel with six transmembrane domains [6, 18]. The transmembrane region of polycystin-2 is homologous to polycystin-1, voltage-activated and transient receptor potential (TRP) channel subunits [19].

Ultrastructurally, the cilium is a hair-like structure filled with microtubules and enclosed by the ciliary membrane. Primary cilia contains the outer nine microtubule doublets, but lack the inner pair of microtubules, and that is why it is known as a “9+0” axoneme. The outer microtubule doublet is connected by a structural protein nexin to form a ring in the primary cilium [20]. The basal body or mother centriole is a region in which doublet microtubules rise

from the triplet microtubules and is known as the transition zone. The basal body and centriole join together and form a centrosome, which serves as the cell's main microtubule organization. The ciliary necklace, which is made of a series of membrane proteins at the transition zone, helps differentiate the ciliary membrane from the cell's plasma membrane [21]. The primary cilium also has ciliary pockets on each side of the plasma membrane [22]. These are invaginations into the cell membrane adjacent to the ciliary necklace common to many species. The semi-enclosed area is created by apertures at the transition zone, known as the ciliary sheath, and is thought to restrict protein and lipid entry into the cilium; this area is formed during ciliogenesis [23, 24].

Within an epithelial cilium, the polycystin complex has been proposed to have a role in sensing and mediating flow dependent mechanosensory calcium signaling [11, 12, 25-28]. The involvement of polycystins in primary cilia has further provided that ciliary dysfunction results in abnormal planar cell polarity [29, 30].



Figure 1. Hypothetical model of cystogenesis. The diagram depicts the mechanosensory function of a renal tubular cilium and how cilia dysfunction can lead to cyst formation. The cilium plays an important role as a mechanosensory organelle that transmits extracellular signals such as urine and blood flow into the cell. These signals may provide critical messages to the cell regarding the direction of cell division along the tubule. A mutation in either *PKD1* and/or *PKD2* will result in ciliary dysfunction in sensing fluid movement. The abnormality in ciliary sensing could result in the loss of many signals, including those regulating planar cell polarity. As a result, the direction of cell division becomes randomized, resulting in increased tubular diameter rather than tubular elongation. Consequently, cyst growth will occur in isolated focal manner along the renal tubules. More cysts illustrated from the neighboring nephrons are depicted on the bottom left corner. The diagram was modified from the original [152].

Within the kidney anatomy, planar cell polarity is defined as an organized arrangement of cells in a plane of tissue perpendicular to the apical-basal axis as a direction for the orientation of cell division. Using cystic kidney mouse models, defects in cilia function have been shown to cause abnormal spindle pole orientation during cell division [31]. It is therefore thought that inactivation of ciliary protein would result in abnormal planar cell polarity, which in turn triggers an increase in renal tubular diameter. It is hypothesized that the net result is the initiation of cyst formation (Figure 1).

3. LIVER CYSTS

The most common extrarenal manifestation of ADPKD is the formation of liver cysts, the severity and frequency of which increase with age. Polycystic liver also occurs as a genetically distinct disease in the absence of renal cysts. The prevalence of liver cysts is between 75 to 90% among ADPKD patients [32]. Their prevalence by magnetic resonance imaging is 58% in patients 15 to 24 years old, 85% in 25- to 34-year-olds, and 94% in 35- to 46-year-old subjects [33]. Hepatic cysts are more prevalent and cyst volumes larger in women than in men [34]. In addition, women who have had multiple pregnancies or have used oral contraceptive drugs or estrogen replacement therapy have worse disease outcomes suggesting an estrogen effect on hepatic cyst growth.

4. PANCREATIC CYSTS

The pancreas, involved in secretion of hormones and gastric enzymes, contains a maze of tubules and ducts involved in carrying the enzymes to the intestinal lumen. Ductal epithelial cells secrete bicarbonate to neutralize the acidic chyme from the stomach [35]. True congenital cysts of the pancreas are rare. Multiple congenital pancreatic cysts are mostly associated with ADPKD, although they are also found in cystic fibrosis, von Hippel-Lindau disease, Ivemark syndrome, and Meckel Gruber syndrome [36]. In the case of ADPKD, a connection between cilia defects and pancreatic pathologies suggested a link with pancreatic lesions. In some cases, remnants of chronic pancreatitis are found in approximately 10% of patients with ADPKD [37, 38]. In addition, other studies in mutant mice with defects in cilia formation have shown several pancreatic abnormalities, including exocrine cell atrophy, ductal dilation, and collagen deposition [39, 40]. Moreover, another mutant mouse study demonstrated that the absence of cilia in pancreatic cells produces pancreatic lesions that resemble those found in patients with chronic pancreatitis or cystic fibrosis [41].

5. DIVERTICULITIS

Patients with ESRD due to ADPKD have a higher prevalence (20%) of colon diverticulitis than do those with ESRD due to other etiologies (3%) [41]. Not only do ADPKD-ESRD patients have a higher incidence of diverticulitis, they also have a higher complication rate associated with colon diverticulitis [42]. Colon perforation, fistula formation, intra-abdominal

abscess, and generalized peritonitis are frequently diagnosed in ADPKD patients [43, 44]. However, colonic diverticula are usually asymptomatic. Major complications of diverticulitis, including perforation-related peritonitis, sepsis and shock, occur in only a small percentage of patients.

6. PAIN

Pain is the most frequent complaint among ADPKD patients. A recent study of 171 ADPKD patients reported that 71.3% of the patients had lower back pain, the most common site of pain in this group [45]. About 30% of this group had radiculopathy symptoms. The second most common site for pain was reported to be the abdomen. Abdominal pain was reported by 61% of ADPKD patients. The character of pain described by these patients varied from a dull ache (49.5%), uncomfortable fullness (42.7%), stabbing pain (40.4%), and cramping pain (33.0%). Chronic headache and chest pain were also reported in these patients. Pain in ADPKD can occur acutely or persist becoming chronic. Cyst rupture or hemorrhage, infected cysts and renal stones are considered the most common phases of acute pain [32]. Progressive kidney enlargement may cause dull, chronic pain by stretching of the renal capsule. Larger renal volumes accompanied by asymmetric, hypertrophic lumbosacral muscle spasm are likely to be the basis of chronic back pain.

7. HYPERTENSION

With the availability of renal replacement therapies for patients reaching ESRD, cardiovascular complications have emerged as the major cause of death in patients with ADPKD [46, 47]. Hypertension is diagnosed in 50–70% of patients usually before any substantial reduction in glomerular filtration rate is observed [48]. Hypertension relates to progressive kidney enlargement and to ESRD, but in some studies has been reported to be an independent risk factor for progression to ESRD [49]. In addition, hypertension occurs at a much earlier age in ADPKD patients than in the general population [50]. About 10–20% of children with ADPKD develop hypertension, and the majority of adults are hypertensive before any loss of kidney function. The median age for diagnosis of hypertension in ADPKD was 32 years for males and 34 years for females, compared to a median age of 45–55 years in patients with essential hypertension [50, 51]. Not surprisingly, the occurrence of hypertension is greater in both male and female ADPKD patients when their affected parents are hypertensive [32]. The pathogenesis of hypertension in ADPKD patients is complex and multifactorial [52].

7.1. Endothelin

Endothelin-1 (ET-1) has been reported to exert multiple effects on renal physiology [53]. In addition, there is good evidence that many renal cell types, including tubular cells, synthesize and are affected by ET-1, indicating its role as an autocrine or a paracrine factor [54, 55]. ET-1 has also been demonstrated to play a significant role in human renal tubular cells, and it can

stimulate collagen I gene expression in human renal interstitial fibroblasts [56, 57]. Moreover, tubular cell proliferation has been reported to be an early feature of precystic tubules in human ADPKD and many rodent PKD models [58].

Several studies have shown that patients with ADPKD have high expression of ET-1 in the renal cystic epithelium [57, 59]. ET-1 is also found to be present in cyst fluid [60]. Another study demonstrated that patients with ADPKD have increased plasma levels of ET-1 compared with healthy controls and patients with essential hypertension [61]. Moreover, endothelium-dependent relaxation is impaired and endothelial nitric oxide synthase activity is decreased in normotensive patients with ADPKD [62]. These alterations cause up-regulation of ET-1 and dysfunction of the NO system, resulting in arterial vasoconstriction [63].

Other studies have demonstrated a physiological role for ET-1 acting via tubular ET_B receptor to regulate sodium and water excretion in kidney collecting ducts [64, 65]. Activation of tubular ET_B receptors inhibits vasopressin action, thereby promoting diuresis and sodium excretion by inhibiting Na⁺/K⁺ ATPase and/or epithelial sodium channels [66]. As a result, ET_B inhibition can potentiate vasopressin action in cysts derived from the collecting duct and consequently stimulate cyst growth [67, 68]. It is also noteworthy that the increased sympathetic activity and circulating ET-1 levels could result from the stimulation of intrarenal renin-angiotensin-aldosterone system due to progressive cyst enlargement, thereby leading to systemic hypertension [69].

7.2. Primary Cilia and Nitric Oxide

Primary cilia regulated nitric oxide (NO) production play an important role in the regulation of vascular tone [70, 71]. In a blood vessel, an abrupt increase in blood pressure or shear stress can be detected by mechanosensory proteins localized in the cilia [72, 73]. Extracellular fluid mechanics can then be transduced and translated into a complex of intracellular signaling, which in turn activates endothelial nitric oxide synthase (eNOS), an endothelial enzyme that synthesizes NO. The released NO diffuses from endothelial cells to neighboring smooth muscle cells, thus promoting vasodilation.

Both polycystin-1 and polycystin-2 are expressed in the endothelial and vascular smooth muscle cells of all major vessels [74, 75]. Mutations in both *PKD1* and *PKD2* have been shown to contribute to hypertension [76, 77], in part by the failure to convert an increase in mechanical blood flow into cellular NO biosynthesis [72, 73]. It has been shown that impaired endothelial dependent relaxation from aorta cells of *PKD1* knockout mice, due to a defect in (NO) release from the endothelium (Figure 2), correlates with a decrease in Ca²⁺ dependent endothelial NO synthesis activity [78]. We also previously reported the loss of response to fluid-shear stress in mouse endothelial cells with knockdown or knockout of *PKD2* [72]. In addition to the mouse data, polycystin-2 null endothelial cells generated from *PKD2* patients that do not show polycystin-2 in the cilia are unable to sense fluid flow. This further indicates that overall major effect of endothelial cilia function is to decrease total peripheral resistance, thereby lowering the blood pressure through the production of NO.

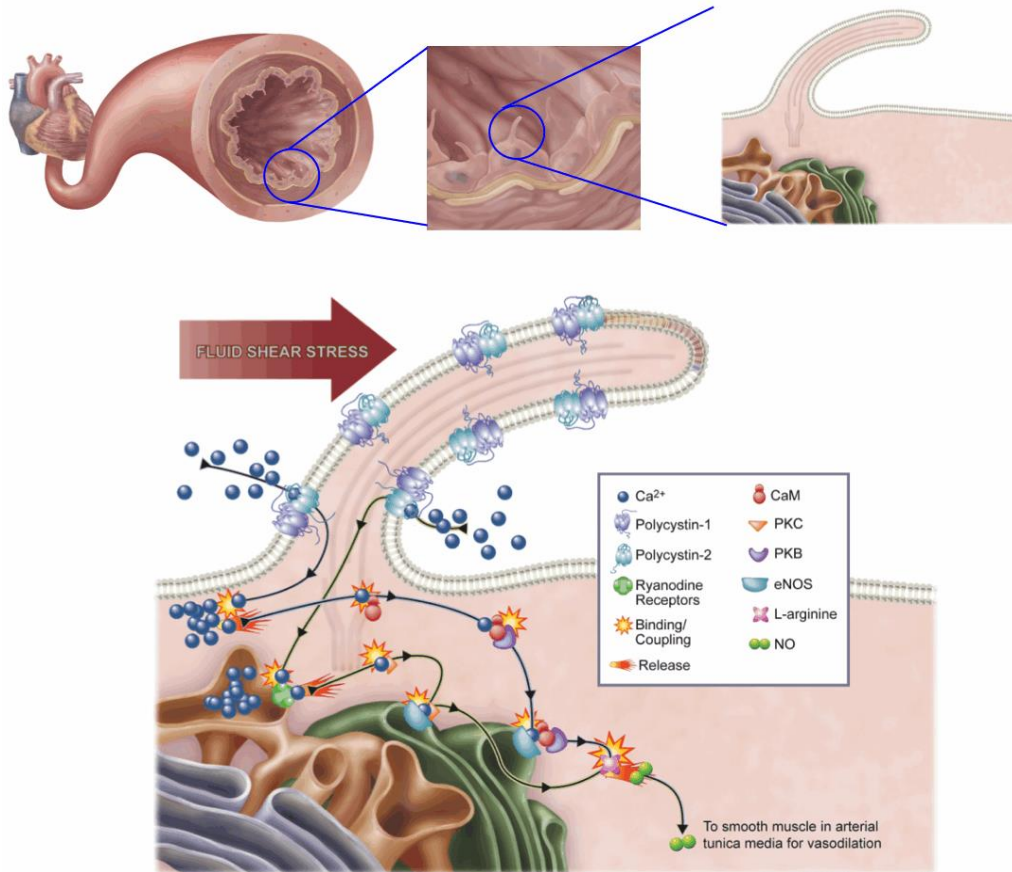


Figure 2. The role of mechanosensory cilia and nitric oxide production in ADPKD. The biochemical production and release of nitric oxide (NO) is dependent on the function of endothelial cilia in the vasculature. In ADPKD, dysfunctional cilia are not able to mechanically sense blood flow, and NO is not produced, resulting in increased blood pressure. The bending of cilia by fluid-shear stress activates the mechanosensory polycystins complex and initiates biochemical synthesis and release of NO. This biochemical cascade involves extracellular calcium influx (Ca^{2+}), followed by the activation of various calcium-dependent proteins including calmodulin (CaM), protein kinase C (PKC) and Akt/PKB). This illustration was modified from the original [153].

A final contributor to loss of vascular tone regulation could be reduction in NO bioavailability secondary to increased reactive oxygen species at least in *PKD2* heterozygous smooth muscle cells [79].

7.3. Angiotensin

Changes in renal structure may play an important role in the pathogenesis of hypertension in ADPKD patients (Figure 3). Cyst enlargement in ADPKD is associated with medial vascular changes and compression of the adjacent parenchyma with resultant areas of renal ischemia and activation of the renin-angiotensin-aldosterone system (RAAS).

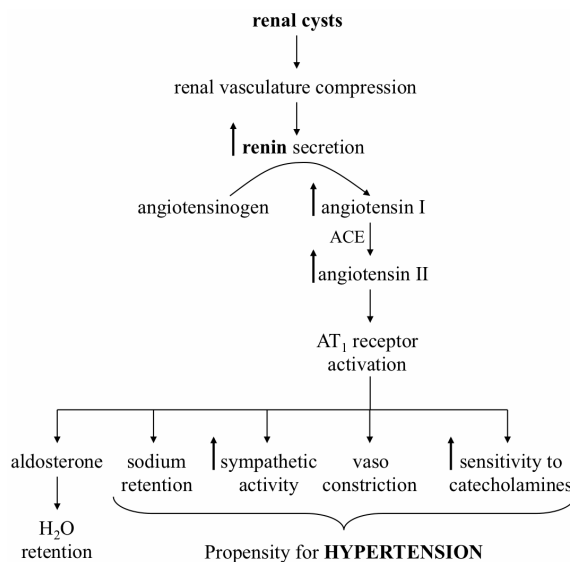


Figure 3. RAAS regulation in ADPKD. Renal cysts compress and disrupt the vascular network in the kidney, which leads to ischemic kidney. It was proposed that this would increase the release of renin from the juxtaglomerular apparatus. The increase in renin secretion would eventually accelerate the conversion of angiotensinogen to angiotensin I, which is further converted by angiotensin-converting enzyme (ACE) to angiotensin II. As a result, the angiotensin receptor (AT₁) is activated initiating cascades of responses resulting in hypertension. This illustration was modified from the original [154].

In support of this, hypertensive patients with ADPKD demonstrate significantly greater renal volumes than patients with normal blood pressure [32, 80]. Other components of the RAAS, including angiotensinogen, angiotensin converting enzyme (ACE), angiotensin II receptor and angiotensin II peptide, have also been detected in cysts and dilated tubules in ADPKD kidneys [81]. Activation of the RAAS has been found in normotensive and hypertensive PKD patients, regardless of their blood pressure and renal function [82]. Not surprisingly, the high levels of circulating angiotensin II in PKD patients have been shown to contribute to the development of vascular hypertrophy, which is further implicated in vascular remodeling [83]. In addition, changes in the vasculature during the course of the PKD progression have been observed in both human [84, 85] and animal studies [86, 87].

Increased sympathetic activity has been reported in hypertensive patients with ADPKD, regardless of renal function [88, 89], suggesting that sympathetic hyperactivity could contribute to the pathogenesis of hypertension in ADPKD patients. ACE-I ramipril and the beta-blocker metoprolol are both effective as first-line therapies in hypertensive PKD patients [90]. It is recommended that more aggressive blood pressure control with these agents is necessary in order to be beneficial for ADPKD patients [32]. It should be noted that angiotensin can stimulate the sympathetic nervous system and vice-versa.

8. LEFT VENTRICULAR HYPERTROPHY

Left ventricular hypertrophy (LVH) is well known as a powerful independent risk factor for cardiovascular morbidity and mortality [91]. Increased left ventricular mass index (LVMI)

is associated with worse renal and patient outcomes in ADPKD [92]. Chapman and colleagues further reported in their study that LVH was found in 48% of hypertensive subjects with ADPKD [93]. Their study also showed a significant correlation between hypertension and LVMI, which has been demonstrated in both children and adults with ADPKD. In addition, children whose blood pressure was within the upper quartile of the normal range were found to have a significantly greater LVMI than those with lower blood pressure [94].

Bardaji and colleagues showed that young normotensive ADPKD patients and preserved renal function had increased LVMI and Doppler abnormalities consistent with early diastolic dysfunction in a cross-sectional study of three different groups of ADPKD patients [95]. Another study by Oflaz and colleagues showed that biventricular diastolic dysfunction was present in both hypertensive and normotensive ADPKD patients with well-preserved renal function [96]. Verdecchia et al. reported that hypertensive patients whose nocturnal blood pressure remains elevated demonstrate higher LVMI compared to those whose blood pressure falls at night [97]. Furthermore, Li Kam et al. demonstrated that hypertensive patients with ADPKD who have normal renal function or mild renal impairment have significantly lower nocturnal decreases in blood pressure compared to patients with essential hypertension [98]. Another study has shown that the nocturnal fall in blood pressure was attenuated even in young normotensive ADPKD patients [99]. In this study, a higher LVMI was closely related to the ambulatory systolic blood pressure in normotensive patients. It has also been shown that insulin resistance is significantly associated with LVMI in healthy relatives and patients with *PKD1* mutations independent of other factors known to increase LVMI such as age, body weight, systolic blood pressure and albuminuria [100]. Thus, the stimulation of angiotensin II and the sympathetic nervous system due to hyperinsulinemia may contribute to increased LVMI in *PKD1* patients [100, 101]. However, other factors besides hypertension, including anemia, obesity, and sodium intake, as well as increased activity of the renin-angiotensin-aldosterone system and other genetic factors, may be associated with LVH, in both hypertensive and normotensive ADPKD patients [93]. The etiology of LVH is likely to be multifactorial but hypertension still plays a major role in its development [102].

9. ANEURYSM

ADPKD patients have a higher prevalence (4.0-11.7%) of intracranial aneurysms than the general population (1.0%) [103, 104]. In addition, four to seven percent of patients with ADPKD die from intracranial aneurysm rupture, and such deaths occur at a younger age than in sporadic cases [105]. Intracranial aneurysm rupture seems to be more common in certain families with ADPKD [106, 107]. This indicates that a family history can be an important tool in assessing the risk of aneurysm rupture in ADPKD. Extracranial aneurysm, such as in coronary arteries, abdominal aorta, renal artery and splenic artery, has been reported in ADPKD patients suggesting that these are primary abnormalities [108, 109]. Other potential cardiovascular features reported in patients with ADPKD include biventricular diastolic dysfunction, endothelial dysfunction, increased carotid intima-media thickness, impaired coronary flow and cardiac valvular defects.

10. THERAPEUTIC TREATMENTS FOR CARDIOVASCULAR COMPLICATIONS

Early and effective treatment of hypertension is highly recommended for the prevention of cardiovascular complications in ADPKD patients. Since LVH and hypertension contribute significantly to cardiovascular morbidity and mortality, controlling these factors can positively impact patient health and survival. Antihypertensive treatment with an ACE inhibitor has been shown to reverse LVH over a seven-year follow-up period, thus decreasing an important risk factor for cardiovascular death in ADPKD patients [110]. A seven-year prospective, randomized study in 75 hypertensive patients with ADPKD and LVH compared the effects of rigorous and standard blood pressure control (<120/80 mmHg versus 135–140/85–90 mmHg) on LVH and renal function. Ecker and Schrier suggest that both strategies decreased LVH significantly [46]. In addition, rigorous blood pressure control was considerably more effective in decreasing LVMI than was standard blood pressure control. More patients in the rigorous-control group (71%) achieved normal LVMI than in the standard group (44%). A subgroup analysis showed that patients who received the ACE inhibitor enalapril experienced a significantly greater decrease in LVH than patients who received the calcium-channel blocker amlodipine, despite similar blood pressure control. On this basis, a blood pressure goal of less than 120/80 mmHg has been recommended for patients with ADPKD with hypertension and LVH [46].

11. MODERN THERAPIES TO HALT PROGRESSION OF RENAL CYSTS

Since ADPKD accounts for up to 10% of patients on renal replacement therapy, an effective disease-modifying drug would have significant implications. The identification of *PKD1* and *PKD2* has led to an explosion in knowledge identifying new disease mechanisms and testing new drugs. Currently, there are three major treatment strategies to treat or reduce progression of kidney failure in ADPKD: to reduce cAMP levels, inhibit cell proliferation, and reduce fluid secretion [111, 112].

11.1. cAMP

Cyclic AMP (cAMP) elevation has an inhibitory effect on cell growth in normal kidney epithelial cells, while it stimulates cell proliferation in ADPKD cells [113]. The molecular basis of this may be Protein kinase-A activation of the B-Raf/MEK/ERK signaling pathway [114]. It has also been proposed that hyper-phosphorylation of polycystin-2 by protein kinase-A can contribute to cystic kidney formation by loss of PC2 inhibition of cell cycle progression [115]. Elevated cAMP also results in increased fluid secretion and cyst enlargement by stimulating the apical CFTR channel and specific basolateral transporters [114, 116]. Vasopressin V2 receptor (V2R) is a major regulator of cAMP production and adenylyl cyclase activity in the principal cells lining the collecting ducts [116]. Nagao et al. showed that high water intake can suppress vasopressin and decrease cyst and renal volumes in PCK rats, with a reduced activity

of the cAMP dependent B-Raf/MEK/ERK pathway [117]. The strongest evidence for a pathogenic role of vasopressin in cyst growth comes from a study which demonstrated that deletion of vasopressin in PCK rats by breeding these with Brattleboro rats results in lower renal cAMP levels and near complete inhibition of cystogenesis [118]. The V2R antagonist OPC-31260 substantially reduced cAMP concentrations and inhibited cyst development in several rodent cystic kidney models [119-121]. Tolvaptan, an analogue with higher potency and selectivity for the human V2R, was equally effective in reducing renal cysts in PCK rats [122]. In a recently concluded phase III clinical trial, Tolvaptan slowed the increase in total kidney volume and the decline in GFR over a three-year period in patients with ADPKD [123]. There was however a significant drop-out rate in the treated group and a few patients developed liver enzyme abnormalities which reversed on cessation of treatment. It is worth mentioning that these drugs had no effect on liver cysts, due to the absence of VPV2R in the liver [34].

11.2. mTOR

The mTOR pathway was shown to be directly regulated by primary cilia [124]. In addition, mTOR signaling can be regulated by different signaling inputs and leads to changes in activity of many cellular processes that drive cyst growth [125]. The polycystin-1 protein directly interacts with the Tuberous Sclerosis Complex-2 (TSC2) protein. TSC2 and Tuberous Sclerosis Complex-1 (TSC1) protein are normally found together in a complex. Bonnet et al. demonstrated that combined mutations in *PKD1* and either *Tsc1* or *Tsc2* in compound heterozygous mice was associated with a more severe renal cystic phenotype than in mice with either mutation alone [126]. mTOR activity is regulated by TSC1-TSC2 complex through several cellular inputs [124, 127]. Thereafter, mTOR regulates protein synthesis, lipid biosynthesis, hypoxia response, *de novo* ceramide synthesis, PKC, AKT, fluid secretion, glycosphingolipid metabolism and ion balance, all of which are dysregulated in PKD. Particularly, the activity of mTORC1, mTORC2, PKC, AKT, ERK, IGF-1, CFTR and EGF-1 are all increased in ADPKD patients [124]. It has therefore been proposed that mTOR inhibition can delay cystic growth and expansion in ADPKD kidneys. More specifically, mTOR kinase activity is aberrantly increased in ADPKD patients [128]. Treatments with mTOR inhibitor, such as rapamycin, sirolimus or everolimus, decrease renal cyst size and improve kidney function in cystic kidney models [129, 130] although not in humans [131, 132].

11.3. EGFR

Members of the epidermal growth factor (EGF)-family bind to ErbB (EGFR)-family receptors, which play an important role in the regulation of various fundamental cell processes including cell proliferation and differentiation [133]. Although ErbB2 and ErbB4 have been detected in developing ureteric buds, EGFR is the predominant ErbB receptor expressed in normal adult mammalian kidney tubules [134, 135]. In addition, EGF-ErbB receptor-mediated interaction is a key element in renal tubular cell proliferation, not only in normal kidneys but also in cyst formation and enlargement. EGF is a well-known mitogen for normal renal epithelia and has been shown to stimulate hyperproliferation in renal cystic epithelia [136, 137]. As mentioned above, EGF and EGF-immunoreactive peptide species are secreted into the apical

medium of cultured ADPKD epithelia, and high mitogenic concentrations of EGF have been measured in ADPKD cyst fluids [1]. Interestingly, the increase of EGF-1 can result in ERK activation through Ras and (B-Raf) Raf signaling pathways, which could in turn regulate the TSC1-TSC2 complex in ADPKD patients [125]. Administration of EGFR inhibitors, such as EKI-785 and EKB-569, in some renal cystic models decreases kidney weights and cyst volumes, suggesting a therapeutic potential of EGFR inhibition in ADPKD treatment [138]. In addition, EGF-like growth factors such as TGF- α and heparin-binding EGF have been found to be abnormally expressed in human ADPKD epithelial cells [133, 139].

11.4. Other Potential Targets

Besides cAMP, mTOR and EGF, there are other potential drugs targeting other signaling pathways such as SR, AP-1, c-Src, Raf, MEK, ERK, A3AR, CFTR and IGF-1 [113, 125], all of which are abnormally regulated in patients with ADPKD (Figure 4). A number of drugs showing promise in preclinical models have been or are being tested in clinical trials (Table 1).

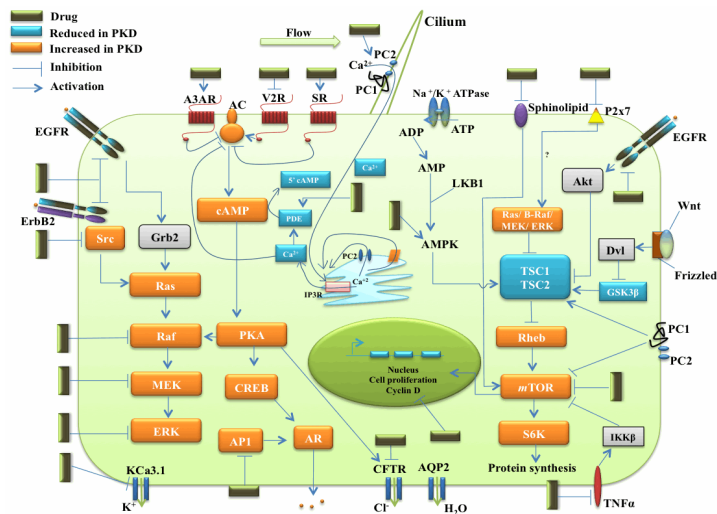


Figure 4. The signaling pathway and drug targets in ADPKD cystic cells. The diagram illustrates the mechanism that polycystin-1 (PC1), polycystin-2 (PC2), signaling proteins, molecules, receptors and drugs exert on signaling pathways leading to cyst formation. The green box indicates drug targets proposed for ADPKD. The blue box indicates the reduced molecules and signaling proteins in ADPKD. The orange box indicates the increased signaling proteins in ADPKD, which are thought to be responsible for an increase in cell proliferation including cAMP, EGFR, Ras/Raf/ERK, Src, CFTR, AC, and mTOR activity. In addition, EGFR activation is also enhanced by amphiregulin (AR) that is abnormally expressed in cystic cells through cAMP, CREB and AP1 signaling. Furthermore, altered EGFR and cAMP signaling stimulate mTOR activity by activation of Akt and Ras/Raf/ERK, which inhibit the TSC1/TSC2 complex. The sphingolipid, Na⁺/K⁺ ATPase, Wnt and P2x7 purinergic receptors are also involved in the regulation of mTOR activity. Na⁺/K⁺ ATPase also regulates the KCa3.1 receptor. However, the abnormal cAMP accumulation contributes to the activation of the Ras/Raf/ERK signaling pathway directly or by the activation of Src, which is able to interact with EGFR in its EGFR/ErbB2 heterodimer form. Other receptors that are involved in ADPKD include adenosine receptor-3A (A3AR), vasopressin receptor-2 (V2R) and somatostatin receptor (SR), which regulate activity of adenylate cyclase (AC). This illustration was adapted from the original [112].

Table 1. Current clinical trials in ADPKD

	Drugs	Signaling pathway	Study design	Treatment duration (months)	Clinical phase trial	Clinical trial status	Clinical trials Gov. identifier	Ref.
1	Octreotide	Analogue of somatostatin known to inhibit cAMP pathway in ADKPD and ADPLD	Randomized, double-blind, Placebo controlled and crossover	12	phase 2/3	Completed /failed	NCT00426153	[140, 141]31
2	Lanreotide	Analogue of somatostatin known to inhibit cAMP pathway in ADKPD and ADPLD	Randomized, double-blind and placebo controlled	6	phase 2/3	Completed	NCT00565097	[142]
3	Sirolimus	mTOR inhibitor	Randomized and penlabel	18	phase 2/3	Completed/ failed	NCT00346918	[132]
4	Sirolimus (SIRENA study)	mTOR inhibitor	Randomized, open-label and crossover	6	phase 2	Completed	NCT00491517	[143]
5	Everolimus	mTOR inhibitor	Multicenter, randomized, double-blind, placebo-controlled	24	phase 3	Completed/ failed	NCT00414440	[131]
6	Tolvaptan	V2-receptor antagonist	parallel-arm, double blind and placebo controlled	36	phase 3	Completed	NCT00428948	[123]

Table 1. (Continued)

	Drugs	Signaling pathway	Study design	Treatment duration (months)	Clinical phase trial	Clinical trial status	Clinical trials Gov. identifier	Ref.
7	Lisinopril and Telmisartan (HALT PKD)1	ACE inhibitor [lisinopril] and angiotensin II receptor blocker (ARB) [telmisartan]	Randomized, double-blind and placebo controlled	72	phase 3	Ongoing	NCT00283686	[144, 145]
8	Pravastatin1	A (HMG coA) reductase inhibitors	Randomized, doubleblind, placebo-controlled	36	phase 3	Ongoing	NCT00456365	[146, 147]
9	Somatostatin	cAMP inhibitor pathway in ADKPD and ADPLD	Randomized, single-blind and placebo controlled	36	phase 3	Ongoing	NCT00309283	not available
10	Bosutinib	c-Src inhibitor	Randomized, double-blind and placebo controlled	24	phase 2	Ongoing	NCT01233869	[148]
12	Triptolide	Restore intracellular Ca ²⁺ signaling and inhibit cell proliferation	Randomized and open-label	36	phase 2	Ongoing	NCT00801268	[149-151]

SUMMARY

Since the discovery of *PKD1* and *PKD2*, there has been tremendous progress in studying disease pathophysiology. As a result, many promising drugs have been and are being developed to slow halt or reverse the progression of cystic disease. Given the complexity of disease and the important extrarenal features, it is likely that a range of specific drugs will be needed to treat patients with ADPKD.

REFERENCES

- [1] P. D. Wilson, Polycystic kidney disease. *The New England journal of medicine* 350, 151 (Jan 8, 2004).
- [2] W. E. Braun, Autosomal dominant polycystic kidney disease: emerging concepts of pathogenesis and new treatments. *Cleve Clin. J. Med.* 76, 97 (Feb, 2009).
- [3] US Renal Data System: USRDS 2008 Annual Data Report. *The National Institutes of Health, National Institute of Diabetes and Digestive and Kidney Disease* Bethesda, MD, (2008).
- [4] Polycystic kidney disease: the complete structure of the PKD1 gene and its protein. The International Polycystic Kidney Disease Consortium. *Cell* 81, 289 (Apr 21, 1995).
- [5] The polycystic kidney disease 1 gene encodes a 14 kb transcript and lies within a duplicated region on chromosome 16. The European Polycystic Kidney Disease Consortium. *Cell* 77, 881 (Jun 17, 1994).
- [6] T. Mochizuki *et al.*, PKD2, a gene for polycystic kidney disease that encodes an integral membrane protein. *Science* 272, 1339 (May 31, 1996).
- [7] C. J. Ward *et al.*, The gene mutated in autosomal recessive polycystic kidney disease encodes a large, receptor-like protein. *Nature genetics* 30, 259 (Mar, 2002).
- [8] V. E. Torres, P. C. Harris, Polycystic kidney disease in 2011: Connecting the dots toward a polycystic kidney disease therapy. *Nature reviews. Nephrology* 8, 66 (Feb, 2012).
- [9] P. C. Harris, V. E. Torres, in *GeneReviews*, R. A. Pagon, T. D. Bird, C. R. Dolan, K. Stephens, M. P. Adam, Eds. (Seattle (WA), 1993).
- [10] P. C. Harris, V. E. Torres, Polycystic kidney disease. *Annual review of medicine* 60, 321 (2009).
- [11] S. M. Nauli *et al.*, Loss of polycystin-1 in human cyst-lining epithelia leads to ciliary dysfunction. *Journal of the American Society of Nephrology: JASN* 17, 1015 (Apr, 2006).
- [12] S. M. Nauli *et al.*, Polycystins 1 and 2 mediate mechanosensation in the primary cilium of kidney cells. *Nature genetics* 33, 129 (Feb, 2003).
- [13] B. K. Yoder, X. Hou, L. M. Guay-Woodford, The polycystic kidney disease proteins, polycystin-1, polycystin-2, polaris, and cystin, are co-localized in renal cilia. *Journal of the American Society of Nephrology : JASN* 13, 2508 (Oct, 2002).
- [14] R. Sandford *et al.*, Comparative analysis of the polycystic kidney disease 1 (PKD1) gene reveals an integral membrane glycoprotein with multiple evolutionary conserved domains. *Human molecular genetics* 6, 1483 (Sep, 1997).

- [15] J. Hughes *et al.*, The polycystic kidney disease 1 (PKD1) gene encodes a novel protein with multiple cell recognition domains. *Nature genetics* 10, 151 (Jun, 1995).
- [16] J. Van Adelsberg, S. Chamberlain, V. D'Agati, Polycystin expression is temporally and spatially regulated during renal development. *The American journal of physiology* 272, F602 (May, 1997).
- [17] L. Geng *et al.*, Distribution and developmentally regulated expression of murine polycystin. *The American journal of physiology* 272, F451 (Apr, 1997).
- [18] S. Gonzalez-Perrett *et al.*, Polycystin-2, the protein mutated in autosomal dominant polycystic kidney disease (ADPKD), is a Ca^{2+} -permeable nonselective cation channel. *Proceedings of the National Academy of Sciences of the United States of America* 98, 1182 (Jan 30, 2001).
- [19] P. Koulen *et al.*, Polycystin-2 is an intracellular calcium release channel. *Nature cell biology* 4, 191 (Mar, 2002).
- [20] B. S. Muntean, S. Jin, S. M. Nauli, Primary Cilia are Mechanosensory Organelles with Chemosensory Roles. *Mechanical Stretch and Cytokines, Mechanosensitivity in Cells and Tissues* 5, 201 (2012).
- [21] N. B. Gilula, P. Satir, The ciliary necklace. A ciliary membrane specialization. *The Journal of cell biology* 53, 494 (May, 1972).
- [22] A. Molla-Herman *et al.*, The ciliary pocket: an endocytic membrane domain at the base of primary and motile cilia. *Journal of cell science* 123, 1785 (May 15, 2010).
- [23] H. L. Kee *et al.*, A size-exclusion permeability barrier and nucleoporins characterize a ciliary pore complex that regulates transport into cilia. *Nature cell biology* 14, 431 (Apr, 2012).
- [24] R. Rohatgi, W. J. Snell, The ciliary membrane. *Current opinion in cell biology* 22, 541 (Aug, 2010).
- [25] S. Wang *et al.*, Fibrocystin/polyductin, found in the same protein complex with polycystin-2, regulates calcium responses in kidney epithelia. *Molecular and cellular biology* 27, 3241 (Apr, 2007).
- [26] C. Xu *et al.*, Human ADPKD primary cyst epithelial cells with a novel, single codon deletion in the PKD1 gene exhibit defective ciliary polycystin localization and loss of flow-induced Ca^{2+} signaling. *American journal of physiology. Renal physiology* 292, F930 (Mar, 2007).
- [27] S. H. Low *et al.*, Polycystin-1, STAT6, and P100 function in a pathway that transduces ciliary mechanosensation and is activated in polycystic kidney disease. *Developmental cell* 10, 57 (Jan, 2006).
- [28] V. Chauvet *et al.*, Mechanical stimuli induce cleavage and nuclear translocation of the polycystin-1 C terminus. *The Journal of clinical investigation* 114, 1433 (Nov, 2004).
- [29] E. Fischer *et al.*, Defective planar cell polarity in polycystic kidney disease. *Nature genetics* 38, 21 (Jan, 2006).
- [30] V. Patel *et al.*, Acute kidney injury and aberrant planar cell polarity induce cyst formation in mice lacking renal cilia. *Human molecular genetics* 17, 1578 (Jun 1, 2008).
- [31] W. A. AbouAlaiwi, S. Ratnam, R. L. Booth, J. V. Shah, S. M. Nauli, Endothelial cells from humans and mice with polycystic kidney disease are characterized by polyploidy and chromosome segregation defects through survivin down-regulation. *Human molecular genetics* 20, 354 (Jan 15, 2011).

- [32] A. Masoumi, B. Reed-Gitomer, C. Kelleher, M. R. Bekheirnia, R. W. Schrier, Developments in the management of autosomal dominant polycystic kidney disease. *Therapeutics and clinical risk management* 4, 393 (Apr, 2008).
- [33] K. T. Bae *et al.*, Magnetic resonance imaging evaluation of hepatic cysts in early autosomal-dominant polycystic kidney disease: the Consortium for Radiologic Imaging Studies of Polycystic Kidney Disease cohort. *Clinical journal of the American Society of Nephrology: CJASN* 1, 64 (Jan, 2006).
- [34] V. E. Torres, P. C. Harris, Y. Pirson, Autosomal dominant polycystic kidney disease. *Lancet* 369, 1287 (Apr 14, 2007).
- [35] S. Abdul-Majeed, S. M. Nauli, Polycystic diseases in visceral organs. *Obstetrics and gynecology international* 2011, 609370 (2011).
- [36] J. H. Grendell, T. H. Ermak, Anatomy, histology, embryology, and developmental anomalies of the pancreas. *Sleisenger and Fordtran's Gastrointestinal and Liver Disease*, 761 (1998).
- [37] O. Basar *et al.*, Recurrent pancreatitis in a patient with autosomal-dominant polycystic kidney disease. *Pancreatology* 6, 160 (2006).
- [38] R. Torra *et al.*, Ultrasonographic study of pancreatic cysts in autosomal dominant polycystic kidney disease. *Clinical nephrology* 47, 19 (Jan, 1997).
- [39] D. A. Cano, N. S. Murcia, G. J. Pazour, M. Hebrok, Orpk mouse model of polycystic kidney disease reveals essential role of primary cilia in pancreatic tissue organization. *Development* 131, 3457 (Jul, 2004).
- [40] J. H. Moyer *et al.*, Candidate gene associated with a mutation causing recessive polycystic kidney disease in mice. *Science* 264, 1329 (May 27, 1994).
- [41] D. A. Cano, S. Sekine, M. Hebrok, Primary cilia deletion in pancreatic epithelial cells results in cyst formation and pancreatitis. *Gastroenterology* 131, 1856 (Dec, 2006).
- [42] S. Y. Li *et al.*, Recurrent retroperitoneal abscess due to perforated colonic diverticulitis in a patient with polycystic kidney disease. *Journal of the Chinese Medical Association: JCMA* 72, 153 (Mar, 2009).
- [43] E. D. Lederman, G. McCoy, D. J. Conti, E. C. Lee, Diverticulitis and polycystic kidney disease. *The American surgeon* 66, 200 (Feb, 2000).
- [44] E. D. Lederman, D. J. Conti, N. Lempert, T. P. Singh, E. C. Lee, Complicated diverticulitis following renal transplantation. *Diseases of the colon and rectum* 41, 613 (May, 1998).
- [45] Z. H. Bajwa, K. A. Sial, A. B. Malik, T. I. Steinman, Pain patterns in patients with polycystic kidney disease. *Kidney international* 66, 1561 (Oct, 2004).
- [46] T. Ecder, R. W. Schrier, Cardiovascular abnormalities in autosomal-dominant polycystic kidney disease. *Nature reviews. Nephrology* 5, 221 (Apr, 2009).
- [47] G. M. Fick, A. M. Johnson, W. S. Hammond, P. A. Gabow, Causes of death in autosomal dominant polycystic kidney disease. *Journal of the American Society of Nephrology: JASN* 5, 2048 (Jun, 1995).
- [48] A. B. Chapman, R. W. Schrier, Pathogenesis of hypertension in autosomal dominant polycystic kidney disease. *Seminars in nephrology* 11, 653 (Nov, 1991).
- [49] A. B. Chapman, K. Stepniakowski, F. Rahbari-Oskoui, Hypertension in autosomal dominant polycystic kidney disease. *Advances in chronic kidney disease* 17, 153 (Mar, 2010).

- [50] C. L. Kelleher, K. K. McFann, A. M. Johnson, R. W. Schrier, Characteristics of hypertension in young adults with autosomal dominant polycystic kidney disease compared with the general U.S. population. *American journal of hypertension* 17, 1029 (Nov, 2004).
- [51] R. W. Schrier, A. M. Johnson, K. McFann, A. B. Chapman, The role of parental hypertension in the frequency and age of diagnosis of hypertension in offspring with autosomal-dominant polycystic kidney disease. *Kidney international* 64, 1792 (Nov, 2003).
- [52] S. M. Nauli, X. Jin, B. P. Hierck, The mechanosensory role of primary cilia in vascular hypertension. *International journal of vascular medicine* 2011, 376281 (2011).
- [53] M. Y. Chang, A. C. Ong, Endothelin in polycystic kidney disease. *Contributions to nephrology* 172, 200 (2011).
- [54] Y. Kawanabe *et al.*, Cilostazol prevents endothelin-induced smooth muscle constriction and proliferation. *PloS one* 7, e44476 (2012).
- [55] Y. Kawanabe, S. M. Nauli, Endothelin. *Cellular and molecular life sciences: CMLS* 68, 195 (Jan, 2011).
- [56] A. C. Ong *et al.*, An endothelin-1 mediated autocrine growth loop involved in human renal tubular regeneration. *Kidney international* 48, 390 (Aug, 1995).
- [57] A. C. Ong *et al.*, Human tubular-derived endothelin in the paracrine regulation of renal interstitial fibroblast function. *Experimental nephrology* 2, 134 (Mar-Apr, 1994).
- [58] M. Y. Chang *et al.*, Haploinsufficiency of Pkd2 is associated with increased tubular cell proliferation and interstitial fibrosis in two murine Pkd2 models. *Nephrology, dialysis, transplantation: official publication of the European Dialysis and Transplant Association - European Renal Association* 21, 2078 (Aug, 2006).
- [59] B. Hoher *et al.*, Renal endothelin system in polycystic kidney disease. *Journal of the American Society of Nephrology: JASN* 9, 1169 (Jul, 1998).
- [60] C. Munemura, J. Uemasu, H. Kawasaki, Epidermal growth factor and endothelin in cyst fluid from autosomal dominant polycystic kidney disease cases: possible evidence of heterogeneity in cystogenesis. *American journal of kidney diseases: the official journal of the National Kidney Foundation* 24, 561 (Oct, 1994).
- [61] R. Giusti *et al.*, Plasma concentration of endothelin and arterial pressure in patients with ADPKD. *Contributions to nephrology* 115, 118 (1995).
- [62] D. Wang, J. Iversen, S. Strandgaard, Endothelium-dependent relaxation of small resistance vessels is impaired in patients with autosomal dominant polycystic kidney disease. *Journal of the American Society of Nephrology: JASN* 11, 1371 (Aug, 2000).
- [63] M. A. Al-Nimri *et al.*, Endothelial-derived vasoactive mediators in polycystic kidney disease. *Kidney international* 63, 1776 (May, 2003).
- [64] Y. Ge, P. K. Stricklett, A. K. Hughes, M. Yanagisawa, D. E. Kohan, Collecting duct-specific knockout of the endothelin A receptor alters renal vasopressin responsiveness, but not sodium excretion or blood pressure. *American journal of physiology. Renal physiology* 289, F692 (Oct, 2005).
- [65] Y. Ge *et al.*, Collecting duct-specific knockout of endothelin-1 alters vasopressin regulation of urine osmolality. *American journal of physiology. Renal physiology* 288, F912 (May, 2005).

- [66] D. E. Kohan, The renal medullary endothelin system in control of sodium and water excretion and systemic blood pressure. *Current opinion in nephrology and hypertension* 15, 34 (Jan, 2006).
- [67] V. E. Torres, P. C. Harris, Autosomal dominant polycystic kidney disease: the last 3 years. *Kidney international* 76, 149 (Jul, 2009).
- [68] M. Y. Chang, E. Parker, M. El Nahas, J. L. Haylor, A. C. Ong, Endothelin B receptor blockade accelerates disease progression in a murine model of autosomal dominant polycystic kidney disease. *Journal of the American Society of Nephrology: JASN* 18, 560 (Feb, 2007).
- [69] R. W. Schrier, Renal volume, renin-angiotensin-aldosterone system, hypertension, and left ventricular hypertrophy in patients with autosomal dominant polycystic kidney disease. *Journal of the American Society of Nephrology: JASN* 20, 1888 (Sep, 2009).
- [70] S. Abdul-Majeed, S. M. Nauli, Dopamine receptor type 5 in the primary cilia has dual chemo- and mechano-sensory roles. *Hypertension* 58, 325 (Aug, 2011).
- [71] S. Abdul-Majeed, B. C. Moloney, S. M. Nauli, Mechanisms regulating cilia growth and cilia function in endothelial cells. *Cellular and molecular life sciences: CMLS* 69, 165 (Jan, 2012).
- [72] W. A. AbouAlaiwi *et al.*, Ciliary polycystin-2 is a mechanosensitive calcium channel involved in nitric oxide signaling cascades. *Circulation research* 104, 860 (Apr 10, 2009).
- [73] S. M. Nauli *et al.*, Endothelial cilia are fluid shear sensors that regulate calcium signaling and nitric oxide production through polycystin-1. *Circulation* 117, 1161 (Mar 4, 2008).
- [74] A. C. Ong *et al.*, Coordinate expression of the autosomal dominant polycystic kidney disease proteins, polycystin-2 and polycystin-1, in normal and cystic tissue. *The American journal of pathology* 154, 1721 (Jun, 1999).
- [75] S. Abdul-Majeed, S. M. Nauli, Calcium-mediated mechanisms of cystic expansion. *Biochimica et biophysica acta* 1812, 1281 (Oct, 2011).
- [76] N. Hateboer *et al.*, Location of mutations within the PKD2 gene influences clinical outcome. *Kidney international* 57, 1444 (Apr, 2000).
- [77] N. Hateboer *et al.*, Comparison of phenotypes of polycystic kidney disease types 1 and 2. European PKD1-PKD2 Study Group. *Lancet* 353, 103 (Jan 9, 1999).
- [78] S. Hassane *et al.*, Pathogenic sequence for dissecting aneurysm formation in a hypomorphic polycystic kidney disease 1 mouse model. *Arteriosclerosis, thrombosis, and vascular biology* 27, 2177 (Oct, 2007).
- [79] Z. L. Brookes *et al.*, Pkd2 mesenteric vessels exhibit a primary defect in endothelium-dependent vasodilatation restored by rosiglitazone. *American journal of physiology. Heart and circulatory physiology* 304, H33 (Jan 1, 2013).
- [80] P. A. Gabow *et al.*, Renal structure and hypertension in autosomal dominant polycystic kidney disease. *Kidney international* 38, 1177 (Dec, 1990).
- [81] M. Loghman-Adham, C. E. Soto, T. Inagami, L. Cassis, The intrarenal renin-angiotensin system in autosomal dominant polycystic kidney disease. *American journal of physiology. Renal physiology* 287, F775 (Oct, 2004).
- [82] A. B. Chapman, P. A. Gabow, Hypertension in autosomal dominant polycystic kidney disease. *Kidney international. Supplement* 61, S71 (Oct, 1997).

- [83] E. A. McPherson *et al.*, Chymase-like angiotensin II-generating activity in end-stage human autosomal dominant polycystic kidney disease. *Journal of the American Society of Nephrology: JASN* 15, 493 (Feb, 2004).
- [84] P. J. Azurmendi *et al.*, Early renal and vascular changes in ADPKD patients with low-grade albumin excretion and normal renal function. *Nephrology, dialysis, transplantation: official publication of the European Dialysis and Transplant Association - European Renal Association* 24, 2458 (Aug, 2009).
- [85] C. T. Itty, A. Farshid, G. Talaulikar, Spontaneous coronary artery dissection in a woman with polycystic kidney disease. *American journal of kidney diseases: the official journal of the National Kidney Foundation* 53, 518 (Mar, 2009).
- [86] C. Boulter *et al.*, Cardiovascular, skeletal, and renal defects in mice with a targeted disruption of the Pkd1 gene. *Proceedings of the National Academy of Sciences of the United States of America* 98, 12174 (Oct 9, 2001).
- [87] M. A. Arnaout, Molecular genetics and pathogenesis of autosomal dominant polycystic kidney disease. *Annual review of medicine* 52, 93 (2001).
- [88] J. Neumann, G. Ligtenberg, I. H. Klein, P. J. Blankestijn, Pathogenesis and treatment of hypertension in polycystic kidney disease. *Current opinion in nephrology and hypertension* 11, 517 (Sep, 2002).
- [89] I. H. Klein, G. Ligtenberg, P. L. Oey, H. A. Koomans, P. J. Blankestijn, Sympathetic activity is increased in polycystic kidney disease and is associated with hypertension. *Journal of the American Society of Nephrology: JASN* 12, 2427 (Nov, 2001).
- [90] R. Zeltner, R. Poliak, B. Stiasny, R. E. Schmieder, B. D. Schulze, Renal and cardiac effects of antihypertensive treatment with ramipril vs metoprolol in autosomal dominant polycystic kidney disease. *Nephrology, dialysis, transplantation: official publication of the European Dialysis and Transplant Association - European Renal Association* 23, 573 (Feb, 2008).
- [91] M. J. Koren, R. B. Devereux, P. N. Casale, D. D. Savage, J. H. Laragh, Relation of left ventricular mass and geometry to morbidity and mortality in uncomplicated essential hypertension. *Annals of internal medicine* 114, 345 (Mar 1, 1991).
- [92] P. A. Gabow *et al.*, Factors affecting the progression of renal disease in autosomal-dominant polycystic kidney disease. *Kidney international* 41, 1311 (May, 1992).
- [93] A. B. Chapman *et al.*, Left ventricular hypertrophy in autosomal dominant polycystic kidney disease. *Journal of the American Society of Nephrology: JASN* 8, 1292 (Aug, 1997).
- [94] M. A. Cadnapaphornchai, K. McFann, J. D. Strain, A. Masoumi, R. W. Schrier, Increased left ventricular mass in children with autosomal dominant polycystic kidney disease and borderline hypertension. *Kidney international* 74, 1192 (Nov, 2008).
- [95] A. Bardaji *et al.*, Cardiac involvement in autosomal-dominant polycystic kidney disease: a hypertensive heart disease. *Clinical nephrology* 56, 211 (Sep, 2001).
- [96] H. Oflaz *et al.*, Biventricular diastolic dysfunction in patients with autosomal-dominant polycystic kidney disease. *Kidney international* 68, 2244 (Nov, 2005).
- [97] P. Verdecchia *et al.*, Circadian blood pressure changes and left ventricular hypertrophy in essential hypertension. *Circulation* 81, 528 (Feb, 1990).
- [98] T. C. Li Kam Wa, A. M. Macnicol, M. L. Watson, Ambulatory blood pressure in hypertensive patients with autosomal dominant polycystic kidney disease. *Nephrology,*

dialysis, transplantation: official publication of the European Dialysis and Transplant Association - European Renal Association 12, 2075 (Oct, 1997).

- [99] F. A. Valero *et al.*, Ambulatory blood pressure and left ventricular mass in normotensive patients with autosomal dominant polycystic kidney disease. *Journal of the American Society of Nephrology: JASN* 10, 1020 (May, 1999).
- [100] A. Lumiaho *et al.*, Insulin resistance is related to left ventricular hypertrophy in patients with polycystic kidney disease type 1. *American journal of kidney diseases: the official journal of the National Kidney Foundation* 41, 1219 (Jun, 2003).
- [101] G. Lembo *et al.*, Abnormal sympathetic overactivity evoked by insulin in the skeletal muscle of patients with essential hypertension. *The Journal of clinical investigation* 90, 24 (Jul, 1992).
- [102] T. Ecder, R. W. Schrier, Hypertension and left ventricular hypertrophy in autosomal dominant polycystic kidney disease. *Expert review of cardiovascular therapy* 2, 369 (May, 2004).
- [103] P. M. Ruggieri *et al.*, Occult intracranial aneurysms in polycystic kidney disease: screening with MR angiography. *Radiology* 191, 33 (Apr, 1994).
- [104] A. P. Rocchini, C. Moorehead, S. DeRemer, T. L. Goodfriend, D. L. Ball, Hyperinsulinemia and the aldosterone and pressor responses to angiotensin II. *Hypertension* 15, 861 (Jun, 1990).
- [105] S. Graf *et al.*, Intracranial aneurysms and dolichoectasia in autosomal dominant polycystic kidney disease. *Nephrology, dialysis, transplantation: official publication of the European Dialysis and Transplant Association - European Renal Association* 17, 819 (May, 2002).
- [106] M. M. Belz *et al.*, Recurrence of intracranial aneurysms in autosomal-dominant polycystic kidney disease. *Kidney international* 63, 1824 (May, 2003).
- [107] M. M. Belz *et al.*, Familial clustering of ruptured intracranial aneurysms in autosomal dominant polycystic kidney disease. *American journal of kidney diseases: the official journal of the National Kidney Foundation* 38, 770 (Oct, 2001).
- [108] H. Hadimeri, C. Lamm, G. Nyberg, Coronary aneurysms in patients with autosomal dominant polycystic kidney disease. *Journal of the American Society of Nephrology: JASN* 9, 837 (May, 1998).
- [109] R. Torra *et al.*, Abdominal aortic aneurysms and autosomal dominant polycystic kidney disease. *Journal of the American Society of Nephrology: JASN* 7, 2483 (Nov, 1996).
- [110] T. Ecder *et al.*, Reversal of left ventricular hypertrophy with angiotensin converting enzyme inhibition in hypertensive patients with autosomal dominant polycystic kidney disease. *Nephrology, dialysis, transplantation: official publication of the European Dialysis and Transplant Association - European Renal Association* 14, 1113 (May, 1999).
- [111] M. Y. Chang, A. C. Ong, Mechanism-based therapeutics for autosomal dominant polycystic kidney disease: recent progress and future prospects. *Nephron. Clinical practice* 120, c25 (2012).
- [112] M. Y. Chang, A. C. Ong, New treatments for autosomal dominant polycystic kidney disease. *British journal of clinical pharmacology*, (Apr 18, 2013).
- [113] G. Aguiari, L. Catizone, L. Del Senno, Multidrug therapy for polycystic kidney disease: a review and perspective. *American journal of nephrology* 37, 175 (2013).

- [114] T. Yamaguchi, G. A. Reif, J. P. Calvet, D. P. Wallace, Sorafenib inhibits cAMP-dependent ERK activation, cell proliferation, and in vitro cyst growth of human ADPKD cyst epithelial cells. *American journal of physiology. Renal physiology* 299, F944 (Nov, 2010).
- [115] A. J. Streets, O. Wessely, D. J. Peters, A. C. Ong, Hyperphosphorylation of polycystin-2 at a critical residue in disease reveals an essential role for polycystin-1-regulated dephosphorylation. *Human molecular genetics* 22, 1924 (May 15, 2013).
- [116] S. Terryn, A. Ho, R. Beauwens, O. Devuyst, Fluid transport and cystogenesis in autosomal dominant polycystic kidney disease. *Biochimica et biophysica acta* 1812, 1314 (Oct, 2011).
- [117] S. Nagao *et al.*, Increased water intake decreases progression of polycystic kidney disease in the PCK rat. *Journal of the American Society of Nephrology: JASN* 17, 2220 (Aug, 2006).
- [118] X. Wang, Y. Wu, C. J. Ward, P. C. Harris, V. E. Torres, Vasopressin directly regulates cyst growth in polycystic kidney disease. *Journal of the American Society of Nephrology: JASN* 19, 102 (Jan, 2008).
- [119] V. H. Gattone, 2nd, X. Wang, P. C. Harris, V. E. Torres, Inhibition of renal cystic disease development and progression by a vasopressin V2 receptor antagonist. *Nature medicine* 9, 1323 (Oct, 2003).
- [120] V. E. Torres *et al.*, Effective treatment of an orthologous model of autosomal dominant polycystic kidney disease. *Nature medicine* 10, 363 (Apr, 2004).
- [121] V. H. Gattone, 2nd, R. L. Maser, C. Tian, J. M. Rosenberg, M. G. Branden, Developmental expression of urine concentration-associated genes and their altered expression in murine infantile-type polycystic kidney disease. *Developmental genetics* 24, 309 (1999).
- [122] X. Wang, V. Gattone, 2nd, P. C. Harris, V. E. Torres, Effectiveness of vasopressin V2 receptor antagonists OPC-31260 and OPC-41061 on polycystic kidney disease development in the PCK rat. *Journal of the American Society of Nephrology: JASN* 16, 846 (Apr, 2005).
- [123] V. E. Torres *et al.*, Tolvaptan in patients with autosomal dominant polycystic kidney disease. *The New England journal of medicine* 367, 2407 (Dec 20, 2012).
- [124] C. Boehlke *et al.*, Primary cilia regulate mTORC1 activity and cell size through Lkb1. *Nature cell biology* 12, 1115 (Nov, 2010).
- [125] O. Ibraghimov-Beskrovnaya, T. A. Natoli, mTOR signaling in polycystic kidney disease. *Trends in molecular medicine* 17, 625 (Nov, 2011).
- [126] C. S. Bonnet *et al.*, Defects in cell polarity underlie TSC and ADPKD-associated cystogenesis. *Human molecular genetics* 18, 2166 (Jun 15, 2009).
- [127] J. Huang, C. C. Dibble, M. Matsuzaki, B. D. Manning, The TSC1-TSC2 complex is required for proper activation of mTOR complex 2. *Molecular and cellular biology* 28, 4104 (Jun, 2008).
- [128] J. M. Shillingford *et al.*, The mTOR pathway is regulated by polycystin-1, and its inhibition reverses renal cystogenesis in polycystic kidney disease. *Proceedings of the National Academy of Sciences of the United States of America* 103, 5466 (Apr 4, 2006).
- [129] M. Wu *et al.*, Everolimus retards cyst growth and preserves kidney function in a rodent model for polycystic kidney disease. *Kidney and blood pressure research* 30, 253 (2007).

- [130] P. R. Wahl *et al.*, Inhibition of mTOR with sirolimus slows disease progression in Han:SPRD rats with autosomal dominant polycystic kidney disease (ADPKD). *Nephrology, dialysis, transplantation: official publication of the European Dialysis and Transplant Association - European Renal Association* 21, 598 (Mar, 2006).
- [131] G. Walz *et al.*, Everolimus in patients with autosomal dominant polycystic kidney disease. *The New England journal of medicine* 363, 830 (Aug 26, 2010).
- [132] A. L. Serra *et al.*, Sirolimus and kidney growth in autosomal dominant polycystic kidney disease. *The New England journal of medicine* 363, 820 (Aug 26, 2010).
- [133] N. N. Zheleznova, P. D. Wilson, A. Staruschenko, Epidermal growth factor-mediated proliferation and sodium transport in normal and PKD epithelial cells. *Biochimica et biophysica acta* 1812, 1301 (Oct, 2011).
- [134] F. Zeng, A. B. Singh, R. C. Harris, The role of the EGF family of ligands and receptors in renal development, physiology and pathophysiology. *Experimental cell research* 315, 602 (Feb 15, 2009).
- [135] F. Zeng, M. Z. Zhang, A. B. Singh, R. Zent, R. C. Harris, ErbB4 isoforms selectively regulate growth factor induced Madin-Darby canine kidney cell tubulogenesis. *Molecular biology of the cell* 18, 4446 (Nov, 2007).
- [136] J. Du, P. D. Wilson, Abnormal polarization of EGF receptors and autocrine stimulation of cyst epithelial growth in human ADPKD. *The American journal of physiology* 269, C487 (Aug, 1995).
- [137] P. D. Wilson, J. Du, J. T. Norman, Autocrine, endocrine and paracrine regulation of growth abnormalities in autosomal dominant polycystic kidney disease. *European journal of cell biology* 61, 131 (Jun, 1993).
- [138] V. E. Torres *et al.*, EGF receptor tyrosine kinase inhibition attenuates the development of PKD in Han:SPRD rats. *Kidney international* 64, 1573 (Nov, 2003).
- [139] G. Aguiari *et al.*, Polycystin-1 regulates amphiregulin expression through CREB and AP1 signalling: implications in ADPKD cell proliferation. *J. Mol. Med. (Berl)* 90, 1267 (Nov, 2012).
- [140] M. C. Hogan *et al.*, Randomized clinical trial of long-acting somatostatin for autosomal dominant polycystic kidney and liver disease. *Journal of the American Society of Nephrology: JASN* 21, 1052 (Jun, 2010).
- [141] T. V. Masyuk, A. I. Masyuk, V. E. Torres, P. C. Harris, N. F. Larusso, Octreotide inhibits hepatic cystogenesis in a rodent model of polycystic liver disease by reducing cholangiocyte adenosine 3',5'-cyclic monophosphate. *Gastroenterology* 132, 1104 (Mar, 2007).
- [142] L. van Keimpema *et al.*, Lanreotide reduces the volume of polycystic liver: a randomized, double-blind, placebo-controlled trial. *Gastroenterology* 137, 1661 (Nov, 2009).
- [143] N. Perico *et al.*, Sirolimus therapy to halt the progression of ADPKD. *Journal of the American Society of Nephrology: JASN* 21, 1031 (Jun, 2010).
- [144] V. E. Torres *et al.*, Analysis of baseline parameters in the HALT polycystic kidney disease trials. *Kidney international* 81, 577 (Mar, 2012).
- [145] A. M. Sengul, Y. Altuntas, A. Kurklu, L. Aydin, Beneficial effect of lisinopril plus telmisartan in patients with type 2 diabetes, microalbuminuria and hypertension. *Diabetes research and clinical practice* 71, 210 (Feb, 2006).

- [146] M. A. Cadnapaphornchai *et al.*, Effect of statin therapy on disease progression in pediatric ADPKD: design and baseline characteristics of participants. *Contemporary clinical trials* 32, 437 (May, 2011).
- [147] R. W. Schrier, Optimal care of autosomal dominant polycystic kidney disease patients. *Nephrology (Carlton)* 11, 124 (Apr, 2006).
- [148] J. Elliott, N. N. Zheleznova, P. D. Wilson, c-Src inactivation reduces renal epithelial cell-matrix adhesion, proliferation, and cyst formation. *American journal of physiology. Cell physiology* 301, C522 (Aug, 2011).
- [149] S. J. Leuenroth, N. Bencivenga, H. Chahboune, F. Hyder, C. M. Crews, Triptolide reduces cyst formation in a neonatal to adult transition Pkd1 model of ADPKD. *Nephrology, dialysis, transplantation: official publication of the European Dialysis and Transplant Association - European Renal Association* 25, 2187 (Jul, 2010).
- [150] S. J. Leuenroth, N. Bencivenga, P. Igarashi, S. Somlo, C. M. Crews, Triptolide reduces cystogenesis in a model of ADPKD. *Journal of the American Society of Nephrology: JASN* 19, 1659 (Sep, 2008).
- [151] S. J. Leuenroth *et al.*, Triptolide is a traditional Chinese medicine-derived inhibitor of polycystic kidney disease. *Proceedings of the National Academy of Sciences of the United States of America* 104, 4389 (Mar 13, 2007).
- [152] R. J. Kolb, S. M. Nauli, Ciliary dysfunction in polycystic kidney disease: an emerging model with polarizing potential. *Frontiers in bioscience: a journal and virtual library* 13, 4451 (2008).
- [153] W. A. Abou Alaiwi, S. T. Lo, S. M. Nauli, Primary cilia: highly sophisticated biological sensors. *Sensors (Basel)* 9, 7003 (2009).
- [154] S. Ratnam, S. M. Nauli, Hypertension in Autosomal Dominant Polycystic Kidney Disease: A Clinical and Basic Science Perspective. *Int. J. Nephrol Urol.* 2, 294 (2010).

Chapter 84

HEREDITARY HAEMORRHAGIC TELANGIECTASIA OR RENDU-OSLER-WEBER SYNDROME

***Roberto Zarrabeitia¹, Cristina Amado¹,
Virginia Albiñana² and Luisa-María Botella²***

¹Sierrallana Hospital, Reference Hospital for HHT in Spain, Torrelavega,
Santander, Spain. IFIMAV

²Center for Biological Investigations (CIB), Spanish Research Council,
(CSIC). Madrid. CIBERER

ABSTRACT

Hereditary Haemorrhagic Telangiectasia (HHT) or Rendu Osler Weber syndrome (OMIM 187300/ORPHA774) is a vascular hereditary autosomic dominant multiorganic dysplasia. Prevalence is in between 1 to 5,000/8,000 inhabitants around 65,000 in Europe, and 200,000 in USA); although due to founder effect, and insulation, it is higher in some regions as the Jura in France, Funen Island in Denmark and Caribbean Dutch Antilles where the prevalence may be 1 in 1,200 inhabitants. Diagnosis is based on the clinical criteria of Curaçao (Shovlin et al., 2000): epistaxis, telangiectases, first degree relative with HHT, and visceral arteriovenous malformations (AVMs), mainly in lung, liver and brain. For a positive diagnosis, 3 out of the 4 previous criteria are required. A positive genetic test implies also a positive diagnosis.

The clinical diagnosis requires then a detailed medical screening, with involvement of different medical specialties. Penetrance of the disease is variable increasing with age.

Pulmonary arteriovenous malformations (PAVMs) occur in approximately 50% of patients, hepatic involvement in up to 70%, brain AVMs in 10% and spinal in 1%. However the most frequent clinical manifestation of HHT is epistaxis (nose bleeding) normally from light to moderate that affects 93% of patients and is present before the age of 21 in 90% of cases (Faughnan et al., 2009).

The genetic origin of the disease is due to mutations of genes involved in the TGB- β pathway, critical for the normal development of blood vessels (Fernández et al. 2006). The first gene identified was *Endoglin* (*ENG*), responsible for the 39-59% of the HHT cases (HHT 1); shortly after, *ALK1*/*ACVRL1* was discovered to be involved in 25-57% of cases (HHT2). In around 2% of the HHT patients, the mutation is located in the *MADH4*/*Smad4*

gene leading to a combined syndrome of Juvenile Polyposis and HHT (JPHT). A third and a fourth locus have been mapped on chromosomes 5 and 7 with no genes identified at the moment. Endoglin plays a key role in vasculogenesis and arterial/venous differentiation in embryos, as well as in angiogenesis and neovascularization processes in the adult; ALK1 is responsible for the events occurring during the activation phase of angiogenesis. Haploinsufficiency is accepted as the mechanism of pathogenicity for the HHT.

INTRODUCTION

Hereditary Haemorrhagic Telangiectasia (HHT) or Rendu Osler Weber syndrome (OMIM 187300/CIE9 448.0/ORPHA774) is a genetic disease with dominant inheritance pattern leading to a vascular dysplasia characterized by the presence of lesions ranging from small vessel enlargements (telangiectases) to complex arteriovenous malformations (AVMs) or shunts that can potentially affect any organ but mainly lungs, brain, liver and gastrointestinal tract. First described [1] by Henry Gawen Sutton in 1864, it was Henry Rendu in 1886 who recognized it as a different entity from the haemophilia [2]. William Bart Osler in 1901 and Frederick Parks Weber in 1907 published the first series of cases [3,4] but it was in 1909, when Hanes introduced the actual terminology of HHT [5]. Microscopical anatomical findings rely on the presence in a capillary vessel of a direct communication between artery and vein with absence of the intermedium capillary net [6,7]. In the initial phase of the development of the telangiectasia, a dilatation of the postcapillary venules in the horizontal upper plexus in the papillary dermis occurs with the presence of a perivascular infiltration of lymphocytes, monocytes and macrophages; in the intermediate phase, when the lesion reaches a diameter of 0.5 mm, the walls of the postcapillary venule enlarge due to the increase in the number of pericytes while the arteriole can be dilated but still connected by a short capillary system with the venule; at the end when the telangiectasia is completely formed (2 mm), the venules are markedly dilated, elongated and twisted occupying the whole dermis.

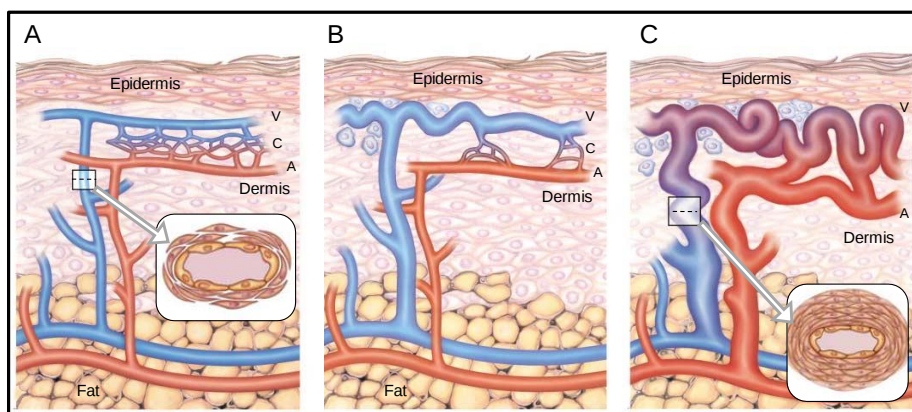


Figure 1. Evolution of a Cutaneous Telangiectasia in Hereditary Hemorrhagic Telangiectasia. **A.** In normal skin, arterioles (A) are connected to venules (V) through multiple capillaries (C). The ultrastructure of a normal postcapillary venule includes the lumen (L), endothelial cells, and two to three layers of surrounding pericytes. **B.** In the earliest stage of cutaneous telangiectasia, a single venule becomes dilated, but it is still connected to an arteriole through one or more capillaries. A perivascular lymphocytic infiltrate is apparent (asterisk). **C.** In a fully developed cutaneous telangiectasia, the venule has become markedly dilated throughout the dermis. The connecting arterioles have also become dilated and communicate directly with the venules without capillaries. The thickened wall of the dilated venules contains more than 10 layers of smooth-muscle cells. (Adapted from Guttmacher et al;1995).

These venules have between 8 and 11 layers of plain muscle and disperse quantities of collagen with no elastic fibers; direct connection between arterioles and venules is observed at this stage with disappearance of the capillary net. The perivascular infiltrate of monocytes can still be present in the abnormal vessels at this time (Figure 1).

HHT IS A RARE DISEASE IN SPITE OF ITS AUTOSOMAL DOMINANCE

According to the literature it can be considered a rare disease as its prevalence ranges from 1:5000-8000 worldwide [8]. It is calculated a population of 62500 affected individuals in Europe, and near one million people all over the world. However and due to a founder effect, there are certain areas with a higher prevalence. The most detailed study about prevalence of the disease was presented in 1989 by Plauchu et al., as a result of a research on French population concluding with a medium prevalence in France of 1/8,000 with higher rates in the Jurá area (1/2,351) [9]. The Dutch Antilles is the place with highest prevalence in the world (1/1,1331 in Curaçao and 1/1,331 in Bonaire) [10]. Other areas with high concentration of cases are Akita in northern Japan (1/5,000-8,000) [11], the Fynn county in Denmark (1/1,641-7,246) [12], and more recently in Canary Islands (near 1/3,000) reported in the 10th HHT meeting in Cork (2013) (Table 1).

Table 1. Estimations of prevalence of HHT

Geographic Area	Prevalence	Reference
<i>World estimation</i>	1/5,000-8,000	Govani F et al. Eur J Hum Genet 2009; 17(7):860-871.
<i>World estimation</i>	min 1/10,000	Abdalla J et al. J Medical Genetics 2006; 43(2):97-110.
<i>Europe estimation</i>	1/5,000 - 8,000	Schoen FJ et al. Hum Mutat 2002; 19(2):140-148.
<i>France</i>	1/ 8,345	Plauchu et al. Am J. Med. Genet 1989; 32: 291-297.
<i>Denmark</i>	1/39,216	Kjeldsen AD et al. Chest 1999; 116(2):432-439.
<i>Northern England</i>	1/5,000 -8,000	Begbie ME et al. Postgrad Med J 2003; 79:18-24
<i>Cantabria (Spain)</i>	min 1:12,200	Morales C et al. Acta Otorrinolaring Esp. 1997; 48:625-629
<i>Norway</i>	1/8,000	Dheyauldeen S et al. Am J Rhinol Allergy 2011; 25(4):214-8
<i>Netherlands</i>	1:5,000-10,000.	Letteboer TG et al. Oral Pathol Oral Radiol Endod 2008; 105:38-41
<i>Germany</i>	1:10,000	Geisthoff UW et al. Arch Clin Exp Ophthalmol 2007; 245:1141-1144.
<i>Italy</i>	1/3,500-5,000	Sabbà et al. Minerva Cardiologica 2002; 50:221-238
<i>USA</i>	1/10,000	Guttmacher AE et al. N Engl J Med 2004; 351(22):2333-36
<i>Vermont (USA)</i>	1/16,500	Guttmacher AE et al. N Engl J Med 1995; 333:918
<i>Japan</i>	1/5,000 8,000	Dakeishi M et al. Hum Mutat 2002; 19(2):140-148
<i>Dutch Antilles</i>	1/1,331	Jessurun GA et al. Clin Neurol Neurosurg 1993; 95(3):193-8
<i>Funen (Denmark)</i>	1/3,500	Kjeldsen AD et al. Chest 1999; 116(2):432-9

Table 2. Curaçao clinical criteria

1. Epistaxis	(Spontaneous and recurrent)
2. Telangiectases	(Multiple and in specific locations)
3. Dominant Inheritance Pattern	(a first-degree relative with HHT according to these criteria)
4. Internal Organ Involvement	(pulmonary, hepatic, cerebral, or spinal arteriovenous malformations, gastrointestinal telangiectasias)

HHT CLINICAL DIAGNOSTIC CRITERIA

Nose bleeding [13] is the main and earliest clinical symptom (90% of patients before the age of 21) present in more than 96% of patients. The epistaxis is also the circumstance affecting primarily the quality of life of HHT patients. However, there are certain cases where the pulmonary or the brain AVMs appear before the nose bleeding manifestation, in early ages, mainly children. Diagnosis is based on the clinical Curaçao criteria [14] (Table 2): epistaxis, mucocutaneous telangiectases in typical areas, family dominant pattern and internal organs involvement. Definite illness is diagnosed if three criteria are fulfilled and probable HHT if patient shows two of them. The identification of the causative mutation through molecular study is complementary to the probable or doubtful cases, and definite in relatives of affected families where the mutation for the index case is known. Penetrance is variable but in general it increases with age (90% at the age of 45) [15]. Due to the HHT rare condition, it is highly underdiagnosed all over the world, with a medium delay in diagnosis of patients estimated in about 30 years [16].

CLINICAL SYMPTOMS

Although the most prevalent symptom is the epistaxis due to the rupture of local telangiectases in the nose, almost any organ of the body can be affected, occasionally with important associated morbi-mortality. In a study carried out on a Danish population of HHT patients [12] the mortality observed on HHT patients under 60 years old was higher than the expected for control population within the same age range. Pattern presentations are highly variable among families, even considering individuals of the same family. However, regarding genotype-phenotype correlation, pulmonary and brain AVMs have been observed with higher incidence on HHT type 1 patients, while HHT2 patients are in higher risk of gastrointestinal and hepatic involvement [17].

EPISTAXIS

The inspired air stream shear stress on the nasal mucose causes the rupture of these malformations with bleeding, present in more than 96% of patients. They normally constitute the earliest and more frequent manifestation of the disease. About 50% of HHT patients will present epistaxis before the age of 20. The Kiesselbach area is normally the most affected and usually the intensity increases with age. However sex, climate, diet, stress and pharmacological treatments can influence the frequency and quantity of nose bleeds [18].

There are several scales trying to quantify epistaxis severity: the Sadick scale [19] that considers frequency and quantity, and more recently the HHT epistaxis severity score (HHT-ESS) [20] that estimates frequency, quantity, characteristics of the bleeding, need for medical attention and the presence of anemia.

MUCOCUTANEOUS TELANGIECTASES

They are present in more than 75% of HHT patients and appear normally from the second decade of life. They are usually round, 1-3 mms reddish spots that disappear with vitropression and normally located on lips, tongue, palate, fingers, face and ears (Figure 2).



Figure 2. Typical telangiectases in HHT.

Apart from the presence of a mutation in heterozygous condition, it has been postulated the need for a “second hit” as physical trauma, wounds, or exposition to physical agents to trigger the telangiectases appearance. Due to the extinction of the capillary net and the arteriolization of the circulation, jet-bleedings are frequent when telangiectases break. Sometimes variations on the microscopic capillaries have been observed with a capillaroscopy [21] preceding the onset of macroscopic lesions, this circumstance which can be useful to clinically diagnose HHT in pediatric ages.

PULMONARY AVMS

It is estimated that 25-30% of HHT patients hold pulmonary arteriovenous malformations (PAVMs) while 90% of general patients with PAVMs suffer in fact from HHT. These arteriovenous malformations in the HHT lung patients are morphological and histologically indistinguishable from those present in non HHT ones; however PAVMs in HHT patients are more frequently multiple (35-65% of cases) and bilateral (25%) [22]. Pulmonary AVMs can be of two types: simple (80% of total, they have a unique afferent artery draining to a unique efferent vein through a bulbous, aneurismatic sac) and complex (20% of total, they have two or more subsegmentary afferent arteries connecting through an aneurismatic sac with two or more efferent arteries) [23]. Lesions are more frequent in lower pulmonary lobes due to dynamic blood flow and it is calculated that 25% of PAVMs grow gradually between 0.3 and 2 mm per year. This risk of enlargement is higher in adolescence, pregnancy and situations of high hypervolemia [24]. They can cause hypoxemia, rupture with secondary haemothorax or hemoptysis, paradoxical stroke (10-15% risk if not treated) due to right-left extracardiac shunting and central nervous system infections in case of paradoxical septic emboli [25]. Screening for PAVMs is mandatory in suspected or diagnosed HHT patients to prevent neurological complications that normally increase with age and number of lesions [25]. The most sensitive test to disclose PAVMs is cardiac echocardiography with agitated saline solution that can be graded to evaluate the possibility of finding large AVMs susceptible to be treated [26].

HEPATIC AVMS

Although up to 75% of HHT patients will present liver involvement in image tests [27], it is normally asymptomatic (only 8% have severe symptoms associated with hepatic liver malformations (HAVMs), and characterized by the presence of telangiectases or confluent vascular structures, perfusion disturbances, enlargement of hepatic artery and venous-venous or portal-venous shunting. In cases of severe liver involvement and due to the double irrigation of the organ, we could find three types of shunts: arterio-hepatic (between hepatic artery and suprahepatic veins: these can lead to heart high-output failure), arterio-portal (between hepatic artery and porta vein: these can lead to cirrhosis) and porto-venous (between porta vein and suprahepatic veins) [28]. Heart failure is the most frequent manifestation of HAVMs (63%), portal hypertension occurs in 17% of severe cases and biliary ischemia in 19% of patients (normally associated with higher shunt grades). Other symptoms are portal-systemic encephalopathy (4%) and mesenteric ischemia. Liver biopsy is contraindicated in these patients (it must be considered also the highest prevalence of nodular focal hyperplasia due to HAVMs which can be confused with cirrotic nodules) and endoscopic cholangiography should be also avoided [29].

BRAIN AVMS

Between 10-20% of HHT patients present brain AVMs, representing a much higher incidence than the non HHT population. Lesions can range from telangiectases (40-50%), arteriovenous malformations (20%), to arteriovenous fistulae (20%), which represent direct connections between artery and vein with high flow, dural arteriovenous fistulae (5%) and cavernomatous malformations (5%) [30]. In HHT patients most lesions tend to be multiple (50% of individuals show two or more malformations) [31], with cortical location in almost all cases. Up to 23% of HHT patients present neurological symptoms as migraine, epilepsy, stroke, abscesses or hemorrhage [32]. Risk of bleeding remains controversial, with series that show less possibility compared with the general population (0.5% per year) [33], and others demonstrating over 20 times higher risks, depending on the type of vascular malformation [34] (normally lesions with deep venous drainage, unique drainage vein or high pressures on afferent arteries). Several cases of bleeding in newborn and young children have been described generally secondary to rupture of high flow pial fistulae. Screening with angio nuclear magnetic resonance is advisable at the moment of diagnosis of HHT and in newborns of parents with HHT and positive genetic test.

SPINAL AVMS

Spinal cord lesions in HHT patients are perimedullary intradural vascular malformations with direct connection between artery and vein and high flow [35]. They can be classified into micro and macro fistulae (the latter more frequent in HHT patients) [36]. The presence of a spinal macrofistula in paediatric ages could indicate suspected HHT. Symptomatology can be secondary to bleeding, compression, vascular robbery or venous thrombosis. Screening for

spinal AVMs in young women could be indicated before pregnancy, or prior to delivery to prevent the puncture of lesions in case of epidural anesthesia.

GASTROINTESTINAL BLEEDING

Gastrointestinal symptoms are related with the development of telangiectases in the digestive tract. These normally tend to appear from the fifth-sixth decade of life and the presence of ferropenic anemia due to chronic bleeding is the rule. Although up to 80% of HHT patients can have lesions, only 13-33% present gastrointestinal bleeding [37] and symptoms can very often mistaken with swallowed blood from epistaxis. Digestive tract involvement is higher in the upper areas (stomach and first parts of the small bowel with normally multiple telangiectases) while the quantity of lesions present in jejunum and ileum correlate with the number of telangiectases present in duodenum [38]. Gastrointestinal tract involvement should be suspected when anemia is present in the context of mild epistaxis and existence of telangiectases shown by endoscopic studies (gastroscopy, colonoscopy or videocapsule). Management is difficult mainly in distant telangiectases. Electrocoagulation of lesions should be only tried in a limited number of cases. Pharmacological management is similar to that used for nose bleeding. Surgery should be delayed as the last therapeutic solution for refractory bleeding. HHT patients with mutations in Smad 4 should be considered for early and periodical endoscopic screening to prevent development of local malignancy [39].

HHT Genetics

Considering molecular basis, two loci are involved by mutation in more than 90% of HHT cases. The first gene identified was *endoglin*, that maps to chromosome 9 [40-42] and is responsible of 39-59% of the total HHT cases. Afterwards *ACVRL1* (activin like kinase type 1 or ALK1) was described [43,44], and maps to chromosome 12, being involved in 25%-57% of the HHT reports. Mutations in *ENG* and *ACVRL1* give rise to HHT1 and HHT2 types, respectively. In 2% of HHT patients the mutation is located in the *MADH4* gene (Smad4) leading to a combined syndrome of Juvenile familial polyposis (JPHT) and HHT [45]. A third and fourth locus with a role in the genesis of HHT have been localized on chromosomes 5 and 7 with no identified genes, at the moment [46,47]. The pathogenical mechanism proposed for the HHT is the haploinsufficiency of endoglin, ALK1 or Smad 4 [48,49]. The mutated counterpart gives rise to either RNA which suffers decay for a premature stop codon, codes for a protein which does not reach the membrane surface, or, even reaching the membrane, the product is functionally inactive in the TGF- β signaling cascade. In summary, the endothelial cell harbors only at the most, half of the functional protein (ENG, ALK1 or Smad4) and this amount is not enough to meet the physiological needs of the cell.

A geographical variation has been observed with higher prevalence of mutations in *ALK1* in the Mediterranean regions [50,51] and on *Endoglin* in North America and Northern countries of Europe [52].

The mutations in *Endoglin* and *ALK1* are numerous, of different types including insertions, deletions, duplications, nucleotide changes, splice defects, loss of exons, and loss of the whole

gene. The mutation database of Arup laboratories is in constant modification and with increasing number of mutations <http://arup.utah.edu/database>.

The three identified genes mutated in HHT (ACVRL1, ENG and MADH4) encode for proteins involved in the TGF- β signalling pathway [8].

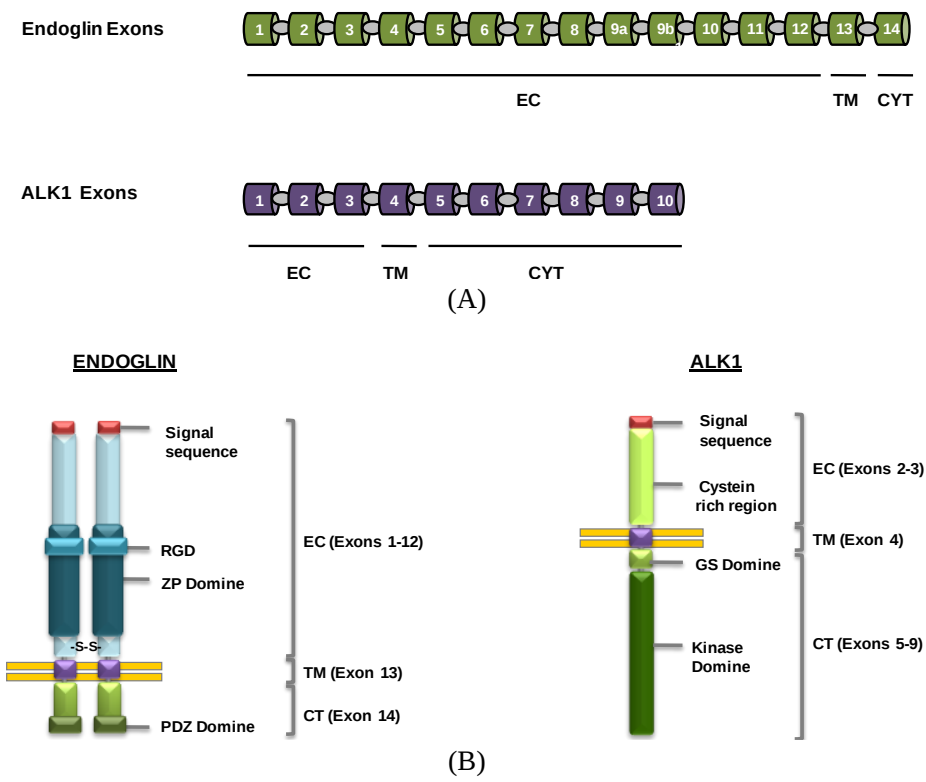


Figure 3. (A). Schematic distribution of exons in *Endoglin* and *ACVRL1* genes (EC: extracellular domain, CYT: cytoplasmic domain, TM: transmembrane domain). (B). Schematic distribution of different regions in Endoglin and ALK1 proteins.

STRUCTURE OF ENDOGLIN AND ALK1

Both endoglin and ALK1 are type I trans-membrane proteins. The *Endoglin* gene is 40 kb long and was localized on chromosome 9q33-34 by linkage studies and *in situ* hybridization [40,41]. It is composed of 14 exons (exons 1 to 12 encode the extracellular domain, exon 13 encodes the transmembrane domain, and exon 14 encodes the cytoplasmic domain). On the other hand *ACVRL1* gene spans 13 Kb and mapped on 12q11-q14 and it is composed of 9 exons (exons 1 to 3 encode the extracellular domain, exon 4 encodes the transmembrane domain, and exon 5 to 10 encodes the cytoplasmic domain [43,44] (Figure 3A, 3B).

Endoglin is expressed as a 180-kDa disulfide-linked homodimer [53]. It contains a large extracellular domain of 561 amino acids, highly glycosylated mainly in asparagines residues. Structurally, endoglin belongs to the Zona Pellucida (ZP) family of proteins that share a ZP domain of 260 amino acid residues at their extracellular region [54,55]. The third region does not show any significant homology to other protein family/domain and thereby has been named

“orphan” domain. A transmembrane region, spanning 25 hydrophobic residues, acts as a linker between the ectodomain and the cytosolic region.

ALK1 is a transmembrane protein of approximately 55 kDa with an N-glycosylated ectodomain of 97 amino acids carrying a cysteine-rich small sequence which likely confers the appropriate structural conformation to capture the ligand. The ALK1 cytoplasmic region of 362 amino acids contains (i) a GS domain, a conserved 30 amino acids glycine/serine-rich sequence involved in the regulation of the receptor activation and (ii) a serine/threonine kinase domain. Phosphorylation of serine/threonine residues of ALK1 in the GS domain by the type II receptor (T β RII) leads to a conformational change in ALK1 that allows phosphorylation of the downstream signaling molecules Smad1, Smad5 or Smad8 [56].

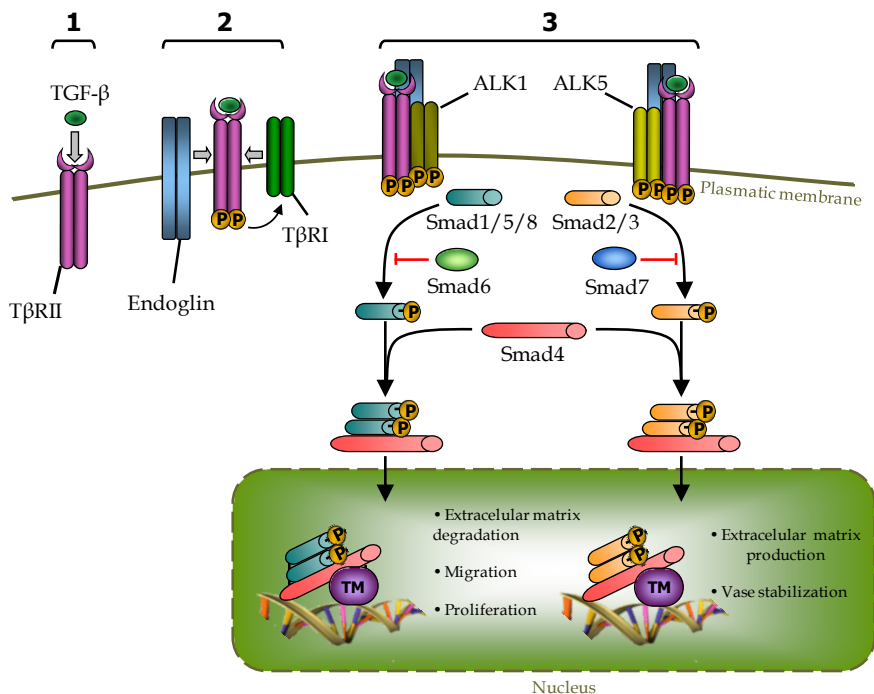


Figure 4. The TGF- β signalling pathway.

THE TGF- β PATHWAY

Transforming growth factor- β 1 (TGF β -1) is the prototypic member of a large family of evolutionarily conserved pleiotropic secreted cytokines, which also includes the activins and bone morphogenetic proteins (BMPs). Individual family members have crucial roles in multiple processes throughout development and in the maintenance of tissue homeostasis in adult life [57]. Not surprisingly, therefore, subversion of signalling by TGF β family members has been implicated in many human diseases, including cancer, fibrosis, autoimmune and vascular diseases [58-60].

TGF- β signals through a heteromeric complex of type I (RI) and type II (RII) which are transmembrane serine/threonine kinase receptors (Figure 4). Although the core of the TGF- β

receptor complex is formed by the association of RI and RII, it may also contain auxiliary receptors such as endoglin and betaglycan.

First, TGF- β binds RII with a high affinity. This TGF- β /RII complex then recruits RI. Once the heteromeric complex TGF- β /RII/RI is formed, a domain of RI is phosphorylated by RII [57,61]. This phosphorylation of serine/threonine residues leads to RI activation, in turn propagating the signal through a cascade of intracellular effectors which belong to the Smad protein family. There are three different types of Smads: receptor regulated (R-Smads), common mediator (Co-Smads) and inhibitory (I-Smads) Smads. R-Smads like Smad1, Smad2, Smad3, Smad5 and Smad8 are phosphorylated and activated by RI and these activated R-Smads bind subsequently to the Co-Smad, Smad4. The R-Smad/Co-Smad complexes translocate to the nucleus where they contribute to the transcriptional activation of target genes [62]. The I-Smads (Smad6 and Smad7) prevent R-Smad phosphorylation by competing with R-Smads for receptor interaction through recruitment of ubiquitin ligases to the activated receptor leading to its proteosomal degradation, or by recruiting phosphatases that inactivate RI [63,64] (Figure 4).

Several members of the TGF- β superfamily, including TGF- β 1, TGF- β 3, activin-A, BMP-2, BMP-7 and BMP-9 are able to bind endoglin and/or ALK1. This binding triggers the Smad-dependent downstream signalling [65,66].

In ECs, endoglin modulates ligand binding and signaling by association with ALK1 and ALK5 [65-68]. Thus, endoglin inhibits the TGF- β /ALK5/Smad3-mediated cellular responses such as the increased expression of the plasminogen activator inhibitor 1 (PAI-1). By contrast, endoglin promotes the ALK5/Smad2-mediated upregulation of endothelial nitric oxide synthase (eNOS) as well as the TGF- β 1/ALK1-mediated increase of Id1. Interestingly, endoglin inhibits the BMP-9/ALK1 signaling in ECs. Overall, endoglin appears to be a critical modulator of the balance between ALK1 and ALK5 signalling [69-71].

Different studies support the view that endoglin and ALK1 participate in a common signaling pathway that is critical for EC responses to TGF- β family members [70-72]. This conclusion agrees with the fact that pathogenic mutations in *ENG* or *ACVRL1* genes result in HHT and that ALK1 and endoglin null mice have similar vascular phenotypes [73].

FUNCTIONAL IMPLICATIONS OF ENDOGLIN AND ALK1

Endoglin and ALK1 are expressed in endothelial cells (ECs), which are the primary cell target in HHT. Endoglin is expressed at low levels in resting ECs, but at high levels in endothelial proliferating cells at sites of active angiogenesis and during embryogenesis [72]. Other cell types that express endoglin at their surface are macrophages, erythroid precursors in bone marrow, syncytiotrophoblasts and several cell types closely related to the cardiovascular system such as smooth muscle cells of atherosclerotic plaques and cardiac fibroblasts [71].

Upregulated expression of endoglin was found in inflamed or infected tissues, healing wounds, psoriatic skin, synovial arthritis, upon vascular injury and in tumoral vessels [71,74,75]. Under hypoxic conditions, the hypoxia inducible factor-1 (HIF-1) complex binds a functional consensus hypoxia responsive element (HRE) in the *ENG* gene promoter [76]. TGF- β signaling, via Smad transcription factors, also potently stimulates endoglin expression [77,78]. Whereas hypoxia alone moderately stimulates endoglin transcription, addition of TGF- β 1 under hypoxic conditions results in a transcriptional cooperation between both signalling

pathways, leading to a marked stimulation of endoglin expression. This synergic stimulation involves the formation of a transcriptional multicomplex containing Smad3/Smad4, Sp1, and HIF-1 [76]. Upon vascular injury, a transcriptional activation of endoglin mediated by the cooperative interaction between Sp1 and KLF6 transcription factors has been reported [74]. By contrast, tumor necrosis factor-alpha (TNF- α) decreases endoglin protein levels in ECs [71].

Endoglin is also implicated in the cytoskeletal organization. The cytoplasmic tail of L-endoglin interacts with members of the LIM domain-containing family of proteins, including zyxin and ZRP-1 (zyxin-related protein-1), involved in regulating cytoskeleton, assembly and cell motility [79]. The organization of the capillary network during angiogenesis depends on the structure of ECs so that in the vasculature of HHT patients a disorganized cytoskeleton is prone to cell breaking with changes in shear stress and blood pressure. This might lead to vessel haemorrhages and eventual disappearance of the capillary network, as occurs in HHT [80].

On the other hand, expression of endoglin in the tumor cells appears to play an important role in the progression of cancer, influencing cell proliferation, motility, invasiveness and tumorigenicity [72,81,82]. In addition, *in vitro* and *in vivo* experiments in which endoglin expression is modulated, have provided evidence supporting endoglin as a tumor suppressor [83]. Interestingly, increased levels of soluble endoglin have been detected in plasma, serum and urine from patients with different pathologies, including preeclampsia and cancer [72]. This soluble endoglin comes from a proteolytic shedding of the membrane bound protein. MT-1 is the metalloprotease responsible for the cleavage [84]. Circulating soluble endoglin is a reliable marker of preeclampsia and is associated with poor prognosis in cancer. Whereas it has been postulated a pathogenic role for soluble endoglin in preeclampsia due to its anti-angiogenic activity, the role of soluble endoglin in tumor progression remains to be established [83]. Decreased levels of soluble endoglin have been detected in plasma samples of HHT patients, as a reflect of decreased endoglin on the membrane surface [85].

ALK1 expression has been reported not only in highly vascularized tissues including lung, placenta, and heart, but also at specific sites of epithelial-mesenchymal interactions, and in other cell types such as monocytes, microglia, skin fibroblasts, stellate hepatic cells, chondrocytes, neural crest stem cells and more recently myoblasts [67,71]. Nonetheless, most studies to date suggest that its major roles are related to the endothelial specific expression pattern. ALK1 is involved in angiogenesis and a regulatory region of ACVRL1 gene is sufficient for endothelial expression in arteries feeding ischemic tissues [86]. The characterization of ACVRL1 promoter and the study of its transcriptional regulation has begun to be elucidated by Garrido-Martín et al. (2011) and its transcriptional activation upon vascular injury, through Sp1 and KLF6 cooperation has been recently described [87].

GENERAL CLINICAL RECOMMENDATIONS FOR MANAGEMENT OF HHT PATIENTS

Given the risk of multisystemic affection in HHT patients, the international guidelines [88] recommend the performance of a screening protocol to disclose the presence of internal organ involvement. These protocols are in constant revision due to the appearance of new evidences but they constitute a very appropriate initial approach to the correct management of the disease. Although, there is a lack of wide series and clinical trials, due to the rare character

of the disease, and differences regarding health systems; experts recommend to perform in adults a routine blood test, genetic test if available (to identify the causative mutation, cardiac echocardiography with contrast (if positive, a computed tomography scan to identify pulmonary AVMs and subsequent diagnostic/therapeutic angiography), liver doppler ultrasound study, brain angio-magnetic resonance and ENT evaluation. In case of children, this screening protocol tries to disclose brain AVMs (angio- magnetic resonance) and if the child is asymptomatic, pulmonary screening could be delayed till adolescence.

TREATMENT

There is not a definite treatment for HHT and due to its quality of rare disease and the lack of normalized clinical trials, the support therapies either pharmacological, surgical or combination of both are based in the “state of the art” practice.

Epistaxis

The pharmacological handling of the nose bleeding, almost similar as in the case of gastrointestinal bleeding, and in addition to the general measures of local moisturing, tamponage (preferably with neumatic or autodissolving devices) and iron supplies in case of anemia, relies on six possible alternatives:

1. *Antifibrinolytic therapy*: ϵ -aminocaproic and mainly tranexamic acid [89], locally inhibitors of the fibrinolysis on the wall of the telangiectases with stabilization of the clot and, as *in vitro* observed, increase of the quantity of endoglin and transcriptional induction of endoglin and ALK1 mRNA [80]. A clinical trial has been carried out showing decrease in frequency and quantity of nose bleeding with 1.5 g/day of oral tranexamic acid [90].
2. *Antiangiogenic therapy*: the vascular endothelial growth factors (VEGFs) are specific mitogens for vascular endothelial cells and essential in the process of angiogenesis and lymphangiogenesis in most of physiological and pathological conditions. Some series of cases have been reported about improvement of nose bleeding in HHT with bevacizumab, a recombinant humanized monoclonal antibody to VEGF-A ligand, both topical and systemic, but with no long-term effect is observed after its removal [91-93]. Also thalidomide, a suppressor of several cytokines and angiogenic factors (VEGF, TNF- α and interleukine 6) has been shown to improve nose bleeding in short series of HHT patients [94] and also in non HHT patients with angiodysplasias in the gastrointestinal tract [95].
3. *Immunosuppressant drugs*: several cases of HHT patients treated with sirolimus or tacrolimus due to concomitant pathology (transplants) showed improvement on bleeding severity [96]. This effect is likely due to a promoter increase of *ENG* and *ALK1* TGF- β pathway dependent [97].

4. *Anti-oxidant agents*, such as N-acetylcysteine. In a pilot study developed on 43 HHT patients treated daily with 600 mg/day of N-acetylcysteine for an average of 11 weeks, nose bleeding decreased [98,99].
5. *Hormonal treatment*: hormonal therapy with ethynilestradiol alone or combined with progesterone can be useful to alleviate nose and gastrointestinal bleeding in HHT patients [100]. Also danazol seems to have similar results and to be better for male HHT population to avoid estrogenic feminizing side effects [101]. Based on results with tamoxifen on HHT women with breast cancer and on series of patients that showed improvement in bleeding severity [102,103], a study with raloxifene (a selective estrogen receptor modulator indicated for treatment and prevention of postmenopausal osteoporosis and with advantages on cardiovascular profile and prevention of breast cancer), was carried out on postmenopausal HHT women with HHT at a dose of 60 mgrs./day. After a medium period of 18 months 72% of patients improved symptoms considering frequency and quantity of epistaxis. In parallel, molecular bases of raloxifene effects on cells were studied, concluding that raloxifene binds the promotor of endoglin and ALK1, stimulating their transcription [104]. With these data, raloxifene was designed as the first orphan drug for HHT by the European Medicines Agency and the Food and Drug Administration in 2009 (EMA/OD/138/09, EU/3/10/730, FDA/10/3099).
6. Other approaches have been performed with propranolol, a β -blocker that is currently considered the most efficient drug treatment for the infantile haemangiomas. Antiangiogenic properties of propranolol have been tested in endothelial cells to check the availability for topical use in HHT patients showing its capacity to decrease cellular migration and tube formation and also apoptotic effects [105]. Isolate reports with other drugs such as bleomycin have been presented with variable results.

It must be considered the fact that most of the previously cited treatment options (mainly hormonal treatment, antifibrinolytics and antiangiogenics) could lead as a side effect to thromboembolism phenomena. There are reports on HHT patients that show high amounts of VIII coagulation factor and an increased number of deep venous thrombosis and pulmonary thromboembolism [106], so careful estimation of risk-benefit must be considered when taking in consideration the use of any of them. On the other hand in HHT patients with atrial fibrillation or prosthetic heart valves or any circumstance with the necessity or option to be treated with antiplatelets or anticoagulants, risk-benefit should be considered prior being this circumstance a relative but not absolute contraindication for its administration.

In case of failing of the pharmacological approach or impossibility of its administration, a surgical approach (combined with systemic drugs or alone) should be considered. The possibilities in this field can be “minor surgery”: electrocoagulation of the local lesions with argon plasma [107] or laser combined or not with topical estrogens or local injection of bevacizumab [108]; local sclerotherapy with ethoxysclerol that has been proved to be a good option in patients with mild-moderate bleeding [109] or “major surgery”: septodermoplasty [110] with a substitution of the damaged mucosa by a cutaneous implant (it normally improves the clinic but many patients refer dryness and bad smell); nostrils closure [111] has been shown in some series to be a good option and quite well tolerated. Supraselective embolization of maxillary artery branches is reserved only for cases of unstoppable bleeding or preparation for additional techniques as its effects is normally very short in time.

Pulmonary AVMs

In the case of adult patients with HHT and pulmonary arteriovenous malformations, embolization [112] of lesions, even if they are diffuse [113], is recommended (with coils or Amplatzer devices) to reduce the potential risk of paradoxical stroke and septic paradoxical embolism. Malformations with an afferent artery over 3mm were supposed to be in a higher risk of complications but recent articles postulate that risk does not depend on this size and all accessible lesions should be treated [114]. Follow up after treatment is based on computer tomography studies and recommended in the next 6-12 months. In the paediatric population, management is different as the risk of stroke is lower and in most cases if the patient is asymptomatic, treatment is delayed until puberty [115]. All patients with pulmonary AVMs, treated or not, should receive antibiotic prophylaxis in case of procedures with risk of bacteremia, mainly of the buccal area, to prevent development of abscesses [116]. Pulmonary hypertension has been related in some cases with ALK1 mutations but in most cases of HHT it is secondary to liver/heart failure, and it must be managed currently [117].

Liver AVMs

Although more than 70% of HHT patients can present liver involvement, only 8% show symptoms. Pharmacological management of clinical symptomatology associated with high-output heart failure, cirrhosis and cholangitis is the initial measure but in severe cases liver transplantation has been proved the alternative with the highest survival rates [118]. Other alternatives as hepatic artery embolization appear to have higher morbi-mortality rates and worse survival prediction. Some trials with antiangiogenics (bevacizumab) mainly in cases of heart failure have been performed with good outcomes [119]. Follow-up of liver affected HHT patients show the need to undergo a close control mainly for cases with hepatic AVMs as there is favourable response to treatment (either pharmacological/surgical) in 63% of cases and indicating the heart failure and the development of arrhythmias as the most associated symptoms [120].

Brain AVMs

Cerebral arteriovenous malformations when disclosed can be managed as in non HHT patients with embolisation, stereotactic radiation and surgery or combinations of previous [121,122]. There are not long-term results on wide series of these patients and recommendations rely on the need of expertise for the management of these patients and to treat in prevention those lesions with higher risk to bleed (i.e., high-flow pial fistulae). Cerebral abscesses in HHT patients are mostly related with the presence of pulmonary AVMs and normally caused by anaerobic bacteria (mainly *Streptococcus*); they must be treated as in non HHT patients (antibiotics and/or surgery) [123].

Pregnancy

Most of deliveries in HHT pregnant women proceed normally, however, due to dynamic blood flow specific circumstances pregnancy is in a higher risk of enlargement of pulmonary AVMs and bleeding [124], cardiac heart failure [125], epistaxis and mucose telangiectases worsening due to the hormonal natural state while there is no evidence about management of brain AVMs during pregnancy, recommending the guidelines to wait till delivery to perform the specific treatment if needed in this case. Recognition of the HHT condition and the presence of pulmonary AVMs before pregnancy increases rates of survival however pregnant HHT women should be treated as high risk obstetric patients [126].

Pediatrics and HHT

There is no evidence about HHT increased risk for birth defects [127], however a high prevalence of arteriovenous malformations have been disclosed in children at early ages or even at birth [128] bringing on the table the necessity of checkup at this age. Screening for asymptomatic HHT children is under debate but due to the potentially life-threatening manifestations should be considered mainly for brain AVMs [129].

REFERENCES

- [1] Sutton, H. Epistaxis as an indication of impaired nutrition, and of degeneration of the vascular system. *Med Mirror*, 1864, 769-81. Rendu, H. Épistaxis répétées chez un sujet porteur de petits angiomes cutanées et muqueux. *Gaz Hop*, 1896, 1322-3.
- [2] Osler, W. On a family form of recurring epistaxis, associated with multiple telangiectases of the skin and mucous membranes. *Bull Johns Hopkins Hops*, 1901, 12, 333-7.
- [3] Weber, F. Multiple hereditary developmental angiomas (telangiectases) of the skin and mucous membranes associated with recurring haemorrhages. *Lancet*, 1907, 2, 160-2.
- [4] Hanes, F. Multiple hereditary telangiectases causing hemorrhage (hereditary hemorrhagic telangiectasia). *Bull Johns Hopkins Hosp*, 1909, 20, 63-73.
- [5] Ikawa, T; Tokunaga, T; Matsumura, M; Imayama, S. An electron microscopic study in Osler's disease. *J Clin Electron Microscopy*, 1983, 16, 5-6
- [6] Jahnke, V. Ultrastructure of hereditary telangiectasia. *Arch Otolaryng*, 1970, 91, 262-5.
- [7] Govani, FS; Shovlin, CL. Hereditary hemorrhagic telangiectasia: a clinical and scientific review. *Eur J Hum Genet*, 2009, 17(7), 860-71.
- [8] Plauchu, H; de Chadarevian, JP; Bideau, A; Robert, JM. Age-related clinical profile of hereditary hemorrhagic telangiectasia in an epidemiologically recruited population. *Am J Med Genet*, 1989, 32, 291-7.
- [9] Jessurun, GA; Kamphuis, DJ; van der, Zande, FH; Nossent, JC. Cerebral arteriovenous malformations in the Netherlands Antilles; high prevalence of hereditary hemorrhagic telangiectasia-related single and multiple cerebral arteriovenous malformations. *Clin Neurol Neurosurg*, 1993, 95(3), 193-8.

- [10] Dakeishi, M; Shioya, T; Wada, Y; et al. Genetic epidemiology of hereditary hemorrhagic telangiectasia in a local community in the northern part of Japan. *Hum Mutat*, 2002, 19(2), 140-148.
- [11] Kjeldsen, AD; Vase, P; Green, A. Hereditary hemorrhagic telangiectasia: a population-based study of prevalence and mortality in Danish patients. *J Intern Med*, 1999, 245(1), 31-9.
- [12] Os, AA; Friedman, CM; White, RI, Jr. The natural history of epistaxis in hereditary hemorrhagic telangiectasia. *Laryngoscope*, 1991, 101, 977-80
- [13] Shovlin, CL; Guttmacher, AE; Buscarini, E; et al. Diagnostic criteria for hereditary hemorrhagic telangiectasia (Rendu Osler Weber syndrome). *Am J Genet*, 2000, 91, 66-7.
- [14] Guttmacher, AE; Marchuuk, DA; White, RI, Jr. Hereditary hemorrhagic telangiectasia. *N Engl J Med*, 1995, 333, 918-24.
- [15] Pierucci, P; Lenato, GM; Suppressa, P; et al. A long diagnostic delay in patients with hereditary hemorrhagic telangiectasia: a questionnaire-based retrospective study. *Orphanet J rare Dis*, 2012, 7, 33.
- [16] Lesca, G; Olivieri, C; Burnichon, N; et al. Genotype-phenotype correlations in hereditary hemorrhagic telangiectasia, data from the French-Italian HHT network. *Genet Med*, 2007, 9, 14-22.
- [17] Aassar, O; Friedman, C; White, R. The natural history of epistaxis in hereditary hemorrhagic telangiectasia. *Laryngoscope*, 1991, 101, 977-80.
- [18] Sadick, H; Naim, R; Oulmi, J; et al. Plasma surgery and topical estriol, effects on the nasal mucosa and long-term results in patients with Osler's disease. *Otolaryngol Head Neck Surg*, 2003, 129, 233-8.
- [19] Hoag, JB; Terry, P; Mitchell, S; et al. An epistaxis severity score for hereditary hemorrhagic telangiectasia. *Laryngoscope* 2010, 120, 838-43.
- [20] Mager, JJ; Westermann, CJ. Value of capillary microscopy in the diagnosis of hereditary hemorrhagic telangiectasia. *Arch Dermatol*, 2000, 136, 732-4.
- [21] White, RI; Mitchell, SE; Barth, KH; et al. Angioarchitecture of pulmonary arteriovenous malformations: an important consideration before embolotherapy. *Am J Roentgenol*, 1983, 140, 681-6.
- [22] Gossage, JR; Kanj, G. Pulmonary arteriovenous malformations. A state of the art review. *Am J Respir Crit Care Med*, 1998, 158, 643-61.
- [23] Dines, D; Arms, R; Bernatz, P; et al. Pulmonary arteriovenous fistulas. *Mayo Clin Proc*, 1974, 49, 460-5.
- [24] McDonald, J; Bayrak-Toydemir, P; Pyeritz, RE. Hereditary hemorrhagic telangiectasia: an overview of diagnosis, management and pathogenesis. *Genet Med*, 2011, 7, 607-16.
- [25] Parra, JA; Bueno, J; Zaruza, J; et al. Graded contrast echocardiography in pulmonary arteriovenous malformations. *Eur Respir J*, 2010, 35, 1-7.
- [26] Ianora, AAS; memeo, M; Sabba, C et al. Multi-detector Row Helical CT assessment of hepatic involvement. *Radiology*, 2004, 230, 250-9.
- [27] García-Tsao, G; Korzenik, JR; Young, L; et al. Liver disease in patients with hereditary hemorrhagic telangiectasia. *N Engl J Med*, 2000, 343, 931-6.
- [28] Buscarini, E; Danesino, C; Plauchu, H; et al. High prevalence of hepatic focal nodular hyperplasia in subjects with hereditary hemorrhagic telangiectasia. *Ultrasound in Medicine & biology*, 2004, 30, 1089-97.

- [29] Matsubara, S; Manzia, JL; Ter, Brugge, K; et al. Angiographic and clinical characteristics of patients with cerebral arteriovenous malformations associated with hereditary hemorrhagic telangiectasia. *AJNR Am J Neuroradiol*, 2000, 21, 1016-20.
- [30] Dowd, CF; William, L; Ter Brugge, K; et al. Quantitative assesment brain arteriovenous malformation multiplicity predicts the diagnosis of hereditary hemorrhagic telangiectasia. *Stroke*, 2012, 43, 72-78.
- [31] Nishida, T; Faughnam, ME; Krings, T; et al. Brain arteriovenous malformations associated with hereditary hemorrhagic telangiectasia: gene-phenotype correlations. *Am J Med Genet A*, 2012, 158, 2829-34.
- [32] McDonald, R; Stoodley, M; Weir, B. Vascular malformations of the central nervous system. *Neurosurgery Quaterly*, 2001, 11, 231-47.
- [33] Easey, AJ; Wallace, GMF; Hugues, JMB; et al. Should asymptomatic patients with hereditary hemorrhagic telangiectasia (HHT) be screened for cerebral vascular malformations? Data from 22061 years of patient life. *J Neurol Neurosurg Psychiatry*, 2003, 74, 743-8.
- [34] Cullen, S; Alvarez, H; Rodesch, G; et al. Spinal arteriovenous shunts presenting before 2 years of age, analysis of 13 cases. *Child's Nervous System*, 2006, 22, 1103-10.
- [35] Rodesch, G; Hurth, M; Alvarez, H; et al. Classification of spinal cord arteriovenous shunts, proposal for a reappraisal: the Bicetre experience with 155 consecutive patients treated between 1981 and 1999. *Neurosurg*, 2002, 51, 374-80.
- [36] Kjeldsen, AD; Kjeldsen, J. Gastrointestinal bleeding in patients with hereditary hemorrhagic telangiectasia. *Am J Gastroenterol*, 2000, 95, 415-8.
- [37] Proctor, DD; Henderson, KZ; Dziura, JD; et al. Enteroscopic evaluation of the gastrointestinal tract in symptomatic patients with hereditary hemorrhagic telangiectasia. *J Clin Gastroenterol*, 2005, 39, 115-9.
- [38] Schwenter, F; Faughnam, ME; Gradinger, AB; et al. Juvenile polyposis, hereditary hemorrhagic telangiectasia and early onset colorectal cancer in patients with SMAD4 mutation. *J Gastroenterol*, 2012, 47, 795-804.
- [39] McDonald, MT; Papenberg, KA; Ghosh, S; Glatfelter, AA; Biesecker, BB; Helmbold, EA; Markel, DS; Zolotor, A; McKinnon, WC; Vanderstoep, JL; Jackson, CE; Iannuzzi, M; Collins, FS; Boehnke, M; Porteous, ME; Gutmacher, AE, y; Marchuk, DA. A disease locus for hereditary haemorrhagic telangiectasia maps to chromosome 9q33-34. *Nat Genet*, 1994, 6, 197-204.
- [40] Fernandez-Ruiz, E; St-Jacques, S; Bellon, T; Letarte, M, y; Bernabeu, C. Assignment of the human endoglin gene (END) to 9q34-->qter. *Cytogenet Cell Genet*, 1993, 64, 204-7.
- [41] McAllister, KA; Grogg, KM; Johnson, DW; Gallione, CJ; Baldwin, MA; Jackson, CE; Helmbold, EA; Markel, DS; McKinnon, WC; Murrell, J; McCormick, MK; Pericak-Vance, MA; Heutink, P; Oostra, BA; Haitjema, T; Westerman, CJJ; Porteous, ME; Gutmacher, AE; Letarte, M, y; Marchuk, DA. Endoglin, a TGF-beta binding protein of endothelial cells, is the gene for hereditary haemorrhagic telangiectasia type 1. *Nat Genet*, 1994, 8, 345-51.
- [42] Johnson, DW; Berg, JN; Baldwin, MA; Gallione, CJ; Marondel, I; Yoon, SJ; Stenzel, TT; Speer, M; Pericak-Vance, MA; Diamond, A; Gutmacher, AE; Jackson, CE; Attisano, L; Kucherlapati, R; Porteous, ME, y; Marchuk, DA. Mutations in the activin receptor-like kinase 1 gene in hereditary haemorrhagic telangiectasia type 2. *Nat Genet*, 1996, 13, 189-95.

- [43] Johnson, DW; Berg, JN; Gallione, CJ; McAllister, KA; Warner, JP; Helmbold, EA; Markel, DS; Jackson, CE; Porteous, ME y; Marchuk, DA. A second locus for hereditary hemorrhagic telangiectasia maps to chromosome 12. *Genome research*, 1995, 5, 21-8.
- [44] Gallione, CJ; Repetto, GM; Legius, E; Rustgi, AK; Schelley, SL; Tejpar, S; Mitchell, G; Drouin, E; Westermann, CJ y; Marchuk, DA. A combined syndrome of juvenile polyposis and hereditary haemorrhagic telangiectasia associated with mutations in MADH4 (SMAD4). *Lancet*, 2004, 363, 852-9.
- [45] Cole, SG; Begbie, ME; Wallace, GM, y; Shovlin, CL. A new locus for hereditary haemorrhagic telangiectasia (HHT3) maps to chromosome 5. *J Med Genet*, 2005, 42, 577-82.
- [46] Bayrak-Toydemir, P; McDonald, J; Akarsu, N; Toydemir, RM; Calderon, F; Tuncali, T; Tang, W; Miller, F, y; Mao, R. A fourth locus for hereditary hemorrhagic telangiectasia maps to chromosome 7. *Am J Med Genet*, 2006, 140, 2155-62.
- [47] Marchuk, DA. Genetic abnormalities in hereditary hemorrhagic telangiectasia. *Curr Opin Hematol*, 1998, 5, 332-8.
- [48] Shovlin, CL. Molecular defects in rare bleeding disorders, hereditary haemorrhagic telangiectasia. *Thromb Haemost*, 1997, 78, 145-50.
- [49] Fernandez-L, A; Sanz-Rodriguez, F; Zarrabeitia, R; Perez-Molino, A; Morales, C; Restrepo, CM; Ramirez, JR; Coto, E; Lenato, GM; Bernabeu, C, y; Botella, LM. Mutation study of Spanish patients with hereditary hemorrhagic telangiectasia and expression analysis of Endoglin and ALK1. *Hum Mutat*, 2006, 27, 295.
- [50] Olivieri, C; Mira, E; Delu, G; Pagella, F; Zambelli, A; Malvezzi, L; Buscarini, E, y; Danesino, C. Identification of 13 new mutations in the ACVRL1 gene in a group of 52 unselected Italian patients affected by hereditary haemorrhagic telangiectasia. *J Med Genet*, 2002, 39, E39.
- [51] Abdalla, SA; Pece-Barbara, N; Vera, S; Tapia, E; Paez, E; Bernabeu, C, y; Letarte, M. Analysis of ALK-1 and endoglin in newborns from families with hereditary hemorrhagic telangiectasia type 2. *Hum Mol Genet*, 2000, 9, 1227-37.
- [52] Gougos, A, y; Letarte, M. Primary structure of endoglin, an RGD-containing glycoprotein of human endothelial cells. *J Biol Chem*, 1990, 265, 8361-4.
- [53] Jovine, L; Darie, CC; Litscher, ES; Wassarman, PM. Zona pellucida domain proteins. *Annu Rev Biochem*, 2005, 74, 83-114.
- [54] Llorca, O; Trujillo, A; Blanco, FJ; Bernabeu, C. Structural model of human endoglin, a transmembrane receptor responsible for hereditary hemorrhagic telangiectasia. *J Mol Biol*, 2007, 365, 694-705.
- [55] Gordon, KJ; Blobe, GC. Role of transforming growth factor-beta superfamily signaling pathways in human disease. *Biochim Biophys Acta*, 2008, 1782, 197-228.
- [56] Massague, J. TGF-beta signal transduction. *Annu Rev Biochem*, 1998, 67, 753-91.
- [57] Pardali, E; Goumans, MJ, y; ten, Dijke, P. Signaling by members of the TGF-beta family in vascular morphogenesis and disease. *Trends Cell Biol*, 2010, 20, 556-67.
- [58] Van, Meeteren, LA, y; Ten, Dijke, P. Regulation of endothelial cell plasticity by TGF-beta. *Cell Tissue Res*, 2011, 347, 177-86.
- [59] Bobik, A. Transforming growth factor-betas and vascular disorders. *Arterioscler Thromb Vasc Biol*, 2006, 26, 1712-20.

- [60] Wieser, R; Wrana, JL, y; Massague, J. GS domain mutations that constitutively activate T beta R-I, the downstream signaling component in the TGF-beta receptor complex. *EMBO J*, 1995, 14, 2199-208.
- [61] Massague, J; Seoane, J, y; Wotton, D. Smad transcription factors. *Genes Dev*, 2005, 19, 2783-810.
- [62] Itoh, S, y; ten, Dijke, P. Negative regulation of TGF-beta receptor/Smad signal transduction. *Curr Opin Cell Biol*, 2007, 19, 176-84.
- [63] Valdimarsdottir, G; Goumans, MJ; Itoh, F; Itoh, S; Heldin, CH, y; ten, Dijke, P. Smad7 and protein phosphatase 1alpha are critical determinants in the duration of TGF-beta/ALK1 signaling in endothelial cells. *BMC Cell Biol*, 2006, 7, 16.
- [64] Guerrero-Esteo, M; Sanchez-Elsner, T; Letamendia, A, y; Bernabeu, C. Extracellular and cytoplasmic domains of endoglin interact with the transforming growth factor-beta receptors I and II. *J Biol Chem*, 2002, 277, 29197-209.
- [65] Blanco, FJ; Santibañez, JF; Guerrero-Esteo, M; Langa, C; Vary, CP, y; Bernabeu, C. Interaction and functional interplay between endoglin and ALK-1, two components of the endothelial transforming growth factor-beta receptor complex. *J Cell Physiol*, 2005, 204, 574-84.
- [66] Velasco, S; Alvarez-Muñoz, P; Pericacho, M; Dijke, PT; Bernabéu, C; López-Novoa, JM; Rodríguez-Barbero, A. L- and S-endoglin differentially modulate TGFbeta1 signaling mediated by ALK1 and ALK5 in L6E9 myoblasts. *J Cell Sci*, 2008, 121, 913-919.
- [67] Santibañez, JF; Letamendia, A; Perez-Barriocanal, F; Silvestri, C; Saura, M; Vary, CP; Lopez-Novoa, JM; Attisano, L, y; Bernabeu, C. Endoglin increases eNOS expression by modulating Smad2 protein levels and Smad2-dependent TGF-beta signaling. *J Cell Physiol*, 2007, 210, 456-68.
- [68] David, L; Feige, JJ; Bailly, S. Emerging role of bone morphogenetic proteins in angiogenesis. *Cytokine Growth Factor Rev*, 2009, 20, 203-212.
- [69] Lebrin, F; Mummery, CL. Endoglin-mediated vascular remodeling: mechanisms underlying hereditary hemorrhagic telangiectasia. *Trends Cardiovasc Med*, 2008, 18, 25-32.
- [70] Bernabeu, C; Conley, BA; Vary, CP. Novel biochemical pathways of endoglin in vascular cell physiology. *J Cell Biochem*, 2007, 102, 1375-1388.
- [71] Bernabeu, C; Lopez-Novoa, JM, y; Quintanilla, M. The emerging role of TGF-beta superfamily coreceptors in cancer. *Biochim Biophys Acta*, 2009, 1792, 954-73.
- [72] Abdalla, SA, y; Letarte, M. Hereditary haemorrhagic telangiectasia, current views on genetics and mechanisms of disease. *J Med Genet*, 2006, 43, 97-110.
- [73] Botella, LM; Sanchez-Elsner, T; Sanz-Rodriguez, F; Kojima, S; Shimada, J; Guerrero-Esteo, M; Cooreman, MP; Ratzu, V; Langa, C; Vary, CP; Ramirez, JR; Friedman, S, y; Bernabeu, C. Transcriptional activation of endoglin and transforming growth factor-beta signaling components by cooperative interaction between Sp1 and KLF6, their potential role in the response to vascular injury. *Blood*, 2002, 100, 4001-10.
- [74] Fonsatti, E; Nicolay, HJ; Altomonte, M; Covre, A; Maio, M. Targeting cancer vasculature via endoglin/cd105: a novel antibody-based diagnostic and therapeutic strategy in solid tumors. *Cardiovasc Res*, 2010, 86(1), 12-9.

- [75] Sanchez-Elsner, T; Botella, LM; Velasco, B; Langa, C, y; Bernabeu, C. Endoglin expression is regulated by transcriptional cooperation between the hypoxia and transforming growth factor-beta pathways. *J Biol Chem*, 2002, 277, 43799-808.
- [76] Botella, LM; Sanchez-Elsner, T; Rius, C; Corbi, A, y; Bernabeu, C. Identification of a critical Sp1 site within the endoglin promoter and its involvement in the transforming growth factor-beta stimulation. *J Biol Chem*, 2001, 276, 34486-94.
- [77] Rius, C; Smith, JD; Almendro, N; Langa, C; Botella, LM; Marchuk, DA; Vary, CP, y; Bernabeu, C. Cloning of the promoter region of human endoglin, the target gene for hereditary hemorrhagic telangiectasia type 1. *Blood*, 1998, 92, 4677-90.
- [78] Sanz-Rodriguez, F; Guerrero-Esteo, M; Botella, LM; Banville, D; Vary, CP, y; Bernabeu, C. Endoglin regulates cytoskeletal organization through binding to ZRP-1, a member of the Lim family of proteins. *J Biol Chem*, 2004, 279, 32858-68.
- [79] Fernandez-L, A; Garrido-Martin, EM; Sanz-Rodriguez, F; Ramirez, JR; Morales-Angulo, C; Zarrabeitia, R; Perez-Molino, A; Bernabeu, C, y; Botella, LM. Therapeutic action of tranexamic acid in hereditary haemorrhagic telangiectasia (HHT), regulation of ALK-1/endoglin pathway in endothelial cells. *Thromb Haemost*, 2007, 97, 254-62.
- [80] Perez-Gomez, E; Eleno, N; Lopez-Novoa, JM; Ramirez, JR; Velasco, B; Letarte, M; Bernabeu, C, y; Quintanilla, M. Characterization of murine S-endoglin isoform and its effects on tumor development. *Oncogene*, 2005, 24, 4450-61.
- [81] Wong, VC; Chan, PL; Bernabeu, C; Law, S; Wang, LD; Li, JL; Tsao, SW; Srivastava, G; Lung, ML. Identification of an invasion and tumor-suppressing gene, Endoglin (ENG), silenced by both epigenetic inactivation and allelic loss in esophageal squamous cell carcinoma. *Int J Cancer*, 2008, 123, 2816-2823.
- [82] Pérez-Gómez, E; Villa-Morales, M; Santos, J; Fernández-Piqueras, J; Gamallo, C; Dotor, J; Bernabéu, C; Quintanilla, M. A role for endoglin as a suppressor of malignancy during mouse skin carcinogenesis. *Cancer Res*, 2007, 67, 10268-10277.
- [83] Hawinkels, LJ; Kuiper, P; Wiercinska, E; Verspaget, HW; Liu, Z; Pardali, E; Sier, CF, y; ten, Dijke, P. Matrix metalloproteinase-14 (MT1-MMP)-mediated endoglin shedding inhibits tumor angiogenesis. *Cancer Res*, 2010, 70, 4141-50.
- [84] Ojeda-Fernandez, L; Barrios, L; Rodriguez-Barbero, A; Recio-Poveda, L; Bernabeu, C; Botella, LM. *Clin Chim Acta*, 2010, Apr 2, 411(7-8), 494-9.
- [85] Li, X; Yonenaga, Y; Seki, T. Shortened ALK1 regulatory fragment maintains a specific activity in arteries feeding ischemic tissues. *Gene Ther*, 2009, 16, 1034-1041,
- [86] Garrido-Martin, EM; Blanco, FJ; Fernandez-L, A; Langa, C; Vary, CP; Lee, UE; Friedman, SL; Botella, LM, y; Bernabeu, C. Characterization of the human Activin-A receptor type II-like kinase 1 (ACVRL1) promoter and its regulation by Sp1. *BMC Mol Biol*, 2010, 11, 51.
- [87] Faughnam, ME; Palda, V; Garcia-Tsao, G; et al. International guidelines for the diagnosis and management of hereditary haemorrhagic telangiectasia. *J Med Genet* 2011, 48, 73-87.
- [88] Sabba, C; Morelli, GA; Logrono, LA. Treatment of bleeding in hereditary hemorrhagic telangiectasia with aminocaproic acid. *N Eng J Med*, 1994, 330, 1789-90.
- [89] Morales, Angulo, C; Pérez, del; Molino, A; Zarrabeitia, R; et al. Treatment of epistaxis in hereditary hemorrhagic telangiectasia (Rendu Osler Weber disease) with tranexamic acid. *Acta Otorrinolaringol Esp*, 2007, 58, 129-32.

- [90] Flieger, D; hainke, S; Fischbach, W. Dramatic improvement in hereditary hemorrhagic telangiectasia after treatment with the vascular endothelial growth factor (VEGF) antagonist bevacizumab. *Ann Hematol*, 2006, 85, 631-2.
- [91] Simonds, J; Miller, F; mandel, J; Davidson, TH. The effect of bevacizumab (Avastin) treatment on epistaxis in hereditary hemorrhagic telangiectasia. *Laryngoscope*, 2009, 119, 988-92.
- [92] Chavan, A; Schumann-Binarsch, S; Luthe, L. Systemic therapy with bevacizumab in patients with HHT. *Vasa*, 2013, 42, 106-10.
- [93] Lebrin, F; Srun, S; Raymond, K; et al. Thalidomide stimulates vessel maturation and reduces epistaxis in individuals with hereditary hemorrhagic. *Nat Med*, 2010, 16, 420-8.
- [94] Garrido, A; Sayago, M; López, J. Thalidomide in refractory bleeding due to gastrointestinal angiodysplasias. *Rev Esp Enferm Dig*, 2012, 104, 69-71.
- [95] McAlister, VC. Regression of cutaneous and gastrointestinal telangiectasia with sirolimus and aspirin in a patient with hereditary hemorrhagic telangiectasia. *Ann Intern Med*, 2006, 144, 226-7.
- [96] Albiñana, V; Sanz-Rodriguez, F; Recio-Poveda, L; et al. Immunosuppressor FK506 increases endoglin and activin receptor-like kinase 1 expression and modulates transforming growth factor- β 1 signaling in endothelial cells. *Mol Pharmacol*, 2011, 79, 833-43.
- [97] Mager, JJ; Gussem, EM; Disch, FJ; et al. A pilot study on the effect of acetylcysteine on epistaxis and quality of life in HHT. *Hematology Meeting Reports*, 2007, 19.
- [98] de, Gussem, EM; Snijder, RJ; Disch, FJ; Zanen, P; Westermann, CJ, y; Mager, JJ. (2009). The effect of N-acetylcysteine on epistaxis and quality of life in patients with HHT, a pilot study. *Rhinology*, 47, 85-8.
- [99] Jameson, JJ; Cave, DR. Hormonal and antihormonal therapy for epistaxis in hereditary hemorrhagic telangiectasia. *Laryngoscope*, 2004, 114, 705-9.
- [100] Korzenik, JR. Danazol in the treatment of gastrointestinal hemorrhage in HHT. *Gastroenterology*, 1995, 108, 207.
- [101] Zacharski, LR; Dunbar, SD; Newsom, WA, Jr. Hemostatic effects of tamoxifen in HHT. *Thrombosis and Haemostasis*, 2001, 85, 371-2.
- [102] Yaniv, E; Preis, M; Shevro, J. Anti-estrogen therapy for HHT-a long term clinical trial. *Rhinology*, 2011, 49, 214-6.
- [103] Albiñana, V; Bernabeu-Herreó, ME; Zarrabeitia, R; et al. Estrogen therapy for hereditary hemorrhagic telangiectasia (HHT):effects of raloxifen on endoglin and ALK1 expression in endothelial cells. *Thromb Haemost*, 2010, 103, 525-34.
- [104] Albiñana, V; Recio-Poveda, L; Zarrabeitia, R; et al. Propranolol as antiangiogenic candidate for the therapy of HHT. *Thromb Haemost* 2012, 108, 41-53
- [105] Shovlin, CL; Sulaiman, NL; Govani, FS; et al. Elevated factor VIII in hereditary haemorrhagic telangiectasia (HHT):association with venous thromboembolism. *Thromb Haemost*, 2007, 98, 1031-9.
- [106] Pagella, F; Matti, E; Chu, F; et al. Argon plasma coagulation is an effective treatment for HHT patients with severe nosebleeds. *Acta Otolaryngol*, 2013, 133, 174-86.
- [107] Karnezis, TT; Davidson, TM. Treatment of hereditary hemorrhagic telangiectasia with submucosal and topical bevacizumab therapy. *Laryngoscope*, 2012, 122, 495-7.

- [108] Morais, D; Millás, T; Zarrabeitia, R; et al. Local sclerotherapy with polydocanol (Aethoxysklerol®) for the treatment of epistaxis in Rendu-Osler-Weber or HHT:15 years of experience. *Rhinology*, 2012, 50, 80-6.
- [109] Harvey, RJ; Kanagalingan, J; Lund, VJ. The impact of septodermoplasty and potassium-titanyl-phosphate (KTP) laser therapy in the treatment of HHT-related epistaxis. *Am J Rhinol*, 2008, 22, 182-7.
- [110] Richer, SL; Geisthoff, VW; Livada, N; et al. The Young's procedure for severe epistaxis from hereditary hemorrhagic telangiectasia. *Am J Rhinol Allergy*, 2012, 26, 401-4.
- [111] Pollak, JS; Saluja, S; Thabet, A; et al. Clinical and anatomic outcomes after embolotherapy of pulmonary arteriovenous malformations. *J Vasc Interv Radiol*, 2006, 17, 35-44.
- [112] Lacombe, P; Lagrange, C; Beauchet, A. Diffuse pulmonary arteriovenous malformations in HHT: long-term results of embolization according to the extent of lung involvement. *Chest* 2009, 135, 1031-7
- [113] Lee, DW; White, RI; Egglin, TK. Embolotherapy of large pulmonary arteriovenous malformations: long-term results. *Ann Thorac Surg*, 1997, 64, 930-9.
- [114] Shovlin, CL; Jackson, JE; Bamford, KB. Primary determinants of ischaemic stroke/brain abscess risks are independent of severity of pulmonary arteriovenous malformations in HHT. *Thorax*, 2008, 63, 259-66.
- [115] Shovlin, C; Bamford, K; Wray, D. Post-NICE 2008:Antibiotic prophylaxis prior to dental procedures for patients with pulmonary arteriovenous malformations and hereditary hemorrhagic telangiectasia. *Br Dent J*, 2008, 205, 531-3.
- [116] Sopena, B; Pérez-Rodríguez, MT; Portela, D. High prevalence of pulmonary hypertension in patients with hereditary hemorrhagic telangiectasia. *Eur J Intern Med*, 2013, 24, 30-4.
- [117] Buscarini, E; Plauchu, H; Garcia-Tsao, G. Liver involvement in hereditary hemorrhagic telangiectasia: consensus recommendations. *Liver Int*, 2006, 26, 1040-6.
- [118] Dupuis-Girod, S; Ginon, I; Saurin, JC. Bevacizumab in patients with hereditary hemorrhagic telangiectasia and severe hepatic vascular malformations and high cardiac output. *JAMA*, 2012, 307, 948-55.
- [119] Buscarini, E; Leandro, G; Conte, D; et al. Natural history and outcome of hepatic vascular malformations in a large cohort of patients with hereditary hemorrhagic telangiectasia. *Dig Dis Sci*, 2011, 56, 2166-78.
- [120] De, Cillis, E; Burdi, N; Bortone, AS; et al. Endovascular treatment of pulmonary and cerebral arteriovenous malformations in patients affected by HHT. *Curr Pharm Des*, 2006, 12, 1243-8.
- [121] Meisel, HJ; Mansmann, U; Alvarez, H; et al. Cerebral arteriovenous malformations and associated aneurysms: analysis of 305 cases from a series of 662 patients. *Neurosurgery*, 2000, 46, 793-800.
- [122] Mathis, S; Dupuis-Girod, S; Plauchu, H; et al. Cerebral abscesses in hereditary haemorrhagic telangiectasia: a clinical and microbiological evaluation. *Clin Neurol Neurosurg*, 2012, 114, 235-40.
- [123] Swietlik, E; Doboszynska, A. Recurrence of arteriovenous malformations with life-threatening complications in a pregnant women with hereditary hemorrhagic telangiectasia. *J Physiol Pharmacol*, 2008, 59, 683-8.

- [124] Goussous, T; Heynes, A; Najarian, K. Hereditary hemorrhagic telangiectasia patient presenting a high cardiac failure during pregnancy. *Cardiol Res Pract*, 2009.
- [125] Shovlin, CL; Sodhi, V; McCarthy, A. estimates of maternal risks of pregnancy for women with hereditary hemorrhagic telangiectasia (Osler-Weber-Rendu syndrome): suggested approach for obstetric services. *BJOG*, 2008, 115, 1108-15.
- [126] Wain, K; Swanson, K; Watson, W; et al. Hereditary hemorrhagic telangiectasia and risks for adverse pregnancy outcomes. *Am J Med Genet A*, 2012, 2009-14.
- [127] Giordano, P; Lenato, GM; Suppresa, P; et al. Hereditary hemorrhagic telangiectasia: arteriovenous malformations in children. *J Pediatr*, 2013, 13.
- [128] Mei-Zahav, M; Letarte, M; Faughnan, ME; et al. Symptomatic children with hereditary hemorrhagic telangiectasia, a pediatric center experience. *Arch Pediatr Adolesc Med*, 2006, 160, 596-601.

Chapter 85

OSTEOGENESIS IMPERFECTA

Colin R. Paterson*

Medicine in the University of Dundee,
Dundee, Scotland

ABSTRACT

Osteogenesis imperfecta is the most common heritable cause of fractures in children. It is not a single disorder but a large group of diseases most, but not all, being caused by defects of the genes coding for collagen. Autosomal dominant inheritance is the most common finding in familial cases but new mutations occur. Autosomal recessive inheritance does occur and mosaicism is recognized.

The great molecular heterogeneity is reflected in great clinical and radiological variation. Some cases are so severe that survival beyond intra-uterine life is impossible. At the other extreme some patients, undoubtedly affected their family history, have few fractures and live normal lives. Fractures often occur spontaneously and previously asymptomatic fractures in various stages of healing are often found radiologically. While most symptomatic fractures are diaphyseal, all types of fractures including metaphyseal fractures, rib fractures and skull fractures do occur.

Modern management involves good orthopaedic surgery; it is particularly important to avoid prolonged immobilisation of limbs to avoid superimposed osteopenia. Drug therapy, particularly with pamidronate, may be appropriate in children with the more severe forms of the disorder. Specific attention may be needed to scoliosis, basilar invagination or deafness. Good occupational therapy to maximise mobility is important. Children with osteogenesis imperfecta have normal intelligence and good education is vital.

* Corresponding Author's Email: c.s.paterson@btinternet.com (Dr. C R Paterson, Temple Oygates, Longforgan, Dundee DD2 5HS, UK; Tel: +44 1382 360240).

INTRODUCTION

Osteogenesis imperfecta (OI) is the most common bone dysplasia causing fractures in childhood. In western countries it has a prevalence of about 1 in 10,000. It occurs in all races.

CLASSIFICATION AND GENETICS

The most widely used classification of OI based on clinical and radiological features is that of Sillence et al. [1]. These authors suggested four major types, each of which has subsequently been subdivided (Table 1). It is important to recognize that there is little relationship between the Sillence type and the underlying mutation. Each Sillence type includes patients with many different mutations. Nevertheless the Sillence classification has been invaluable over the years in allowing consistent reporting of the clinical features of each case.

Table 1. Clinical features of the different types of OI in the Sillence scheme together with principal subsequent additions

Type	Features	Genetics
I*	Mild disease usually. Blue or grey sclerae. Normal growth	Autosomal dominant New mutations frequent
II**	Severe disease leading to multiple intrauterine fractures, stillbirth or early neonatal death	Autosomal recessive New mutations
III	Severe disease with fractures at birth and progressive deformity. Often very short stature. Dentinogenesis imperfecta common	Most commonly new mutations of autosomal dominant disorder
IV*	Mild to moderate severity. Normal sclerae except in infancy	Autosomal dominant. New mutations frequent
V	Moderate to severe bone fragility. Hypertrophic callus after fractures. Calcification of interosseous membrane	Autosomal dominant
VI	Moderate to severe bone fragility and deformity. Distinctive bone histology	Autosomal dominant
VII	Moderate to severe disease	Autosomal recessive

* Subdivided: IA and IVA have normal teeth; IB and IVB have dentinogenesis imperfecta.

** Subdivided: A, B and C in relation to the radiological findings.

In recent years a group of clinicians in Montreal has suggested the addition of three further types, V, VI and VII (Table 1). Type V patients were earlier included in type IV but were distinct in having a tendency to form hyperplastic callus after a fracture and calcification of the interosseous membrane of the forearm. These patients were also distinctive in having an IFTM5 mutation [2]. Patients with OI type VI were also clinically similar to type IV patients but had distinctive bone histology with an excess of osteoid. They had a distinctive mutation in

SERPINF1 [3]. OI type VII is the name given to an unusual form of severe OI inherited in an autosomal recessive manner found in a small group of indigenous people in northern Quebec [4]. It is not thought to be caused by a mutation in the collagen genes.

Even among patients with the same mutation there is considerable variation in clinical severity [5]. Similarly within a single family the severity of the disorder may vary greatly. The reason for these variations is not known. One exception is provided by parental mosaicism in which, for example, a clinically normal parent or a mildly affected parent has a severely affected child [6-8]. A further cause for variation within a family occurs when parents who are heterozygous for a mutation and show minor features of OI have a child who is homozygous for the mutation and severely affected [9].

BIOCHEMICAL CAUSES

The mechanical strength of bone depends in part on its content of type I collagen and abnormalities of collagen have been thought since the 1970s to underlie OI [10]. Chemical studies of collagen produced by cultured fibroblasts from patients with OI showed abnormalities. In addition families with dominantly inherited OI had consistent linkage to the two genes coding for type I collagen, COL1A1 and COL1A2 [10, 11].

In 1985 the first mutation was identified, an internal deletion in COL1A1 that had caused a lethal form of OI [12, 13]. By 2007 no less than 832 distinct mutations had been reported [14]. Most of these are private mutations, mutations unique to one individual or family. Deletions are now thought to be an uncommon cause of OI. Exon skipping, due to mutations at splice donor or acceptor sites, is more common. The most common type of mutation by far causes the substitution of a single amino acid.

The structure of a collagen molecule is a triple helix of two alpha 1 chains and one alpha 2 chain. Both contain glycine residues at every third position. These are important structurally; their replacement by bulkier residues distorts the triple helix and prevents the neat packing of molecules into fibrils. Most of the recognized mutations in OI involve substitutions of glycine [15, 16].

The correlation between the nature of the mutation and the clinical features in affected patients is very imperfect. For example substitution of glycine by serine in the alpha 1 chain causes a more severe disease than the same substitution in the alpha 2 chain [5]. In addition a single mutation in apparently unrelated families can result in substantial variation in the clinical phenotype in terms of severity and associated findings [5]. One further cause of clinical variability is somatic mosaicism; an unaffected or mildly affected parent may be associated with a severely affected child [6, 7]. In general blue sclerae are more associated with COL1A1 mutations, particularly those near the amino terminal. Dentinogenesis imperfecta (DI) was most common with mutations in the central region and near the carboxy terminal both in COL1A1 and COL1A2.

It should be noted that mutations in the genes coding for type 1 collagen are not the only causes of OI. It has long been recognized that a minority of patients with OI have no evidence of a mutation in COL1A1 or COL1A2. Some patients have a clearly autosomal recessive inheritance. Many of these disorders are now known to be caused by mutations affecting the proteins involved in the formation of mature collagen.

The formation of collagen is a complex process. The precursors of the alpha 1 and alpha 2 chains are known as procollagens and have peptide chains at both amino- and carboxy-terminals in addition to the future helical portions. These procollagen chains undergo several further steps including the hydroxylation of some proline residues to hydroxyproline with the enzyme prolyl hydroxylase. The procollagen chains come together to give the triple helix and the additional chains at each end are cleaved off. Mutations affecting one of the cleavage sites cause an unusual form of OI with dense bones [17, 18].

Mutations leading to OI, often as a recessive disorder, have now been identified in genes coding for each of the proteins involved in a complex consisting of prolyl 3-hydroxylase, cartilage-associated proteins (CRTAP) and cyclophilin B [19, 20]. Further genes in which mutations have been shown to cause recessively inherited OI include FKBP10, which also causes Bruck syndrome with contractures, and SERPINF 1, which causes OI type VI [3, 21]. OI type V is an autosomal dominant disorder caused by mutations in the gene IFITM5 which codes for a skeletal protein, BRIL, whose function is not yet known [2]. The many mutations that have been shown to underlie OI are recorded in a continuously updated database [22]. Collagen maturation is dependent on the availability of both copper and ascorbic acid. Collagen defects leading to symptoms including fractures are well recognized in copper deficiency and in scurvy.

CLINICAL FEATURES

While the hallmark of the disorder is the increased tendency to fractures other clinical signs may be present. These include blue or grey sclerae, increased joint laxity, impaired growth, liability to bruising, odd shaped skull, dentinogenesis imperfecta, excessive sweating and premature deafness. Most but not all of these clinical features can be associated with the disorders of collagen which are thought to underlie most cases. Few patients have all these signs and some, undoubtedly affected on the basis of family history, have none.

Fractures

The fractures may occur with little or no recognised trauma or with normal handling. Some asymptomatic fractures may be found when X-rays are taken for other reasons. Fractures usually occur without evidence of any local bruising because little trauma is involved. The increased liability to bruising or petechial haemorrhages is thought to result from the increased fragility of small blood vessels [23]. The paradox that fractures occur with few bruises in a child who also has an increased liability to bruising is shared with other bone disorders that cause fractures [24].

In OI fractures may be unpredictable; there may be long gaps between symptomatic fractures in an individual patient (Figure 1). It should be noted that the fracture rate diminishes in the teenage years in all types of OI. In older men the fracture rate remains low in later life but in women there is an increase in fracture rate, particularly vertebral crush fractures after the menopause (Figure 2) [25].

The fractures of OI may include symptomatic long bone fractures, including transverse, oblique and spiral fractures [26]. Metaphyseal fractures are well recognized and usually asymptomatic (Figure 3).

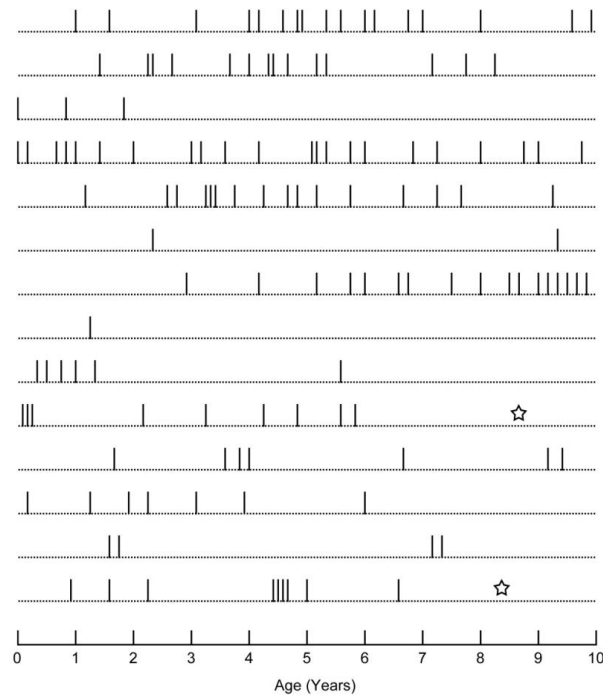


Figure 1. Incidence of fractures in 14 patients with OI type IVA for whom complete records were available for at least eight years. In two cases details were not available for later than the point marked with a star. Data of Paterson et al. 1987 [77].

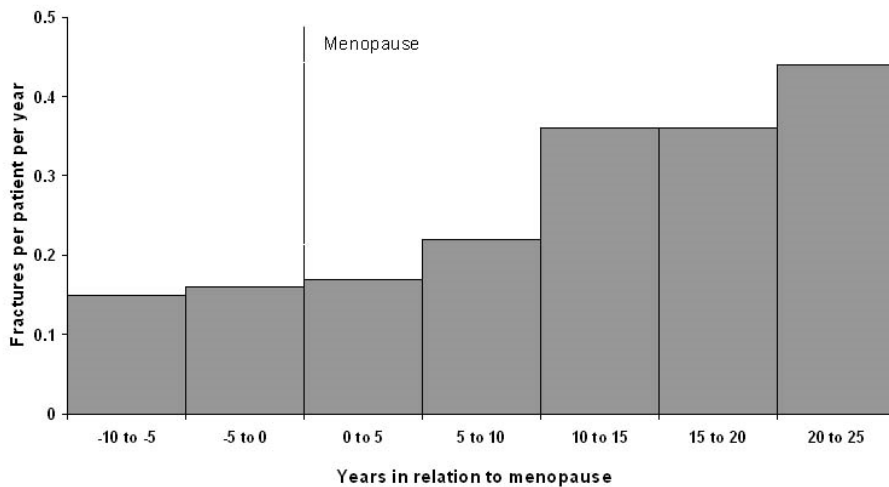


Figure 2. Fractures per patient per year in relation to time of the menopause in 30 women with OI type IA. Data of Paterson et al. 1984 [25].

Skull fractures are well-recognized [27]. Rib fractures occur and are usually asymptomatic. That these do occur is illustrated by Figure 4 which shows a child on the day of birth later thought to have OI type III. Multiple rib fractures of different dates are seen; all occurred *in utero*.



Figure 3. Metaphyseal fractures of both femora in a three day old child later recognized as having OI type IVA.



Figure 4. Chest x-ray of a boy on the day of birth showing multiple rib fractures of different ages, all of which had occurred *in utero*. He was later thought to have OI type III.

One specific cause of symptomatic fractures in OI needs to be highlighted. The standard tests for congenital dislocation of the hip are those of Ortolani and of Barlow. Carrying out such tests in infants with OI, or whose family history implies that they may have OI, is not advisable; fractures of the femur may occur (Figure 5) [28].



Figure 5. Bilateral femoral fractures found after an examination for congenital hip dislocation in an infant girl with a family history of OI type IVB.

Dental Abnormalities

Dentine also contains collagen and a substantial proportion of OI patients have overt dentinogenesis imperfecta (DI). The teeth are discoloured, translucent and fragile. The severity varies greatly among OI patients. Detailed study of the teeth, including scanning electron microscopy, has shown that all OI patients have abnormalities in the teeth [29]. It should be noted that DI can exist without OI; this disorder is known as DI type II [30]. It is also inherited as an autosomal dominant.

Sclerae

The grey or blue sclerae are among the best known features of OI. They are present in about two thirds of patients and play a part in the conventional classification. The scleral discolouration is thought to reflect thinness of scleral collagen so that the underlying pigment shows through. The assessment of scleral colour is not an exact science; attempts to quantify it using a paint colour strip have not been widely used. Also it should be noted that some blue colouration is seen in normal infants.

Premature Deafness

Hearing loss can occur in all types of OI. Most of the patients who develop hearing loss have their first symptom before the age of 40 but a few have a later age of onset (Table 2) [31, 32].

Table 2. Age of onset of hearing impairment in 77 patients out of 133 OI patients aged 17 or more (32)

Age range	Number of patients
0-10	1
10-20	19
20-30	27
30-40	12
40-50	10
50-60	4
60-70	2
70-80	1
80-90	1

Table 3. Proportion of OI patients of each type who had symptoms of hearing impairment at age 30 [31]

Sillence type	Number of patients aged 30 or more	Number of patients with hearing impairment at age 30	Percentage affected at age 30
IA	276	90	33
IB	60	18	30
III	23	12	52
IVA	32	3	9
IVB	52	15	29
Uncertain	18	7	39

Table 3 shows the likelihood of developing deafness in each type of OI. Most surveys have shown that the deafness associated with OI has no single cause. Most patients have a mixed pattern; smaller numbers have sensorineural or conductive patterns [32, 33].

Cardiovascular Problems

Hypertension appears to be common in adults with OI [34]. Valvular heart disorders, particularly aortic and mitral regurgitation, are also found and often require surgery [34, 35].

Excessive Sweating

Excessive sweating has long been recognized as a symptom in children with OI but no explanation can be made from collagen biochemistry. A postoperative febrile response is more common in children with OI than in normal controls but appears to be harmless [36].

Joint Laxity

Increased laxity of joints is seen in a substantial minority of patients with OI. It may lead to confusion with the Ehlers-Danlos syndrome, some forms of which also result from mutations in the genes coding for collagen. Joint laxity may contribute to the disability resulting from OI, for example in permitting hyperextension of the fingers or in causing instability of the ankles.

Neurological Problems

Basilar invagination is an uncommon but very serious complication of OI and some other bone disorders. It consists of the upward displacement of parts of the occipital bone surrounding the foramen magnum leading to pressure from the upper cervical spine on the brainstem. Symptoms include headache in the neck and occiput, often worse with coughing, sneezing or straining, trigeminal neuralgia, giddiness, weakness in the arms or legs and bladder disorders [37]. Among the physical signs are nystagmus, facial spasm, nerve paresis, proprioceptive defects, pyramidal tract signs and papilloedema [37]. The change in the skull shape may lead to changes in the face including recession of the maxilla and prominence of the mandible (prognathism). While basilar invagination occurs in all types of OI, basilar invagination with neurological consequences is most common in OI type IV B [37].

Intracranial haemorrhage may occur in OI [38, 39] and is a significant cause of death [40].

Impaired Growth

Post-natal growth is normal in many cases of OI type I, reduced in most type IV cases and substantially reduced in OI type III. There is no evidence of growth hormone deficiency [41, 42].

Temperament

One striking feature of many OI patients is their generally positive approach to life often despite appreciable disabilities. This has been confirmed formally [43, 44]. Educational achievements are usually normal despite disabilities [43].

Life Expectancy

Life expectancy is normal in OI type IA, marginally reduced in OI types IB, IVA and IVB, and appreciably reduced in OI type III [45]. Most patients with type I and type IV forms of OI have a normal lifespan and die of causes unrelated to their OI. In many type III patients premature deaths are caused, particularly in childhood, by respiratory problems resulting from kyphoscoliosis. OI plays a direct part in deaths due to basilar invagination and intracranial haemorrhage.

RADIOLOGY AND DENSITOMETRY

The wide range of clinical severity in OI is reflected in a wide range of radiological appearances. Patients with OI types II and III usually have obvious abnormalities from and often before the time of birth. These include fractures which have occurred *in utero* (Figure 4). Most patients with type I and type IV OI have radiologically normal bone at birth. Radiological appearances often remain normal throughout life in bones which have not sustained fractures.

One radiological sign may be helpful in diagnosis. Wormian bones are additional bones, usually in the occipital part of the skull, completely surrounded by a suture line. Conventionally more than 10 such bones are regarded as significant for the diagnosis of OI. In a recent study significant numbers of Wormian bones were seen in 35 percent of patients with OI type I, in 78 percent of patients with OI type IV and in 96 percent of patients with OI type III [46]. It is important to note that excessive numbers of Wormian bones may be found in other disorders including cretinism, cleido-cranial dysostosis and Menkes syndrome [47].

Densitometry

Attempts to assess bone density from ordinary X-rays are imprecise and should not be used [48]. It is inappropriate to comment that a “normal” appearance on ordinary X-rays excludes OI or any other bone disease. However it is true that a substantial number of OI patients have low values when formal densitometry is undertaken.

In our experience of adults with OI (for whom we had a well established reference range) most patients were below the mean for the reference group but many had a higher bone density than the lower reference limit (minus 2 standard deviations) [49]. Similar findings were obtained with DXA of the spine [50, 51]. In one study [51] the authors commented that in “some mildly affected patients brittleness may exist with only small reductions in bone mineral content.”

In children there are only small studies and none of the reference ranges used were satisfactory [52-54]. However it is likely that, as in adults, children with OI have lower bone densities compared with controls but with an appreciable overlap. One difficulty in this field is the continuing lack of robust reference ranges appropriate for the densitometers used.

One further difficulty affecting both adults and children is that a fracture with a period of immobilisation may itself lead to diminished bone density of the limb concerned. This may lead to false assumptions about bone density in the underlying condition. This problem is illustrated in Figure 6.



Figure 6. Hand and foot of a young woman with OI type IVB. The hand (left) shows normal appearances while the foot appears severely abnormal. During her childhood and adolescence she had had multiple fractures in the legs.

CLINICAL BIOCHEMISTRY

Routine investigations such as serum calcium, inorganic phosphate and alkaline phosphatase usually show no abnormality in OI. One exception is that serum alkaline phosphatase may rise if there are multiple healing fractures. It is important to note that normal findings in these tests do not exclude OI.

A number of other biochemical abnormalities have been found in some patients with OI. These include a high urine calcium [55], a high urinary excretion of cross-linked collagen peptides [56] and low serum levels of carboxy-terminal propeptides of human type I procollagen (PICP) [57, 58]. None of these investigations has been evaluated as a test for OI.

One recent study suggested that some children with OI may also be deficient in vitamin D [59]. The actual levels were not given but it seems likely that children with any disability are at risk of spending less time exposed to sunshine than their peers.

Table 4. Percentage of positive findings in known cases of osteogenesis imperfecta investigated by collagen analysis and by mutation detection

OI type	Collagen analysis	Mutation detection
Type I	94%	94%
Type III	84%	81%
Type IV	84%	69%
Type unknown	50%	—

Collagen analysis data of Wenstrup et al. [76]. Mutation detection data in personal communication from Dr D Prockop May 2000

Biochemical investigations used for the diagnosis of OI are of two main types. Some examine the chemistry of the collagen from cultured fibroblasts derived from skin biopsies [60]. Others search for mutations in the genes coding for type I collagen. These tests are useful in the confirmation of OI in cases in which it is not clinically obvious. However there is no information on the likelihood of finding abnormalities in a large unselected group of OI patients. As a result we have little information on the frequency of false negatives. Table 4 shows the findings in two studies; they indicate that in both false negatives are most likely in patients in whom the clinical findings are least helpful. It is not correct to say that a negative result excludes OI.

DIFFERENTIAL DIAGNOSES

In the past numerous cases of OI have initially been diagnosed as child abuse [61-64]. Children with OI may have fractures that the parents or carers cannot explain (Figure 7). They may then be found to have other fractures not clinically suspected; they may have multiple fractures of different ages. The bones may otherwise appear to be normal on ordinary X-rays. It is not surprising therefore that cases of misdiagnosis occur. However it is likely that a substantially larger number of cases of misdiagnosis result from other bone disorders which also mimic accepted descriptions of abuse [24]. These include the osteopathy of prematurity and vitamin D deficiency rickets.



Figure 7. Fracture of left femur in a boy of four months, one of twins. Since the bone appeared normal radiologically it was assumed that considerable force was involved and a care order was sought for both boys. However it later became clear that his father, uncle and grandfather all had pale blue sclerae, dentinogenesis imperfecta and a typical history of OI. By the time of the hearing the boy and his twin brother had obvious dentinogenesis imperfecta.

MANAGEMENT

It is likely that occupational therapists, physiotherapists and orthopaedic surgeons have more to contribute to the management of these patients than any drug therapy but it is important to review the current position.

Drug Therapy

Many drugs have been tried in OI including calcium, anabolic steroids, vitamin C, calcitonin and fluoride. None of these has been shown to be of value in properly controlled trials. The practical difficulties in mounting such trials in OI are substantial because of the variable natural history..

Short stature is a common feature of OI and several trials of growth hormone have been undertaken [65]. While growth velocity increased in the treated group the differences were small and there was no evidence that the final height would be improved. There is no evidence that growth hormone deficiency plays any part in the short stature of OI.

Numerous trials have been undertaken in recent years of various bisphosphonates including alendronate, neridronate, olpadronate, risdronate and zoledronic acid. By far the greatest number of studies have involved pamidronate usually by cyclic intravenous infusions [66, 67]. All the studies found that bone turnover decreased and bone mineral density increased. Many studies have reported a decrease in the number of fractures in the more severe types of OI. However, in some of these the comparison was only with pre-treatment fracture rates; it should be remembered that the fracture rate in untreated OI is extremely variable, often decreasing with age.

The underlying premise in many of these studies is that a rise in bone density is in itself a good thing. However it should be recognised that bisphosphonate treatment in anyone will decrease bone turnover and increase bone density. Decreased bone remodelling may have long term disadvantages for fracture risk. At present it seems that cyclic bisphosphonate therapy is probably advantageous for children with the more severe types of OI but not for adults or for children with milder forms of OI.

One report described clinical responses to bone marrow transplantation in children with severe OI [68]. While initial results were encouraging this strategy does not appear to have been pursued since.

In postmenopausal women with OI the case for low-dose hormone replacement therapy probably outweighs the case against.

Pain Relief

In children, as in adults, pain relief is important after fractures. Paracetamol is effective in mild to moderate pain. For more severe pain non-steroidal anti-inflammatory drugs may be helpful as are opioids.

Orthopaedic Surgery

Good orthopaedic care is needed for fractures to avoid deformity, and therefore increased risk of subsequent fractures. However the period of immobilisation should be minimised to reduce the risk of superimposed disuse osteoporosis. Various forms of intramedullary fixation including telescopic rods can be used in children with recurrent fractures [69].

Scoliosis is common in OI. The conventional view is that if the scoliosis does not exceed 45° conservative management is appropriate. Bracing is ineffective. If the scoliosis exceeds 45° it is thought that posterior spinal fusion is justifiable. However surgery is often difficult because of bleeding and the innate fragility of the bone. Newer techniques appear to be effective [70].

Surgery for Basilar Invagination

Basilar invagination with neurological complications is treated surgically by a transoral clivectomy. This not only halts disease progression in most patients but also provides a sustainable long-term functional outcome [71, 72].

Deafness

As with deafness in general the first step involves the use of hearing aids. In OI, particularly with conductive deafness, stapes surgery usually has good long-term results [73, 74]. In appropriate cases cochlear implantation is a safe and feasible procedure [75].

Dental Care

Good dental care is important in OI, particularly in patients with overt dentinogenesis imperfecta. The teeth in this condition are not only discoloured but fragile. Veneers may be useful for cosmetic reasons but do not contribute to strength.

Some cola drinks are particularly undesirable because of their acidity. Supplemental fluoride may be desirable.

Education

Despite the disadvantage of repeated hospital admissions children with OI do well in education; many have above average potential [43]. A good education is of vital importance for the future. In general, mainstream education is usually preferable to schools for the disabled.

ACKNOWLEDGMENTS

I am grateful to Dr. Elspeth Paterson for her careful review of this paper in draft, and to Mr. Jake McLaughlin and Ms. Kirsteen Bovill for excellent secretarial assistance.

REFERENCES

- [1] Sillence DO, Senn A, Danks DM. Genetic heterogeneity in osteogenesis imperfecta. *J. Med. Genet.* 1979;16: 101-16.
- [2] Rauch F, Moffat P, Cheung M et al. Osteogenesis imperfecta type V: marked phenotypic variability despite the presence of the IFITM5 c. 14C>T mutation in all patients. *J. Med. Genet.* 2013;50: 21-4.
- [3] Homan EP, Rauch F, Grafe I et al. Mutations in SERPINF1 cause osteogenesis imperfecta type VI. *J. Bone Miner. Res.* 2011;26: 2798-803.
- [4] Ward LM, Rauch F, Travers R et al. Osteogenesis imperfecta type VII: an autosomal recessive form of brittle bone disease. *Bone* 2002;31: 12-8.

- [5] Rauch F, Lalic L, Roughley P, Glorieux FH. Genotype-phenotype correlations in nonlethal osteogenesis imperfecta caused by mutations in the helical domain of collagen type I. *Eur. J. Med. Genet.* 2010;18: 642-7.
- [6] Cohen-Solal L, Zolezzi F, Pignatti PF, Mottes M. Intrafamilial variable expressivity of osteogenesis imperfecta due to mosaicism for a lethal G382R substitution in the COL1A1 gene. *Mol. Cell. Probes.* 1996;10: 219-25.
- [7] Lund AM, Nicholls AC, Schwartz M, Skovby F. Parental mosaicism and autosomal dominant mutations causing structural abnormalities of collagen I are frequent in families with osteogenesis imperfecta type III/IV. *Acta Paediatr.* 1997;86: 711-8.
- [8] Pyott SM, Pepin MG, Schwarze U, Yang K, Smith G, Byers PH. Recurrence of perinatal lethal osteogenesis imperfecta in sibships: parsing the risk between parental mosaicism for dominant mutations and autosomal recessive inheritance. *Genet. Med.* 2011;13: 125-30.
- [9] De Paepe A, Nuytinck L, Raes M, Fryns J. Homozygosity by descent for a COL1A2 mutation in two sibs with severe osteogenesis imperfecta and mild clinical expression in the heterozygotes. *Hum. Genet.* 1997;99: 478-83.
- [10] Sykes B, Ogilvie D, Wordsworth P et al. Consistent linkage of dominantly inherited osteogenesis imperfecta to the type I collagen loci: COL1A1 and COL1A2. *Am. J. Hum. Genet.* 1990;46: 293-307.
- [11] Mottes M, Cugola L, Cappello N, Pignatti PF. Segregation analysis of dominant osteogenesis imperfecta in Italy. *J. Med. Genet.* 1990;27: 367-70.
- [12] Chu ML, Gargiulo V, Williams CJ, Ramirez F. Multiexon deletion in an osteogenesis imperfecta variant with increased type III collagen mRNA. *J. Biol. Chem.* 1985;260: 691-4.
- [13] Barsh GS, Roush CL, Bonadio J, Byers PH, Gelinas RE. Intron-mediated recombination may cause a deletion in an $\alpha 1$ type I collagen chain in a lethal form of osteogenesis imperfecta. *Proc. Nat. Acad. Sci. USA* 1985;82: 2870-4.
- [14] Marini JC, Forlino A, Cabral WA et al. Consortium for osteogenesis imperfecta mutations in the helical domain of type I collagen: regions rich in lethal mutations align with collagen binding sites for integrins and proteoglycans. *Hum. Mutat.* 2007;28: 209-21.
- [15] Roughley PJ, Rauch F, Glorieux FH. Osteogenesis imperfecta – clinical and molecular diversity. *Eur. Cell. Mater.* 2003;5: 41-7.
- [16] Forlino A, Cabral WA, Barnes AM, Marini JC. New perspectives on osteogenesis imperfecta. *Nature Rev. Endocr.* 2011;7: 540-57.
- [17] Lindahl K, Barnes AM, Fratzl-Zelman N et al. COL1 C-propeptide cleavage site mutations cause high bone mass osteogenesis imperfecta. *Hum. Mutat.* 2011;32: 598-609.
- [18] Takagi M, Hori N, Chinen Y et al. Heterozygous C-propeptide mutations in COL1A1: osteogenesis imperfecta type IIC and dense bone variant. *Am. J. Med. Gene A* 2011;155A: 2269-73.
- [19] Morello R, Bertin TK, Chen Y et al. CRTAP is required for prolyl 3-hydroxylation and mutations cause recessive osteogenesis imperfecta. *Cell* 2006;127: 291-304.
- [20] Van Dijk FS, Nesbitt IM, Zwikstra EH et al. PPIB mutations cause severe osteogenesis imperfecta. *Am. J. Hum. Genet.* 2009;85: 521-7.

- [21] Kelley BP, Malfait F, Bonafe L et al. Mutations in FKBP10 cause recessive osteogenesis imperfecta and Bruck syndrome. *J. Bone Miner. Res.* 2010;26: 666-72.
- [22] Dalgleish RA *Database of osteogenesis imperfecta and Ehlers-Danlos syndrome variants*. www.le.ac.uk/ge/collagen
- [23] Evensen SA, Myhre L, Stormorken H. Haemostatic studies in osteogenesis imperfecta. *Scand. J. Haematol.* 1984;33: 177-9.
- [24] Paterson CR. Bone disease and fractures in early childhood. In: Turner RA, Rogers HO, eds *Child Abuse*. New York: Nova, 2012 pp #27-52.
- [25] Paterson CR, McAllion S, Stellman JL. Osteogenesis imperfecta after the menopause. *N. Engl. J. Med.* 1984;310: 1694-6.
- [26] Dent JA, Paterson CR. Fractures in early childhood: osteogenesis imperfecta or child abuse? *J. Pediatr. Orthop.* 1991;11: 184-6.
- [27] Charnas LR, Marini JC. Communicating hydrocephalus, basilar invagination, and other neurological features in osteogenesis imperfecta. *Neurology* 1993; 43: 2603-8.
- [28] Paterson CR, Beal RJ, Dent JA. Osteogenesis imperfecta: fractures of the femur when testing for congenital dislocation of the hip. *BMJ* 1992;305: 464-6.
- [29] Hall RK, Maniere MC, Palamara J, Hemmerle J. Odontoblast dysfunction in osteogenesis imperfecta: an LM, SEM, and ultrastructural study. *Connect Tissue Res.* 2002;43: 401-5.
- [30] Malmgren B, Lindskog S, Elgadi A, Norgren S. Clinical, histopathologic, and genetic investigation in two large families with dentinogenesis imperfecta type II. *Hum. Genet.* 2004;114: 491-8.
- [31] Paterson CR, Monk EA, McAllion SJ. How common is hearing impairment in osteogenesis imperfecta? *J. Laryngol. Otol.* 2001; 115: 280-2.
- [32] Kuurila K, Kaitila I, Johansson R, Grénman R. Hearing loss in Finnish adult patients with osteogenesis imperfecta, a nationwide survey. *Ann. Otol. Rhinol. Laryngol.* 2002;111:939-46.
- [33] Pillion JP, Shapiro J. Audiological findings in osteogenesis imperfecta. *J. Am. Acad. Audiol.* 2008;19:595-601.
- [34] Radunovic Z, Wekre LL, Diep LM, Steine K. Cardiovascular abnormalities in adults with osteogenesis imperfecta. *Am. Heart J.* 2011;161: 523-9.
- [35] Wong RS, Follis FM, Shively BK, Wernly JA. Osteogenesis imperfecta and cardiovascular diseases. *Ann. Thorac. Surg.* 1995;60: 1439-43.
- [36] Ghert M, Allen B, Davids J, Stasikelis P, Nicholas D. Increased postoperative febrile response in children with osteogenesis imperfecta. *J. Pediatr. Orthop.* 2003;23: 261-4.
- [37] Sillence DO. Craniocervical abnormalities in osteogenesis imperfecta: genetic and molecular correlation. *Pediatr. Radiol.* 1994;24: 427-30.
- [38] Sasaki-Adams D, Kulkarni A, Rutka J, Dirks P, Taylor M, Drake JM. Neurosurgical implications of osteogenesis imperfecta in children. *J. Neurosurg. Pediatrics* 2008;1: 229-36.
- [39] Faqeih E, Roughley P, Glorieux FH, Rauch F. Osteogenesis imperfecta type III with intracranial hemorrhage and brachydactyly associated with mutations in exon 49 of COL1A2. *Am. J. Med. Genet. A.* 2009;149A: 461-5.
- [40] McAllion SJ, Paterson CR. Causes of death in osteogenesis imperfecta. *J. Clin. Path.* 1996;49: 627-30.
- [41] Vetter U, Pontz B, Zauner E, Brenner RE, Spranger J. Osteogenesis imperfecta: a clinical study of the first ten years of life. *Calcif Tissue Int* 1992;50: 36-41.

- [42] Lund AM, Muller J, Skovby F. Anthropometry of patients with osteogenesis imperfecta. *Arch. Dis. Child* 1999;80: 524-8.
- [43] Widmann RF, Laplaza FJ, Bitan FD, Brooks CE, Root L. Quality of life in osteogenesis imperfecta. *Int. Orthop.* 2002;26: 3-6.
- [44] Suskauer SJ, Cintas HL, Marini JC, Gerber LH. Temperament and physical performance in children with osteogenesis imperfecta. *Pediatrics* 2003;111: E153-61.
- [45] Paterson CR, Ogston SA, Henry RM. Life expectancy in osteogenesis imperfecta. *BMJ* 1996;312: 351.
- [46] Semler O, Cheung MS, Glorieux FH, Rauch F. Wormian bones in osteogenesis imperfecta: correlation to clinical findings and genotype. *Am. J. Med. Genet. A* 2010;152A:1681-7.
- [47] Cremin B, Goodman H, Spranger J, Beighton P. Wormian bones in osteogenesis imperfecta and other disorders. *Skeletal Radiol.* 1982;8: 35-8.
- [48] Williamson MR, Boyd CM, Williamson SL. Osteoporosis: diagnosis by plain chest film versus dual photon bone densitometry. *Skeletal Radiol.* 1990;19: 27-30.
- [49] Paterson CR, Mole PA. Bone density in osteogenesis imperfecta may well be normal. *Postgrad. Med. J.* 1994;70: 104-7.
- [50] Deodhar AA, Woolf AD. Bone density measurement in osteogenesis imperfecta may well be important. *Postgrad. Med. J.* 1994;70: 463-4.
- [51] Lund AM, Mølgaard C, Müller J, Skovby F. Bone mineral content and collagen defects in osteogenesis imperfecta. *Acta Paediatr.* 1999;88: 1083-8.
- [52] Zionts LE, Nash JP, Rude R, Ross T, Stott NS. Bone mineral density in children with mild osteogenesis imperfecta. *J. Bone Joint Surg. (Br)* 1995;77-B:143-7.
- [53] Moore MS, Minch CM, Kruse RW, Harcke HT, Jacobson L, Taylor A. The role of dual energy x-ray absorptiometry in aiding the diagnosis of pediatric osteogenesis imperfecta. *Am. J. Orthop.* 1998;27: 797-801.
- [54] Kutilek S, Bayer M. Quantitative ultrasonometry of the calcaneus in children with osteogenesis imperfecta. *J. Paediatr. Child Health* 2010;46: 592-4.
- [55] Chines A, Boniface A, McAlister W, Whyte M. Hypercalciuria in osteogenesis imperfecta: a follow-up study to assess renal effects. *Bone* 1995;16: 333-9.
- [56] Brenner RE, Vetter U, Bollen A, Mörike M, Eyre DR. Bone resorption assessed by immunoassay of urinary cross-linked collagen peptides in patients with osteogenesis imperfecta. *J. Bone Miner. Res.* 1994;9: 993-7.
- [57] Minisola S, Piccioni AL, Rosso R et al. Reduced serum levels of carboxy-terminal propeptide of human type I procollagen in a family with type I-A osteogenesis imperfecta. *Metabolism* 1994;43: 1261-5.
- [58] Prószyńska K, Wiczorek E, Olszaniecka M, Lorenc RS. Collagen peptides in osteogenesis imperfecta, idiopathic juvenile osteoporosis and Ehlers-Danlos syndrome. *Acta Paediatr.* 1996;85: 688-91.
- [59] Kadhim M, Bober MB, Holmes L et al. Vitamin D status in pediatric patients with osteogenesis imperfecta. *Pediatr. Therapeut.* 2011;1:103.
- [60] Marlowe A, Pepin MG, Byers PH. Testing for osteogenesis imperfecta in cases of suspected non-accidental injury. *J. Med. Genet.* 2002;39: 382-6.
- [61] Pandya NK, Baldwin K, Kamath AF, Wenger DR, Hosalkar HS. Unexplained fractures: child abuse or bone disease. *Clin. Orthop. Relat. Res.* 2011;469: 805-12.

- [62] Kocher MS, Dichtel L. Osteogenesis imperfecta misdiagnosed as child abuse. *J. Pediatr. Orthop. B* 2011;20: 440-3.
- [63] D'Eufemia P, Palombaro M, Lodato V et al. Child abuse and osteogenesis imperfecta: how can they still be misdiagnosed? A case report. *Clin. Cases Miner. Bon. Metab.* 2012;9:195-7.
- [64] Haile JT, Carroll VG, Steele RW. Leg pain in an infant. *Clin. Pediatr.* 2010;49:78-81.
- [65] Antoniazzi F, Bertoldo F, Mottes M et al. Growth hormone treatment in osteogenesis imperfecta with quantitative defect of type I collagen synthesis. *J. Pediatr.* 1996;129:432-9.
- [66] Aström E, Jorulf H, Söderhäll S. Intravenous pamidronate treatment of infants with severe osteogenesis imperfecta. *Arch. Dis. Child* 2007;92:332-8.
- [67] Glorieux FH. Experience with bisphosphonates in osteogenesis imperfecta. *Pediatrics* 2007;119:S163-5.
- [68] Horwitz EM, Prockop DJ, Gordon PL et al. Clinical responses to bone marrow transplantation in children with severe osteogenesis imperfecta. *Blood* 2001;97:1227-31.
- [69] Esposito P, Plotkin H. Surgical treatment of osteogenesis imperfecta: current concepts. *Curr. Opin. Pediatr.* 2008;20:52-7.
- [70] Yilmaz G, Hwang S, Oto M et al. Surgical treatment of scoliosis osteogenesis imperfecta with cement augmented pedicle screw instrumentation. *J. Spinal. Disord. Tech.* 2012;doi:10.1097/BSD.0bo13e3182624b76.
- [71] Hayes M, Parker G, Ell J, Sillence D. Basilar impression complicating osteogenesis imperfecta type IV: the clinical and neuroradiological findings in four cases. *J. Neurol. Neurosurg. Psychiatry* 1999;66:357-64.
- [72] Ibrahim AG, Crockard HA. Basilar impression and osteogenesis imperfecta: a 21 year retrospective review of outcomes in 20 patients. *J. Neurosurg. Spine* 2007;7:594-600.
- [73] Hultcrantz M, Sääf M. Stapes surgery in osteogenesis imperfecta: a clinical study of 16 patients. *Adv. Otorhinolaryngol.* 2007;65:222-5.
- [74] Denoyelle F, Daval M, Leboulanger N et al. Stapedectomy in children: causes and surgical results in 35 cases. *Arch. Otolaryngol. Head Neck. Surg.* 2010;136:1005-8.
- [75] Rottevel LJ, Beynon AJ, Mens LH et al. Cochlear implantation in 3 patients with osteogenesis imperfecta: imaging, surgery and programming issues. *Audiol. Neurotol.* 2008;13:73-85.
- [76] Wenstrup RJ, Willing MC, Starman BJ, Byers PH. Distinct biochemical phenotypes predict clinical severity in non lethal variants of osteogenesis imperfecta. *Am. J. Hum. Genet.* 1990;46:975-82.
- [77] Paterson CR, McAllion SJ, Shaw JW. Clinical and radiological features of osteogenesis imperfecta type IVA. *Acta Paediatr. Scand.* 1987;44:187-9.

Chapter 86

AUTOSOMAL DOMINANT DISORDERS ASSOCIATED TO BREAST CANCER

Nelly Margarita Macías-Gómez*

Laboratorio de Genética Humana; Departamento de Salud y Bienestar. Centro
Universitario del Sur. Ciudad Guzmán, Jalisco, México

ABSTRACT

The sudden rise of new biochemical and molecular techniques, have enabled a better understanding of the physiological and biochemical bases of the tumorigenesis, leaving clear that cancer is a “*genetic condition*”. Breast cancer is one of the most common cancers in women, affecting one in six women among 40-59 years in the world; so it has been widely studied. Unlike the majority of genetic diseases, in which the presence of a mutation in a particular gene is sufficient for delineation of a phenotype (monogenic); in breast cancer, the simple presence of a mutation in a particular gene is not enough to explain it. Approximately 90% of breast cancer cases occur sporadically and the majority of cases are caused by mutations in the *BRCA1* or *BRCA2* gene. However, in 5-10% of cases of breast cancer has been identify an autosomal dominant transmission, as well as mutations in specific genes such as *TP53*, *PTEN*, *CHECK2* *STK11* among others, that considerably increase the susceptibility to this condition. Autosomal dominant transmission of this disease has opened a new chapter in cancer medicine since the presence of any of these mutations in a patient, forces the doctor to carry out a deep investigation of the condition in order to establish a prognosis, as well as effective strategies for survival and family prevention. This chapter makes a brief revision of the autosomal dominant disorders Li-Fraumani syndrome, Cowden Disease and Peutz-Jeghers syndrome, which have a high susceptibility to development of breast cancer.

* Corresponding Author's Email: Nellymacias_2000@yahoo.com.mx.

BREAST CANCER

The sudden boom of biochemical and molecular techniques, have enabled us to a better understanding of the physiological and biochemical bases of the tumorigenesis, leaving clear that cancer is a genetic "condition". Breast cancer is a disease in which the different cells from breast gain the ability to grow, multiply and become immortal. Is the most common cancers in humans, so it has been widely studied; however, until now there are many questions about the mechanisms involved in their development where several factors as hormonal, reproductive, life style and inheritance play an important role. Even breast cancer mortality has declined in the last 10-15 years (2.3% per year) due to the innovation of diagnostic techniques and more effective treatments, still is a health problem in reproductive women because each year is reported an incidence of 1 million cases, representing 200,000 cases in the EU (27% of all cancers in female) and 320,000 cases in Europe (31% of all cancers in female) (Neville, 2001, Dumitrescu, 2005). In women from United States, breast cancer is the first cancer which represents 22-32% of all types of cancer, as well as being the second cause of death followed by lung cancer (Thull, 2004, Garber, 2005). According with Globocan in 2008 was reported 1'384, 155 new cases of breast cancer, which represent the 22.9% of all types of cancer and a mortality of 458,503 (13.7%) (<http://globocan.iarc.fr/>). Is estimated that a woman who lives 85 years, has one to nine chance of developing breast cancer, however, this risk is not homogeneous for the entire population, since while some women never develop breast cancer, the risk is increased for others. Epidemiological studies in different populations, have been the identification of well established factors that increase or decrease the risk for developing cancer of the breast, as well as other factors that is necessary to conduct more studies in order to identify their contribution (table 1) (Dumitrescu, 2005, Singletary, 2003).

Table 1. Risk factors to Breast Cancer

Biological and geographic	Age
	Gender
	Geographic localization
	Race/ Ethnicity
	Hormonal/ Reproductive
Life Style	Alcohol/ Folates
	Diet
	Obesity
	Physical activity
Genetics	Familiar History
	Syndrome associated to Breast Cancer development
Others	Endometriosis
	Tobacco
	Breast density
	Bening breast lesion
	Radiation exposure
	Bone density
	Intake of aspirin and INES

Among the genetics risk factors has been well known, that the affection of a family member with breast cancer is a strong risk factor, and this risk is estimated depending on the type of cancer, degree of relationship (first or second degree), age at which the family developed the cancer and the number of relatives affected in the family. In this regards, has been estimated that a women with a familiar history in first degree with breast cancer at 50 years or over, have a increase relative risk (RR) of 1.8 and who develop before at 50 years it's RR is 3.3. Similarly, when there is a history of two affected relatives in the first degree, the RR increases to 3.6 and, 3.9 when affected relatives are more than two (Dumitrescu, 2005, Glover, 2006, Antoniou, 2006). Although approximately 10-20% of breast cancer cases are attributed to hereditary factors, only 5-10% of patients has been identified a specific mutation in high penetrant genes that are transmitted in autosomal dominant manner, and the most widely identified are *Breast Cancer 1 (BRCA1)* and *Breast Cancer 2 (BRCA2)*. Recently, had been delineated a group of syndromes autosomal dominant that share the susceptibility to develop several types of cancer, including breast cancer: the Li-fraumeni syndrome which has mutations in the *TP53* gene (*Tumor protein 53*), Cowden syndrome with mutations in *PTEN* gene (*Phosphatase and tension homolog*) and Petz-Jeghers with mutation in *STK11* gene (*Serine / threonine kinase 11*) (Dumitrescu, 2005).

LI-FRAUMENI SYNDROME

Li-Fraumeni syndrome (*LFS*) (MIM-#151623), described for the first time in 1969, is a heterogeneous condition autosomal dominant, with high penetrance. *LFS* is characterized by the appearance of different types of cancer during the childhood of an individual, as well as several members of the family. The life time risk for cancer develops in patient with *LFS* is of 73% in males and nearly 100% in females, with the high risk of breast cancer accounting for the difference. Additionally, is important to consider the age of patient, due to the specific risk for males is 19% before to 15 years, 27% between 16-45 years and 54% in older than 45 years. The risk for females is 12 before 15 years, 82% between 16-45 years and 100% in older than 45 years (Malkin, 2011; Evans, 1997). Has been described that *LFS* family member diagnosed with cancer has a 15% to develop a second cancer, 4% develop a third cancer and 2% has a fourth cancer (Hisada, 1998; Mai, 2012). One of the main types of cancer who develop *LFS* patients is breast, which is of early presentation (< 40 years) and usually bilateral. *LFS* is responsible for approximately 1% of familial breast cancer syndrome. The median age of diagnosis of first malignancy is 25 and approximately 50% of *LFS*-associated malignancies occur by age 30 years. The median age of onset of breast cancer diagnosis in carriers is about 33 years and no occurred after age 50 years. Overall, woman with *LFS* has a breast cancer risk of 56% by age 45 and greater than 90% by age 60, with a majority of the cancers occurring under age 40. If the *LFS* is done, the NCCN guideline suggest the patient being monthly breast self-examinations at age 18, biannual clinical breast exams to begin at age 20-25, and annual imaging with mammography and /or MRI at the same ages, or 5-10 years prior to the earliest known breast cancer in the family kindred (Gage, 2012). Other cancer types presentation in patients cancer are, sarcoma of soft tissues, gastrointestinal (colon, gastric, pancreatic) cancer, lung cancer, osteosarcoma, hematopoietic cancer such as leukemia and lymphoma, adrenocortical carcinoma, brain tumors. Often is diagnosed in early infancy according to the

diagnostic criteria (table 2), and approximately 50% of cases are diagnosed at the age of 30. Families or individuals that do not have all the criteria necessary to the LFS diagnosis have been termed LFS-Like (LFS-L) (Mai, 2021; Neville, 2001, Thull, 2004, Garber, 2005). For more information about diagnose criteria review: www.nccn.org and www.iarc.fr.

Table 2. Li-Fraumani syndrome and Li-Fraumani-like syndrome criteria for genetic testing (Mai, 2012)

Classic Li-Fraumani syndrome (LFS)	Patient who developed sarcoma before age 45 and Have a family member in first grade with any type of cancer (< 45 years) and A first or second degree relative with any cancer before 45 years or sarcoma at any age
Li-Fraumani- like syndrome	Birch definition: • A proband with any chils cancer or sarcoma, brain tumor or adrenocortical carcinoma diagnosed before age 45 years and • A first or second degree relative with a typical LFS cancer at any age and • A first or second degree relative with any cancer before age 60 years Eeles definition: • Two first or second degree relatives with LFS-related malignancies at any age
Chompret criteria	A proband who has • A tumor belonging to the LFS tumor spectrum (soft tissue sarcoma, osteosarcoma, pre-menopausal breast cancer, brain tumor, adrenocortical carcinoma, leukemia, or bronchoalveolar lung cancer) before age 46 years and • At least one first or second degree relative with an LFS tumor (expect breast cancer if the proband has breast cancer) before age 56 years or with multiple tumors Or • A proband with multiple tumors (except multiple breast tumors), two of which belong to the LFS tumor spectrum and the first of which occurred before age 46 years Or • A proband who is diagnosed with adrenocortical carcinoma or chorois plexus tumor, irrespective of family history

TP53

In most of families with LFS have been identified germline mutations in *TP53* (*Tumor protein p53*) gene. *TP53* is located in 17p13.1, has 25,767 bases and encode for a protein of 393 amino acids (<http://www.genecard.org>). *TP53* is a tumor suppression protein, consider it as “DNA guardian” for its capacity to identify when cell has a DNA damage, stop the cell cycle in order to repair the DNA and, in case that the damage is major, drive the cell thru apoptosis (Walerych, 2012). *TP53* gene significance has been supported on the others things by the

frequent occurrence of breast cancer in LFS patients. In Only the 60-80% of patient with LFS classic has been detectable germline mutation on *TP53*, while patients with LFS-L so not shown detectable *TP53* mutations. The lack of concordance among no *TP53* mutations and the presence of LFS phenotype, can be explained for possible posttranscriptional *TP53* alterations, an complete deletion, the effects of modifier genes, or the possibility of a second locus until now no identified (Malkin, 2011). Additionally *TP53* gene has a higher mutation frequency, higher than any other tumor suppressor gene in human overall. On average, *TP53* is mutated in 31% of all tumors, and in 23% of breast cancer samples, where the principal mutated protooncogene is *PI3KCA* (Walerych, 2012). In some patients had shown germinal mutations on *CHEK2* gene, which encodes for a protein kinase that functions as regulator of the cell cycle, and responds to the damage to the DNA, activating *PT53* and *BRCA1*. In the North of Europe and the United Kingdom, the variant of the gene *CHEK2* 1100delC (deletion of a cytosine at position 1100 of the gene *CHEK*) increases twice the risk for the development of breast cancer in women, and more than ten times in males; In addition to being responsible for 1% of the cases of breast cancer in women and 9% in males. These frequencies are very clear in North America, as it has been the presence of this variant in 0.3% of healthy volunteers, and in 1% of patients with breast cancer; so it is necessary to determine the contribution of the allele *CHEK2* 1100delC, in the development of cancer of the breast (Neville, 2001, Thull, 2004, Garber, 2005).

2. COWDEN SYNDROME

The Cowden Disease or syndrome of multiple hamartomas (*CS - Cowden syndrome*) (MIM #158350), described by Lloyd and Dennis in 1963, is a rare inherited genodermatosis with an autosomal dominant inheritance pattern and strong predominance of female patients (6:1), which may be fortuitous. CS is associated with both malignant lesions as benign that affect cells of the three germ layers and tissues mostly affected are tissue breast, thyroid, uterus, brain, mucocutaneous tissue and Genitourinary tract. It is estimated that 1 in every 200,000- 250,000 individuals are affected with CS, although this frequency may be underestimated due the difficulty to make the diagnosis; with an age at diagnosis ranging from 13 to 65 years. In this regards, the International Cowden Syndrome Consortium (<http://www.nccn.org>) established the diagnosis criteria in 2000, where the major (such as breast carcinoma, follicular thyroid carcinoma, multiple gastrointestinal hamartomas, macrocephaly, among other) and minor criteria (autism spectrum disorder, colon cancer, esophageal glycogenic acanthosis, mental retardation among others) are fundamental at the diagnosis process (Nagy, 2004). The penetrance in CS is related to age, since at the age of 20 years 99% of affected patients present mucocutaneous lesions, which are the most constant and characteristic findings. The risk of breast cancer in patients with CS is 25-50%, represents less than 1% of the cases of breast cancer and occurs between 38 and 46 years of age. Additionally, women with CS have a higher risk (67%) for develop benign breast disease, which include apocrine metaplasia, fibroadenomas, microcysts, adenosis, and hamartoma-like lesions with densely hyalinized collagen (Gage, 2012; Shah, 2012, 2013). Others types of cancer present in patients with CS are the thyroid cancer, typically follicular and papillary occasionally with a risk of 10%, while for cancer endometrial are 5-10%. Due to the high risk for the development of breast cancer,

patients with the SC must be monitored from an early age, starting a clinical breast examination before age 25, annual mammogram from 30 years, or in case to have a familiar with CS diagnosis and breast cancer, is important to make the first mammogram 5 years before at the age of diagnosis of breast cancer in the family, as well as an annual physical review beginning at the age of 18 years monitoring lesions in skin and thyroid, inclusive is important to include a thyroid ultrasound. Other important characteristic reported in patients with CS are craniomegaly which is the most common extracutaneous manifestation (80% incidence), gastrointestinal polyps (approximately 60-85%) with predilection from the esophagus, stomach, and colorectal structures. The small bowel is rarely involved. Polyps are usually small and <5 mm in size; cutaneous fibromas (76%), thyroid abnormalities (62%) and multiple uterine leiomyoma (40%) (Table 3) (Fistarol, 2002; Nagy, 2004; Starink, 1986). According with the mucocutaneous lesions, these include multiple trichilemmomas, oral papillomatosis, facial papules, and acral keratoses. The trichilemmomas, is defined as a benign hamartomas of the outer sheath of hair follicles, are flesh-colored smooth papules, ranging from 1 -5 mm in size and are present predominantly of the face, head and neck, close to the hairline. Other important mucocutaneous lesions described in CS include the hemangiomas, scrotal and furrowed tongue, neuromas, xanthomas, vitiligo, acanthosis nigricans, perioral and acral lentigines, and frequently the speckled pigmentation of the penis (Shah, 2012, 2013).

PTEN (Phosphatase and Tension Homolog)

In approximately 85% of CS cases has been identified mutations in the *PTEN* (Phosphatase and tensin homolog) gene, which is a gene suppressor tumor located in 10q23.31 (MIM 601728) contains 108,818 bases, 9 exons and encodes for the protein Phosphatidylinositol 3,4,5-triphosphate 3-phosphatase, consists of 403 amino acids and is a member of the super family of genes protein-tyrosine phosphatase (PTP). *PTEN* gene encodes a ubiquitously protein with dual-specificity phosphatase is due to contain a tensin like domain as well as catalytic domain similar to that of the dual specificity protein tyrosine phosphatases. Unlike most of protein tyrosine phosphatases, this protein preferentially dephosphorylates phosphoinositide substrates (www.genecards.org). This activity interferes with the progression of on G1 cell cycle through the negative regulation of PI3-kinase/Akt or PKB signaling pathway. At the same time PTEN protein regulates the mitogen-activated protein kinase (MAPK). Deregulation of these two pathways by *PTEN* inactivation or loss, increased the cell survival and uncontrolled cellular proliferation, resulting in tumor development. The mechanism that gives rise to this syndrome is loss of heterozygosis (LOH) in *PTEN* in 20-60% of cases and is explained with Knudson (double hit hypothesis) hypothesis. In the case of sporadic tumors, both alleles are normal at the moment of the conception, however subsequently have a mutation postzygotic - first hit - in one allele; subsequently a new mutation - second hit - in the other allele causes the LOH, causing a loss in the control of cell growth. In the case of hereditary tumors, the heterozygosity is present from the moment of conception and only a postzygotic mutation is necessary for the LOH (Roman, 2012). Has been reported that germline mutations in *PTEN* gene, is the most common muted gene in several cancers and are associated a number of heritable cancer syndromes, whom collectively are referred to as *PTEN Hamartoma Tumor Syndrome* (PHTS) (Pezzolesi, 2007; Tan, 2012). The PHTS includes the CS, Bannayan-Riley-Ruvalcaba syndrome [BRRS, (MIM #153480)] and Lhermitte-Duclos syndrome (Romano,

2012). According to the strict diagnostic criteria for CS, 80% of the patients presented germinal mutations in the gene *PTEN*, from which approximately 8.8% of patient are carriers of a germinal mutation and the exons 5, 7, and 8 considering as hot-spots; and all kinds of mutations including deletions, insertions, splice site mutations and large deletions have been identified in these patients. Even though many efforts had been done in order to correlate the different types of cancer or phenotype with the found mutations has not been identified (So, 2012). In such patients, it is important to discard the *Bannayan-Riley-Ruvalcaba* syndrome (MIM-#153480), since certain germinal mutations in *PTEN* have been identified in 60-65% of patients with East syndrome, by what has been considered as an allelic to the CS condition (Neville, 2001, Thull, 2004, Garber, 2005).

3. PEUTZ-JEGHERS SYNDROME

Peutz-Jeghers Syndrome (PJS) (MIM #175200), is an autosomal dominant condition, which was recognized in 1949 by Jeghers et al. as a localization of intestinal polyposis and pigmentation of the skin and mucous membranes, hence the eponym Peutz- Jeghers syndrome. The PJS has been reported around the world, affecting equally males and females; its estimate incidence is 1:8300 to 1:200,000 live births, and 25% of cases appear to be non familial or sporadic. PJS is caused by germline mutations in the *STK11* (*Serine/threonine kinase*) tumor suppressor gene; and characterized by a predisposition (20.3 increased relative risk) to developing malignant tumors in different organs, including stomach, colon, pancreas, intestine thin, thyroid, breast, lung, and uterus. The incidence of cancer in these patients has been estimated at fifteen times higher than in the general population; and the pathognomonic features are polyps hamartomatous of the gastrointestinal tract (especially in small intestine), with a specific histology, which may result in symptoms of bleeding or obstruction, being the intussusception a major complication, especially in pediatric patients (Shah, 2012, 2013). Extra-intestinal sites of PJS polyps include kidney, ureter, gallbladder, bronchus and nasal passages. The breast cancer is other important malignance in patient with PJS, affecting at 32% to 54% of patients and being the ductal and occasionally the lobular cancer the more frequent. Recent studies reported that 8% of women with PJS developed breast cancer by age 40, and 32% by age 60 (Gage, 2012; Shah, 2012, 2013). In these cases, it has been observed that the *STK11* gene expression is decreased, which correlates with a histologically high-grade, a large tumor and the presence of metastases to lymph nodes; In addition to partnering with a higher rate of relapse and worse prognosis. More features of the PJS neoplasms are of the genital tract, which include the tumor of sex cord with annular tubules (*SCTAT - Sex cord tumor with annular tubules*), followed in tumor cells of sertoli, Mucinous epithelial tumor, serous tumour and mature teratoma of ovary (Neville, 2001, Thull, 2004). Other characteristic clinical feature is the mucocutaneous hyperpigmentation, which usually is present in childhood as dark macules, and are distributed in the lips and perioral region (94%), hands (74%), buccal mucosa (66%), and feet (62%). However, the mucotutaneous hyperpigmentation also can be present in uncommon region as around the eyes, nostrils, and perianal area. Usually the lesions fade during puberty, with the exception of those on the buccal mucosa (Mishra, 2012). The cumulative risk of any cancer is 67-85% by age 70 and the cumulative risk for CRC is 3% (40 years), 5% (50 years), 15% (60 years), and 39% (70 years). The risk to age 70 for cancers of

the pancreas is 11%, utero/ovary/cervix 18%, breast 45% and lung of 17% (Shah, 2012, 2013; Mishra, 2012).

STK11 Gene

The *STK11* (Serine/Threonine Kinase) gene is the responsible of approximately 70-80% of the PJS; is located in 19.13.3, have 39,029 bases distributes in 10 exons and encode a 433 amino acid protein, STK11 (Liu, 2011). STK11 protein is part of the serine/threonine kinase family and is a well known tumor suppressor, which regulates the activity of AMP-activated protein kinase (AMPK) family members, playing an important role in several processes as cell metabolism, cell polarity, apoptosis and DNA damage response, regulating principally the expression of *p53* gene, as well as its targets genes, such as *p21*. Has been estimated a birth prevalence of mutation in STK11 at 1:25,000 to 1:280,000 (Alexander, 2011; Hemminki, 1998; Liu, 2011).

REFERENCES

- Alexander A, Walker CL. The role of LKB1 and AMPK in cellular responses to stress and damage. *FEBS Lett.* 2011;585(7):952-7.
- Antoniou AC, Easton DF. Models of genetic susceptibility to breast cancer. *Oncogene*, 2006b; 25, 5898-5905.
- Dumitrescu RG, cotar it I. Understanding breast cancer risk-where do we stand in 2005? *Journal of Cellular Molecular Medicine*, 2005; 9 (1): 208-221.
- Evans SC, Lozano G. *The Li-Fraumeni syndrome: an inherited susceptibility to cancer.*
- Fistarol SK, Anliker MD, Itin PH. Cowden disease or multiple hamartoma syndrome - cutaneous clue to internal malignancy. *European Journal of Dermatology*; 2002; 12(5): 411-421.
- Gage M, Wattendorf D, Henry LR. Translational advances regarding hereditary breast cancer syndromes. *J Surg Oncol.* 2012;105(5):444-451.
- Garber JE, Offit K. Hereditary cancer predisposition syndromes. *Journal of Clinical Oncology*, 2005; 23 (2): 276-292.
- Glover JN. Insights into the molecular basis of human hereditary breast cancer from studies of the BRCA1 BRCT domain. *Familial cancer*, 2006;5(1):89-93.
- Hemminki A, Markie D, Tomlinson I, Avizienyte E, et al. A serine/threonine kinase gene defective in Peutz-Jeghers syndrome. *Nature.* 1998;391(6663):184-7.
- Hisada M, Garber JE, Fung CY, Fraumeni JF Jr, Li FP. Multiple primary cancers in families with Li-Fraumeni syndrome. *J Natl Cancer Inst.* 1998;90(8):606-611.
- Liu L, Du X, Nie J. A novel de novo mutation in LKB1 gene in a Chinese Peutz Jeghers syndrome patient significantly diminished p53 activity. *Clin Res Hepatol Gastroenterol.* 2011;35(3):221-6.
- Mai PL, Malkin D, Garber JE, Schiffman JD, Weitzel JN, et al. Li-Fraumeni syndrome: report of a clinical research workshop and creation of a research consortium. *Cancer Genet.* 2012;205(10):479-487.

- Malkin D. Li-fraumeni syndrome. *Genes Cancer*. 2011;2(4):475-484. *Mol Med Today*. 1997;3(9):390-395.
- Nagy R, Sweet K, Eng C. Highly penetrant hereditary cancer syndromes. *Oncogene*. 2004;23(38):6445-6470.
- Neville PJ, Morland SJ, Vaziri SAJ, Casey G. The genetics of breast and ovarian cancer. In: Cowell JK (ed). *Molecular genetics of cancer 2nd ed*. 2001:53-80.
- Pezzolesi MG, Zbuk KM, Waite KA, Eng C. Comparative genomic and functional analyses reveal a novel cis-acting PTEN regulatory element as a highly conserved functional E-box motif deleted in Cowden syndrome. *Human Molecular Genetic*, 2007; 16: 1058-1071.
- Rebecca Nagy, Kevin Sweet and Charis Eng. Highly penetrant hereditary cancer syndromes. *Oncogene*; 2004. 23, 6445-6470.
- Romano C and Schepis C. PTEN Gene: A model for Genetics Disease in Dermatology. *The Scientific World Journal*; 2012. 2012: 252457
- Shah KR, Boland CR, Patel M, Thrash B, Menter A. Cutaneous manifestations of gastrointestinal disease: part I. *J Am Acad Dermatol*. 2013;68(2):189.
- Singletary SE. Rating the risk factors for breast cancer. *Annals of Surgery*, 2003; 237 (4): 474-482.
- Starink TM, van der Veen JP, Arwert F, de Waal LP, et al. The Cowden syndrome: a clinical and genetic study in 21 patients. *Clinical Genetic*, 1986; 29 (3): 222-33.
- Tan MH, Mester JL, Ngeoe J, et al. Lifetime Cancer Risks in Individuals with Germline PTEN Mutations. *Clin Cancer Res*, 2012; 18: 400-407.
- Thull DL, Vogel VG. Recognition and management of hereditary breast cancer syndromes. *The Oncologist*; 2004; 9 (1), 13-24.
- Walerych D, Napoli M, Collavin L, Del Sal G. The rebel angel: mutant p53 as the driving oncogene in breast cancer. *Carcinogenesis*. 2012;33(11):2007-2017.

Chapter 87

FRAGILE X SYNDROME

M. Milà*

Servicio de Bioquímica i Genética Molecular.
Hospital Clínic, Barcelona, Spain
CIBER de Enfermedades Raras (CIBERER) Barcelona, Spain
IDIBAPS (Institut de Investigacions Biomèdiques August Pi I Sunyer)
Barcelona, Spain

ABSTRACT

Fragile X syndrome (FXS) is the most common cause of familial intellectual disability and the most commonly known single gene causing autism spectrum disorders. The incidence varies according to the populations; it is well-accepted that 1/4,000 males and 1/6,000 females are affected, and 1/250 females and 1/800 males are carriers. The main clinical manifestations are intellectual disability, dimorphic traits and behavior disturbances. The syndrome is inherited as a dominant X-linked trait with reduced penetrance (80% for males and 30% for women). This syndrome is due to a functional loss of the *FMR1* gene product, Fragile X Mental retardation Protein (FMRP), and, in most cases it is caused by a CGG repeat expansion in the *FMR1* promoter. The repeat is from 6 to 55 CGGs long in the normal population while in patients with FXS the repeat number is over 200 CGGs (full mutation FM). This number of CGGs generally leads to methylation of the repeat and the promoter region, which is accompanied by silencing of the *FMR1* gene. The absence of the FMR1 protein, FMRP, is the cause of the intellectual disability in these patients. Individuals with 55 to 200 CGGs carry a premutation (PM). All affected children have carrier mothers (full mutation [FM] or PM) with a 50% of chance of having another affected child in future pregnancies. Female PM carriers are at risk of developing primary ovarian insufficiency (FXPOI). Elderly PM carriers (males and females) may develop a progressive neurodegenerative disorder called fragile X-associated tremor/ataxia syndrome (FXTAS). The *FMR1* gene is responsible for different disorders depending on the length of the CGG tract and the molecular mechanism. The Fragile X syndrome is due to a loss of function, and FXPOI and FXTAS are due to a toxic gain of function of the

* Corresponding Author's Email: Montserrat Milà; mmila@clinic.ub.es.

mRNA or repeat-associated-non-AUG translation. At present, there is no treatment, although several studies are ongoing in both human and animal models.

The present and the following chapters review the current status of the wide spectrum of different pathologies related to the *FMR1* gene.

Keywords: fragile X syndrome, *FMR1* gene, CGG expansion, intellectual disabilities, autism spectrum disorder

INTRODUCTION

Intellectual disability (ID) affects 1-3% of the general population; approximately half of the ID cases have a familial origin. On the other hand the autism and autism spectrum disorders (ASD) are estimate to affect 1% of general population. The exact causes of ID and ASD are unknown, although it is thought that several complex genetic and environmental factors are involved. At present in around 50% of the cases with ID or ASD, the genetic defect remains unknown being difficult to give genetic counselling.

The first monogenic cause for these two pathologies is the *FMR1* gene which constitute the most common cause of inherited intellectual disability, developmental delay and the most commonly known single gene causing autism spectrum disorders.

This chapter it is focus on the Fragile X syndrome (FXS) reviewing the different aspects clinical, molecular basis, animal models, genetic counseling, and neonatal screening.

FRAGILE X SYNDROME

Fragile X syndrome (FXS) (FXS #MIM300624; ORPHA 908) is the most common cause of inherited intellectual disability (ID), developmental delay and is the most commonly known single gene causing autism spectrum disorders (ASD). The *FMR1* gene (Fragile X mental retardation type 1 gene) is inherited as a X-linked dominant trait with a reduced penetrance of 80% in males and 30-50% in females.

In 1943 Martin and Bell made the first description of the syndrome in males with ID and affecting more than one male in a family. The most remarkable clinical signs were moderate to severe ID, having a long face, prognatism, large ears, machroorchidism, and alterations in connective tissue as well as characteristic behavior.

The name “fragile X syndrome” was attributed to Lubs, in 1969 [1] who observed fragility at the edge of the log arm of the X chromosome. However, this fragility was only observed in a reduced number of metaphases and it was more evident in cultures in which the media was poor in folic acid.

In 1991 three groups working independently identified the gene responsible and the molecular defect [2-5]. Fragile X syndrome was the first human disease to be found to be caused by a dynamic mutation; the main mutation is a CGG expansion in the untranslated region of the first exon of the *FMR1* gene (Fragile X mental retardation type 1).

The incidence is variable depending on the population, although in general an incidence of 1/4,000 males and 1/6,000 females is well accepted, thereby considering FXS as a rare disease [6].

Clinical manifestations also vary depending on the age and sex [7-12] (Figure 1).

Prepubertal males show moderate ID with a delay in the milestones (seated, speech and ambulation) and behavioral disturbances: deficit in attention, hyperactivity, hand biting, autistic behavior, refusal to touch, shyness and gaze avoidance, repetitive movements especially of the hands and echolalia, all of which become more accentuated with age. Although the ID is usually moderate, it is not uncommon to find severe forms with an IQ between 35-55 and psychiatric disorders. In post pubertal males the physical traits are more pronounced, ID is more severe and macroorchidism is present in 80% of the cases.



Figure 1. Phenotypic appearance of two affected brothers with FXS.

The physical traits of FXS are initially quite subtle, and the first warning sign is often a delay in the emergence of language, which can be accompanied by a slight engine, hyperactive behavior, attention deficit, autistic-like behaviors such as flapping and / or hand biting, and poor eye contact. On physical examination, an elongated face with prominent chin and large and protruding ears is observed. Other relevant clinical data include joint hypermobility, especially of small joints, fine and velvety skin, being common to see many wrinkles on the palms of the hands. They usually have a high palate. Regarding the heart, the most common finding is a prolapse of the mitral valve that manifests as a mild to moderate heart murmur which is usually asymptomatic. These and other anomalies are considered mainly due to an alteration of the connective tissue.

Individuals with FXS often have no serious medical problems. Large studies have reported wide discrepancies in the prevalence of some associated medical problems. This fact is probably due to the great variability in these patients. The most prevalent medical problems are: the mitral valve prolapse, recurrent otitis media, seizures, motor tics, strabismus, sleep problems and obstructive sleep apnea. It is known that during childhood otitis (recurrent otitis media), and subsequently, sinusitis are frequent. Convergent strabismus is also often present. Many children have flat feet, but this finding tends to improve with age. Approximately 30% of patients have some degree of gastroesophageal reflux during the first year, the treatment of which depends on the severity of the reflux. In the neurological area about 15% of patients with

FXS epilepsy have some abnormal EEG findings. In a subgroup of patients with FXS the syndrome is similar to the Prader-Willi syndrome with obesity and hyperphagia phenotype.

The speech of these subjects is often repetitive; they behave shyly and look away when spoken to. In addition, they show tactile defensiveness, gaze avoidance, poor concentration, impulsivity and they sometimes present aggressive behavior. A behavioral phenotype of autism is common in males. It is usual to present stereotyped movements with both hands (shaking), self-harming biting that causes callused areas on the hands and they also show defensiveness or over-reaction to certain external sensory stimuli. Poor social skills and socio-emotional reciprocity can be added to this sometimes progressive deterioration and this can lead to a diagnosis of an autism spectrum disorder.

In general girls show attenuated clinical manifestations; mild ID and phenotypic traits are missing or minimal (big ears, joint hypermobility, elongated facies), although some cases presenting similar clinical manifestations to those of males have been reported. From 30-50% of women have ID and most have learning disabilities, behavioral disorders and communicate with shyness, presenting social anxiety, attention problems and educational delay, especially in mathematics.

The clinical diagnosis may be very difficult in girls and women due to the variable emotional and behavioral characteristics and the subtle physical traits [7-11].

Approximately 30–50% of FXS patients meet full DSM-IV-TR criteria for autism, with 60–74% fulfilling criteria for an autism spectrum disorder. Over 90% of individuals with FXS display some form of atypical behavior characteristics of autism including social interaction (e.g., avoidance of eye contact, social withdrawal, social anxiety) and repetitive and stereotyped behaviors [12-13].

The lack of function of the *FMR1* gene is the cause of FXS. This gene is located in the long arm of the X chromosome at Xq27.3, expands 17 exons and 40 kb of genomic DNA. It transcribes an mRNA of 3.9 kb. The translated protein is called *Fragile X Mental retardation Protein* (FMRP), which is necessary from early stages of development and throughout life [14].

In most cases the molecular basis of the syndrome is a dynamic mutation: an expansion and hypermethylation of an unstable CGG trinucleotide repeat located in the first exon of the *FMR1* gene [2-4]. In the general population the number of CGG repeats is polymorphic, presenting from 6 to 54 CGGs, and the adjacent CpG island, which acts as a promoter is non-methylated. This CpG island is located in the 5' untranslated region of the gene and acts as a switch depending on its methylation status. When the gene is active, *FMR1* is transcribed and translated. Alleles harboring 6 to 54 CGGs are considered normal and remain stable upon transmission.

A second class of alleles that overlaps with the upper normal range contains 45 to 54 CGGs. This range, known as the gray zone, can be stable or slightly unstable, being transmitted to subsequent generations with the possibility of expanding to a premutated allele. The clinical implications of this class of alleles remain unclear [15-16]. The next class is premutated alleles with a range of 55 to about 200 CGG repeats. In this situation, the *FMR1* gene is also transcribed and translated because the CpG island is non-methylated. Therefore, premutated carriers have normal or slightly reduced synthesis of FMRP and increased levels of mRNA (2-8 fold more than normal alleles) and they are asymptomatic for FXS. However, these carriers have a risk of having affected offspring since the number of CGG is unstable and tends to increase the CGG number in each cellular division.

People carrying permuted alleles are at risk of developing some disorders characteristic of this status such as: premature ovarian insufficiency (FXPOI), Fragile X tremor ataxia syndrome (FXTAS) and emotional disturbances, among others (see Chapters 2 and 3).

Finally, when the CGG number is over 200 repeats, these alleles are in the FM range. The CpG island and the repeats themselves are methylated and this methylation switches off the gene, blocking transcription and no protein is translated. These individuals are always affected with FXS if they are males, with only 30-50% of females being affected (Figure 2). Moreover, around 15-25% of patients show a pattern, referring to a mixture of PM and FM, which is caused by the somatic instability of the FM in early embryogenesis that can lead to retraction of the expanded repeat. Another kind of mosaicism can be observed in large expansions with incomplete methylation. The mosaicisms allow the expression of some FMRP and, in some cases, have been associated with milder ID in males.

Full Scale IQ scores were inversely correlated with the percentage of methylation and positively correlated with higher FMRP expression. These latter results point toward a positive impact on cognition for FM mosaics, with lower methylation compared to fully methylated individuals. Recent studies have suggested that low expression of FMRP may be sufficient to have a positive impact on cognitive function in these individuals. [17]. Individuals carrying the FM exhibit the classical fragile X phenotype that includes the broad spectrum of clinical manifestations described previously in this chapter and never present premutation-associated disorders such as FXPOI and FXTAS.

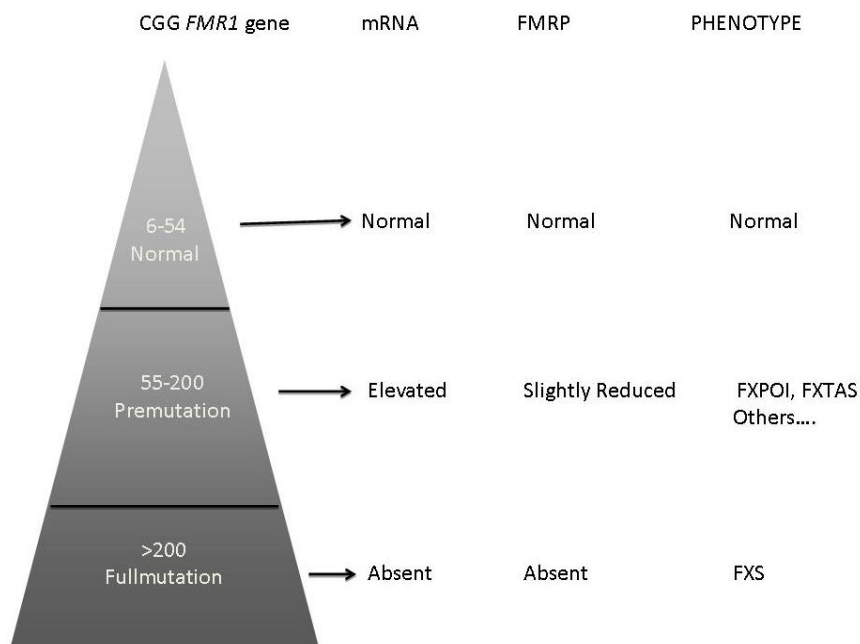


Figure 2. Summary of molecular aspects of the *FMR1* gene.

It is well known that some AGG are interspersed within the CGG expansion in the *FMR1* gene. The biological function of these interruptions seems to stabilize the gene during transmission and decreases the risk of DNA polymerase slippage during DNA replication. Previous studies have reported important data such as: AGG interruptions influence the stability

of the CGG repeats within the *FMR1* gene during parental transmission, the presence or absence of AGG interruptions is not correlated with the transcriptional or translational activity of the gene, and AGG interruption patterns can vary greatly between populations, but are largely inherited without change [18]

In most cases the FXS is due to a CGG repeat expansion in the *FMR1* promoter, but other *FMR1* mutations leading to a loss of function of the gene may also cause FXS or an FXS-like phenotype. Since standard molecular testing does not include sequencing of the *FMR1* coding region, the prevalence of point mutations causing FXS is not well known, although it seems that missense mutations in the *FMR1* gene might account for a considerable proportion of cases in males with FXS-related symptoms, such as those linked to ID and developmental delay. At present, the estimated frequency of point mutations is 1 to 2% of the cases of FXS [19].

The diagnosis of FXS is achieved by molecular analysis that determines the precise number of CGG in the *FMR1* gene [20]. Cytogenetic studies are currently not accepted as a diagnostic test. The quickest diagnostic method is polymerase chain reaction (PCR) using fluorescent labeled primers and subsequent analysis of the product in an image analyzer to determine the exact number of CGG repeats using small quantities of DNA. In males, the absence of PCR product indicates the presence of a pathological expanded allele.

The limitation of this technique is that it does not detect expansions of over 100 repeats and therefore does not detect large PMs or FMs and provides no information about methylation. Another limitation of PCR is that it does not distinguish between a homozygous female (both chromosomes harboring the same CGG number) and a woman with an allele in the normal range and another allele in the large PM range. These problems may be solved by conducting another PCR called Triplet-PCR (TP-PCR) which is able to identify large alleles and to distinguish between a homozygous female and a female carrying pre or full mutated alleles. Another specific PCR is necessary to detect methylation status [21-23].

These techniques are used for prenatal and postnatal diagnosis. DNA obtained from any tissue including saliva, blood, amniotic fluid, or chorionic villi can be used, thus, the choice of technique and tissue depends on the diagnostic strategy based on the clinical and family characteristics.

When an individual harboring an expansion is identified, a cascade family study is required to detect other relatives carrying pre or full mutations. In these cases precise determination of allele size and the presence of the AGG interruption are important to determine the risks of expansion in the carriers. Moreover, the status of methylation or the presence of mosaicism is relevant for genotype-phenotype correlations [24].

Finally, a test based on immunohistochemical methodology using a monoclonal antibody against the FMRP is used to detect the presence of this protein. This study may be useful in large-scale screening of the male population since males with a FM have no or a very low expression of FMRP. This test can also be useful in doubtful cases with a high PM phenotype suggestive of FXS. However, it is not useful in females, since the presence of the second normal X chromosome could modify the expression. [25-26]

In 1991, when the *FMR1* gene was cloned, Southern-blot analysis was the gold standard technique, but it is a laborious technique that requires several days of work, large amounts of DNA and gives imprecise estimates of the number of repetitions.

At present, the use of diagnostic commercial kits provided by several companies simplifies the work and ensures a safe diagnosis replacing the old techniques (Figure 3).

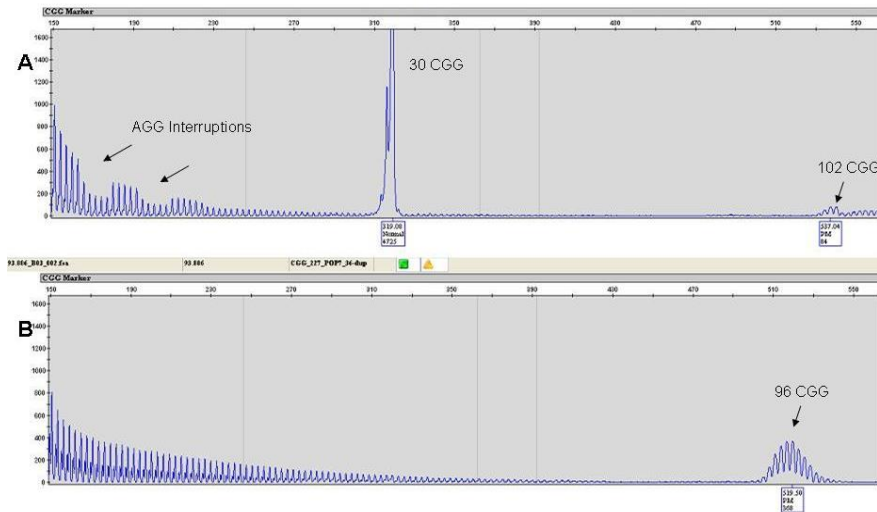


Figure 3. Electropherogram obtained using Amplidex Fragile X kit Asuragen. A) product from a female carrying one allele 30 CCG in normal range and the other allele with 102 CCG in permutation range, the arrows show the position of AGG interruptions and the allele size. B) Electropherogram corresponding to a PM male with 96 CCG and no AGG interruptions, were detected.

FMRP (Fragile X Mental Retardation Protein)

Cognitive impairment is caused by the absence of the FMRP in neurons. FMRP expression is widespread in neurons and spermatogonia. It is ubiquitously expressed from the early stages of development throughout postnatal life. The localization of FMRP is largely cytoplasmatic, being associated with the polyribosomes attached to the endoplasmatic reticulum and with free ribosomes at the bases of dendrites and within dendritic spines [27] FMRP plays a fundamental role in the synapses and normal development of dendrites. In healthy neurons, FMRP modulates the local translation of numerous synaptic proteins; synthesis of these proteins is required for the maintenance and regulation of long-lasting changes in synaptic strength. FMRP exerts profound effects on synaptic plasticity [28].

The *FMR1* has 17 exons with alternative splicing in exons 12, 14, 15 and 17 resulting in the expression of multiple isoforms of FMRP. The distribution of these isoforms is different according to the different brain regions except for the hippocampus and the olfactory bulb. FMRP binds to approximately 4% of all RNAs, regulating protein synthesis; and the lack of FMRP implies an excess basal translation. Approximately one third of all RNAs encoding pre- and postsynaptic proteins are targets of FMRP. This role as a transcription factor may explain the phenotypic complexity of the syndrome and the variable expression.

Animal Models for FXS

Although there are several animal models for FXS, fly and mouse models do not express the protein because the gene has been silenced, but not because of an expansion and hypermethylation.

Nevertheless, these models resemble those of affected patients, showing deficits in spatial learning, defect of prepulse inhibition of acoustic startle response, increased locomotor activity and are more susceptible to having epileptic seizures. These animal models allow the possibility of studying synaptic plasticity in many brain areas and the possibility of performing drugs trials [30-32].

For many years it was thought that PM carriers showed no clinical manifestations, only the risk of transmitting the disease to their offspring. In 1999 the presence of a PM became associated with in FXPOI women, and in 2000 a risk of late onset neurodegenerative disease, Fragile X Tremor Ataxia Syndrome (FXTAS), was associated with male and female PM carriers. Lastly, in the years 2007-2009 a range of phenotypes associated with the PM were described. It is now accepted that FXPOI (MIM#311360) and FXTAS (MIM#300623) are different from FXS and are only associated with the *FMR1* gene PM [33-37] (see Chapters 2-5).

All these associated disorders will be discussed in the next chapters. Nevertheless, it is important to highlight the importance of the PM carriers: It is estimated that more than 1 million people in the USA are *FMR1* carriers, with 20% of these subjects developing FXPOI and 40% of males and 16% of women developing FXTAS. Furthermore, 20% may present cognitive problems such as ASD (8%) and ADHD (30%).

Genetic Counseling

Although genetic counseling is discussed in detail in chapter 6, the following are the main features of this syndrome.

FXS is a genetic disorder that is inherited as a dominant X-linked trait with reduced penetrance (80% for males and 30% for women). In the general population we can find either men or women with alleles in the normal premutated or full mutated range.

Importantly, there are no sporadic cases so that whenever there is an affected child, the mother is an obligated carrier with a 50% of chance of having another affected child in future pregnancies. Members of the family in which the expansion is discarded are not at risk of presenting FXS.

Male carriers of either a PM or a FM always transmit the PM to their daughters, and thus, all carry the PM. The size of the PM is not exactly the same as the parent, and may slightly increase or decrease. Moreover, male carriers never transmit the FM. This is because only a PM and never a FM is found in the sperm. Males carrying a FM expansion transmit this to their daughters while always decreasing the CGG number to the PM range.

In conclusion, since sons inherit the Y chromosome and girls inherit the PM, male offspring are never affected by FXS.

Female carriers transmit the expanded allele to 50% of their offspring, both sons and daughters. Unlike male carriers, women tend to increase the size of the expansions in the passage from one generation to the next, so that a FM is only maternally inherited. The risk of expansion from the PM to the FM in women depends on the size of the PM and the AGG interruptions. Thus, uninterrupted and larger alleles are more likely to become a FM [16]. Indeed, whenever an allele with 100 CGG repeats is transmitted an expansion to full mutation is generated in the next generation. Women with a FM will transmit this allele mutated to 50%

of their offspring (sons and daughters), so that their offspring have a 50% risk of being affected with FXS.

It is important to highlight that despite the incomplete penetrance of the full mutated alleles in females only 50% show clinical manifestations. This fact is very important for genetic counseling especially in cases of prenatal diagnosis, when a female fetus carrying the FM is diagnosed. Nonetheless, we cannot know their phenotype.

A female carrier has different reproductive options: no offspring (adoption), ovidonation, prenatal diagnosis (amniotic fluid, corion villi or cord samples), preimplantation diagnosis or preconceptional diagnosis. (all these options will be discussed in chapter 6).

Genetic counseling in carriers of the PM should include the risk of developing pathologies associated with the PM: PM women have a 20% risk of developing a FXPOI and developing menopause before the age of 40.

PM males and females have a risk of developing FXTAS at over 50 years of age, the penetrance is 40% for males and 16% for females. [reviewed in Chapter 3, 24, 38]

At present it is advisable to request a *FMR1* test in the following cases:

- 1) Boys and girls with ID and autism, since the phenotypic characteristics are subtle in infancy it is recommended to rule out FXS in all cases of ID in whom the etiology is unclear, including a wide range of mild to profound ID, as well as developmental delays, autism, hyperactivity and other behavioral problems.
- 2) Women with infertility and / or ovarian failure before the age of 40 years, especially in cases with high levels of follicle-stimulating hormone (FSH) and if no other cause is confirmed such as ovarian cancer radiation treatment or thyroiditis.
- 3) Men and women with tremor and ataxia. Genetic testing of the *FMR1* gene is recommended in cases of cerebral ataxia with parkinsonism and intention tremor of unknown cause and cognitive decline in a person of 50 years of age.
- 4) Relatives of a diagnosed individual harboring an expansion in the *FMR1* gene. It is also necessary to perform a diagnostic "cascade", ie family studies arising from a former affection, including prenatal diagnosis in cases in which the pregnant woman is a PM or FM carrier.
- 5) Men and women who are ovum or sperm donors, due the high incidence of the PM in the general population.

Newborn Screening (NBS)

The incidence of this syndrome and its significance require its inclusion in screening programs, however, FXS is currently not included in NBS panels in any country due to its controversial particularities.

First of all, differences are observed in the diagnosis of males and females. For many years no test has been available for screening of females because of technical limitations. Secondly, there is no curative treatment for this syndrome, and finally, the screening could lead to a presymptomatic diagnosis of a neurodegenerative disease such FXTAS.

At present, the development of efficient methodologies makes FXS screening feasible for males and females. Indeed, it is possible to perform large population studies spanning the entire spectrum of *FMR1* mutation with a simple PCR.

Although no curative treatment is available, arguments in favor of NBS include the benefits of early detection that allows early intervention with good results and the possibility of cascade screening in these families and providing genetic counseling to all members.

Finally, a large number of clinical trials have been carried out regarding pharmacological interventions and new-targeted treatments will continue to be developed.

The main controversial aspect is that NBS allow the identification of PM carriers. An ever expanding number of medical disorders, although with a reduced penetrance, can occur in these individuals: ASD, ADHD, FXPOI, FXTAS, seizures, hypertension, migraines, sleep apnea, immune-mediated problems, thereby involving presymptomatic diagnosis of some diseases in very early stages of life.

This raises ethical problems which are difficult to solve. Ideally NBS for FXS should be included in NBS programs, however some ethical issues continue to make acceptance by ethical committees difficult [39].

Treatment

Although at present there is no treatment for all of these pathologies, Chapter 7 provides an update of the clinical trials on therapies for all these *FMR1* gene-related disorders and their current status.

The description of new specific treatment for experimental animal models of FXS involving therapeutic targets has allowed the discovery of drugs that are being tested and validated for treatment in humans.

The tests are intended to determine whether any of the compounds can attenuate or even normalize the clinical symptoms in patients with FXS. Until these drugs are authorized symptomatic treatments are able to control part of the disease and should be individually studied in each case to assess their effectiveness in controlling the symptoms present in each patient [40].

Table 1. Summary of the different status of the *FMR1* gene, risk of transmission and risk for the associated pathologies

CGG number	Gender	Status	Risk affected offspring	Risk develop FXPOI, FXTAS
6-45	male	Normal	No	No
	female	Normal	No	No
46-54	male	Normal	No	low
	female	Normal	No	low
55-200	male	Carrier	All daughters carrying premutation	40% FXTAS
	female	Carrier	50% FXS Depending of CGG and AGG number	20% FXPOI 40% FXTAS
>200	male	Affected FXS	All daughters carrying premutation	No
	female	30-50% Affected FXS	50% FXS	No

CONCLUSION

Fragile X syndrome is the most common cause of inherited ID, developmental delay and is a monogenic cause of autism spectrum disorders. The *FMR1* gene is responsible for different disorders: FXS (loss of function), FXPOI and FXTAS (toxic gain of function or repeat-associated-non-AUG). Although no treatment is currently available, early detection facilitates early intervention with good results and the possibility of performing cascade screening in these families and providing genetic counseling to all members.

REFERENCES

- [1] Lubs HA. A marker X chromosome. *Am J Hum Genet.* (1969) May;21(3):231-44.
- [2] Oberlé I, Rousseau F, Heitz D, Kretz C, Devys D, Hanauer A, Boué J, Bertheas MF, Mandel JL. Instability of a 550-base pair DNA segment and abnormal methylation in fragile X syndrome. *Science.* 1991 May 24;252(5009):1097-102.
- [3] Verkerk AJ, Pieretti M, Sutcliffe JS, Fu YH, Kuhl DP, Pizzuti A, et al. Identification of a gene (FMR-1) containing a CGG repeat coincident with a breakpoint cluster region exhibiting length variation in fragile X syndrome (1991). *Cell* 65:905-14.
- [4] Yu S, Pritchard M, Kremer E, Lynch M, Nancarrow J, Baker E et al. (1991). Fragile X genotype characterized by an unstable region of DNA. *Science* 252: 1179-81
- [5] Garber KB, Visootsak J, Warren ST. Fragile X Syndrome. *Eur J Hum Genet* 2008; 16:666-672.
- [6] Hagerman RJ. The physical and behavioural phenotype, in: RJ Hagerman, P Hagerman (Eds.), *Fragile-X Syndrome: Diagnosis, Treatment and Research*, *The Johns Hopkins University Press, Baltimore*, 2002, pp. 3–109
- [7] Farzin F, Perry H, Hessel D, Loesch D, Cohen J, Bacalman S, et al. Autism spectrum disorders and attention-deficit/hyperactivity disorder in boys with the fragile X premutation. *J DevBehavPediatr* 2006; 27:S137–144.
- [8] Kidd SA1, Lachiewicz A2, Barbouth D3, Blitz RK4, Delahunty C5, McBrien D6, Visootsak J7, Berry-Kravis E8. *Fragile X Syndrome: A Review of Associated Medical Problems Pediatrics.* 2014 Oct 6.pii: peds.2013-4301.
- [9] Loesch D1, Hagerman R. Unstable mutations in the FMR1 gene and the phenotypes. *Adv Exp Med Biol.* 2012;769:78-114.
- [10] Turkstra LS, Abbeduto L, Meulenb *Am J Intellect DevDisabil.* Social cognition in adolescent girls with fragile X syndrome. 2014 Jul;119(4):319-39.
- [11] Schneider A, Ligsay A, Hagerman RJ. Fragile X syndrome: an aging perspective. *Dev Disabil Res Rev.* 2013 Aug;18(1):68-74
- [12] McCary LM, Roberts JE. Early identification of autism in fragile X syndrome: a review. *J Intellect Disabil Res.* 2013 Sep;57(9):803-14. Review
- [13] Bailey DB, Raspa M, Holiday D, Bishop E, Olmsted M. Functional skills of individuals with fragile x syndrome: a lifespan cross-sectional analysis. *Am J Intellect Dev Disabil.* 2009 Jul;114(4):289-303

- [14] Pieretti M, Zhang FP, Fu YH, Warren ST, Oostra BA, Caskey CT, Nelson DL. Absence of expression of the FMR-1 gene in fragile X syndrome. *Cell*. 1991 Aug 23;66(4):817-22.
- [15] Madrigal I, Xunclà M, Tejada MI, Martínez F, Fernández-Carvajal I, Pérez-Jurado LA et al. (2011). Intermediate FMR1 alleles and cognitive and/or behavioural phenotypes. *Eur J Hum Genet* 19: 921-3.
- [16] Nolin SL, Sah S, Glicksman A, Sherman SL, Allen E, Berry-Kravis E, Tassone F, Yrigollen C, Cronister A, Jodah M, Ersalesi N, Dobkin C, Brown WT, Shroff R, Latham GJ, Hadd AG. Fragile X AGG analysis provides new risk predictions for 45-69 repeat alleles. *Am J Med Genet A*. 2013 Apr;161A(4):771-8. doi: 10.1002/ajmg.a.35833. Epub 2013 Feb 26.
- [17] Pretto D, Yrigollen CM, Tang HT, Williamson J, Espinal G, Iwahashi CK, Durbin-Johnson B, Hagerman RJ, Hagerman PJ, Tassone F. Clinical and molecular implications of mosaicism in FMR1 full mutations. *Front Genet*. 2014 Sep 17;5:318.
- [18] Yrigollen CM1, Martorell L2, Durbin-Johnson B3, Naudo M2, Genoves J2, Murgia A4, Polli R4, Zhou L5, Barbouth D6, Rupchock A6, Finucane B7, Latham GJ8, Hadd A8, Berry-Kravis E5, Tassone F9. AGG interruptions and maternal age affect FMR1 CGG repeat allele stability during transmission. *J Neurodev Disord*. 2014;6(1):24.
- [19] Handt M, Epplen A, Hoffjan S, Mese K, Epplen JT, Dekomien G Point mutation frequency in the FMR1 gene as revealed by fragile X syndrome screening. *Mol Cell Probes*. 2014 Aug 27 S0890-8508(14)00042-5.
- [20] Biancalana V, Glaeser D, McQuaid S, Steinbach P. EMQN best practice guidelines for the molecular genetic testing and reporting of fragile X syndrome and other fragile X-associated disorders. *Eur J Hum Genet*. 2014 Sep 17. doi: 10.1038/ejhg.2014.185.
- [21] Filipovic-Sadic S, Sah S, Chen L, Krosting J, Sekinger E, Zhang W, Hagerman PJ, Stenzel TT, Hadd AG, Latham GJ, Tassone F. A novel FMR1 PCR method for the routine detection of low abundance expanded alleles and full mutations in fragile X syndrome. *Clin Chem*. 2010 Mar;56(3):399-408.
- [22] Hersh JH, Saul RA; Committee on Genetics. Health supervision for children with fragile X syndrome. *Pediatrics* 2011;127:994-1006.
- [23] Weck KE, Zehnbauser B, Datto M, Schrijver I; CAP/ACMG Biochemical and Molecular Genetics Resource Committee. Molecular genetic testing for fragile X syndrome: laboratory performance on the College of American Pathologists proficiency surveys (2001-2009). *Genet Med* 2012;14:306-12.
- [24] McConkie-Rosell A, Abrams L, Finucane B, Cronister A, Gane LW, Coffey SM, et al. Recommendations from multi-disciplinary focus groups on cascade testing and genetic counseling for fragile X-associated disorders. *J Genet Couns* 2007;16:593-606.
- [25] Castellví-Bel S, Fernández-Burriel M, Rifé M, Jiménez D, Mallolas J, Sánchez A, et al. Detection of the fragile X syndrome protein for the evaluation of FMR1 intermediate alleles. *Hum Genet*. 2000; 107:195-6.
- [26] Ramos FJ, Willemsen R. Diagnosis of the fragile X syndrome by the analysis of FMRP expression in blood and hair roots *Arch Pediatr*. 2003;10:401-2
- [27] Oostra BA1, Willemsen R. FMR1: a gene with three faces. *Biochim Biophys Acta*. 2009 Jun;1790(6):467-77.
- [28] Sidorov MS1, Auerbach BD, Bear MF. Fragile X mental retardation protein and synaptic plasticity. *Mol Brain*. 2013 Apr 8;6:15. doi: 10.1186/1756-6606-6-15.

- [29] Willemsen R, Levenga J, Oostra BA. CGG repeat in the FMR1 gene: sizematters. *Clin Genet*. 2011 Sep;80(3):214-25
- [30] Levenga J, de Vrij FM, Oostra BA, Willemsen R. Potential therapeutic interventions for fragile X syndrome. *Trends Mol Med*. 2010 Nov;16(11):516-27.
- [31] Hunsaker MR, Goodrich-Hunsaker NJ, Willemsen R, Berman RF. Temporal ordering deficits in female CGG KI mice heterozygous for the fragile X premutation. *Behav Brain Res*. 2010 Dec 1;213(2):263-8.
- [32] Santos AR, Kanellopoulos AK, Bagni C. Learning and behavioral deficits associated with the absence of the fragile X mental retardation protein: what a fly and mouse model can teach us. *Learn Mem*. 2014 Sep 16;21(10):543-55.
- [33] Sullivan SD, Welt C, Sherman S. FMR1 and the continuum of primary ovarian insufficiency. *Semin Reprod Med*. 2011 Jul;29(4):299-307 Review.
- [34] Wheeler AC, Raspa M, Green A, Bishop E, Bann C, Edwards A, Bailey DB Jr. Health and reproductive experiences of women with an FMR1premutation with and without fragile X premature ovarian insufficiency. *Front Genet*. 2014 Sep 8;5:300.
- [35] Rodriguez-Revenga L, Madrigal I, Pagonabarraga J, Xunclà M, Badenas C, Kulisevsky J, Gomez B, Milà M. Penetrance of FMR1 premutation associated pathologies in fragile X syndrome families. *Eur J Hum Genet*. 2009 Oct;17(10):1359-62
- [36] Hagerman R1, Hagerman P. *Lancet Neurol*. 2013 Aug;12(8):786-98. Advances in clinical and molecular understanding of the FMR1premutation and fragile X-associated tremor/ataxia syndrome.
- [37] Tassone F, Greco CM, Hunsaker MR, Seritan AL, Berman RF, Gane LW, Jacquemont S, Basuta K, Jin LW, Hagerman PJ, Hagerman RJ.Neuropathological, clinical and molecular pathology in female fragile X premutation carriers with and without FXTAS. *Genes Brain Behav*. 2012 Jul;11(5):577-585.
- [38] *New research on Fragile X syndrome* Jeffrey J Upner (editor) Nova Science Publishers Inc NY ISBN 13 978-1-60021-281-9 Chapter 6 97-117 L Rodriguez-Revenga and Montserrat Milà
- [39] Tassone F. Methylation analysis in newborn screening for fragile X syndrome--reply. *JAMA Neurol*. 2014 Jun;71(6):800-1. doi: 10.1001/jamaneurol.2014.45430.
- [40] Hagerman RJ, Des-Portes V, Gasparini F, Jacquemont S, Gomez-Mancilla B. Translating molecular advances in fragile X syndrome into therapy: a review. *J Clin Psychiatry*. 2014 Apr;75(4):e294-307. doi: 10.4088/JCP.13r08714. Review.

Chapter 88

FRAGILE X-ASSOCIATED PRIMARY OVARIAN INSUFFICIENCY (FXPOI)

Maria-Isabel Tejada

Laboratorio de Genética Molecular, Servicio de Genética, Hospital Universitario Cruces,
BioCruces Health Research Institute, Barakaldo-Bizkaia,
GCV-CIBER de Enfermedades Raras (CIBERER), Spain

ABSTRACT

Fragile X-associated primary ovarian insufficiency (FXPOI) is among the family of disorders caused by the expansion of a CGG triplet repeat in the *FMR1* gene. FXPOI is a new clinical entity in which, carrier premutation (PM) females (with 56 to 200 CGG repeats) present early ovarian dysfunction, with menopause occurring 5 years earlier than non-carrier family members.

It has been estimated that 2-6% of all women with premature ovarian failure and a normal karyotype have a PM. Therefore, *FMR1* testing is recommended in all women with confirmed ovarian failure; i.e., cessation of the menstrual cycle during three-four months with elevated FSH levels of >30U/L.

Despite abundant literature, all authors agree that only a subset of PM carriers develop FXPOI (about 13-26%), with the precise molecular mechanisms of how the *FMR1* premutation causes FXPOI not being well understood. Nonetheless, recent studies have attempted to provide some insight into these mechanisms. This chapter summarizes some of these studies. Among these reports, the most important findings are: 1) the significant positive association of repeat size with POI, demonstrating that women with fewer than 100 repeats have an increased risk of FXPOI, and 2) the hypothesis that the *FMR1* PM may have a toxic RNA gain-of-function effect on ovarian follicle dynamics. These findings have also been demonstrated in rodent models in which the *FMR1* gene protein (FMRP) is highly expressed in oocytes which are important for folliculogenesis. Indeed, the two PM mouse models studied to date have shown evidence of ovarian dysfunction and increased expression of *FMR1* mRNA in the ovary.

Keywords: FMR1, FXPOI, ovarian dysfunction, POF

INTRODUCTION

Fragile X-associated primary ovarian insufficiency (FXPOI) is among the family of disorders caused by the expansion of a CGG triplet repeat in the *FMR1* gene. The most important syndrome caused by this expansion is Fragile X syndrome (FXS).

As seen in Chapter 1, FXS is the most common cause of hereditary intellectual disability (ID) and is one of the most frequent genetic diseases [1]. It is a dominant X-linked disease with incomplete penetrance, affecting approximately one in 2633 men in Spain [2]. This syndrome is caused by the anomalous expansion of a CGG repeat sequence in the 5' untranslated region (UTR) of the X-linked gene *FMR1* located in the FRAXA locus in Xq27.3 [3]. Transmission of this syndrome is complex since the number of CGG repeats in the population appears in four ranges: normal (N), with repeats between 6 and 45; intermediate between 46 and 54; premutation (PM) between 55 and 200; and full mutations (FM) with over 200 repeats (Figure 1). In the latter case, a second mechanism is triggered; that is, the methylation of the *FMR1* gene promoter, resulting in a silencing of thereby not allowing the production of the Fragile X Mental Retardation Protein (FMRP) which is the real cause of the syndrome [4]

Women with a PM do not generally have ID since the *FMR1* gene is not methylated (Figure 2). Nonetheless, these women have the risk of transmitting the syndrome [3], with PMs expanding to a FM in successive generations. Additionally, intermediate or grey alleles may or may not be unstable [5], with any expansion to a FM until successive generations.

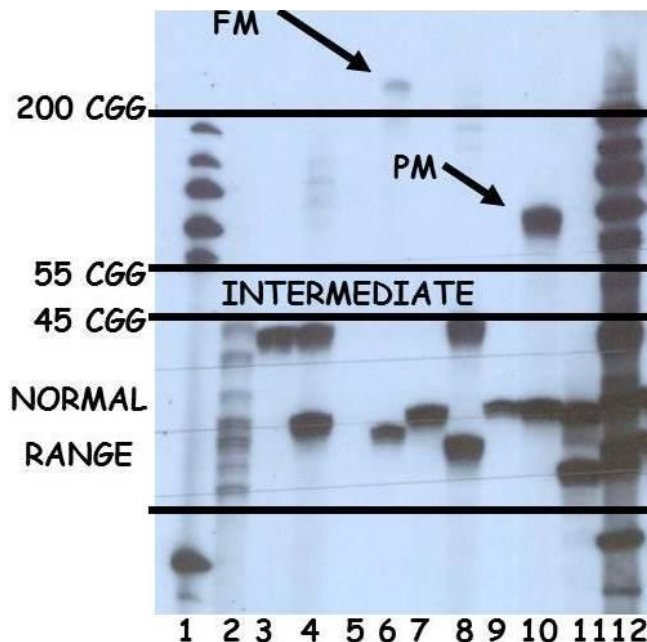


Figure 1. PCR Study of CGG repeats in females. PCR products analyzed by electrophoresis through a denaturing acrylamide gel, blotting, hybridization with an oligo (CGG)₅ and visualized by autoradiography after detection with Digoxigenin and CSPD. Lanes 1 and 12: molecular DNA weight markers. Lane 2: in home ladder. Lanes 4, 8 and 11: heterozygous normal females. Lanes 3, 7 and 9: homozygous normal females. Lane 6: female with the FM and lane 10: female with the PM. (Figure from the laboratory of Dr. Tejada).

Patients with a PM were initially thought to be completely asymptomatic, however over time this belief has changed. Fragile X families have been studied since 1996 [6], showing a higher frequency of premature ovarian failure (POF) compared with the normal population. In 1999, an international collaborative study [7] reported that 16% of the carriers with a PM had POF or presented early menopause (before the age 40). Since then many groups studying FXS have confirmed this disorder in PM carriers [8-10].

Over the years POF resulting in early menopause has been confirmed as a phenotypic characteristic of PM carriers although only about 13-26% of PM carriers actually present this disorder (1:650 females) [11]. On the other hand, an estimated 2-6% of all women with POF and a normal karyotype have a PM [8,12]. Interestingly, FM carriers do not present ovarian dysfunction. To the contrary, the loss of normal ovarian function has been demonstrated in women with intermediate alleles [11,13], and it has even been suggested that women with alleles with more than 30 repeats have a higher risk of having a diminished ovarian reserve [14].

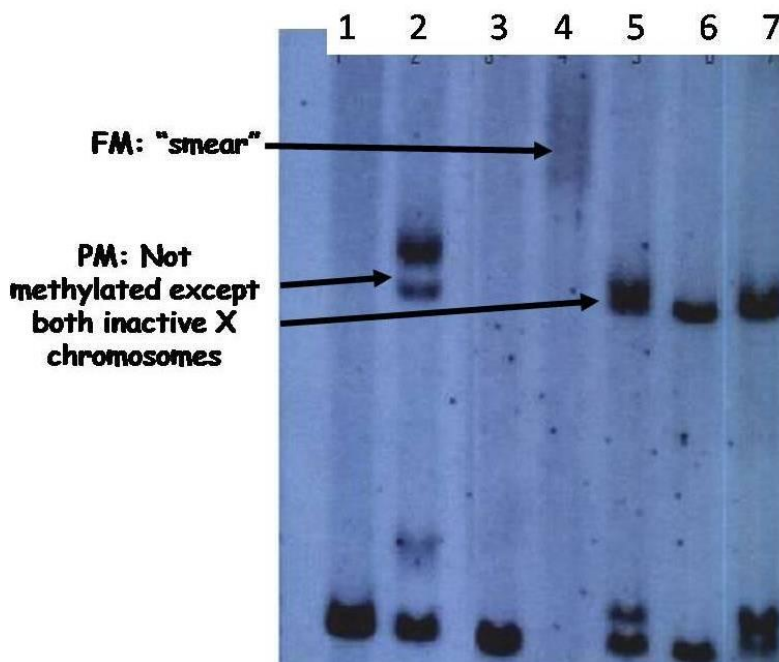


Figure 2. Southern blot analysis. DNA digested with EcoRI and EagI. Hybridization with the StB12.3 probe visualised by autoradiography after detection with Digoxigenin and CSPD. Lanes 1 and 3: normal males. Lanes 2, 5 and 7: females with the PM. Lanes 4: an affected FM male and lane 6: normal female. (Figure from the laboratory of Dr. Tejada).

In 2007, the National Fragile X Foundation defined the term “FXPOI” [15] as a condition in which there is a primary ovarian insufficiency (POI) with menopause occurring at 40 to 45 years of age in association with a PM of the *FMR1* gene [16]. The term –FXPOI– was considered more appropriate than Premature Ovarian Failure (FXPOF). Indeed from the point of view of reproductive counselling, women with FXPOI are able to become pregnant following periods of ovarian failure while those with POF are not. Nonetheless, these pregnancies may lead to a child with FXS.

DEFINING THE DISORDER

Primary ovarian insufficiency is also known as premature menopause, premature ovarian failure, hypergonadotropic amenorrhea and hypergonadotropic hypogonadism, conditions occurring due to chemotherapy, radiation or surgery [17]. At present, the term POI is increasingly used because it better describes impaired ovarian function as a continuum rather than a specific endpoint [18]. FXPOI may be transient or progressive, and usually results in eventual premature menopause. Menopause is the permanent cessation of menses that normally occurs at an average age of 50 years, whereas POF and POI generally describe a disorder consisting of amenorrhea, elevated menopausal level gonadotropins and sex steroid deficiency in women less than 40 years old.

In 2004, Welt and colleagues [19] reported that PM carriers have early ovarian aging. Since then, many different publications have appeared [reviewed in 16,20] demonstrating that FXPOI is the result of low ovarian reserves, which is manifested as a decrease in fertility. On comparison with normal control women FXPOI is associated with increased levels of some hormones, especially the follicle stimulating hormone (FSH) and with the decreased duration of one of the menstrual cycle phases, the so-called follicular phase.

Similarly, Anti-Müllerian hormone (AMH) levels, a marker of the growing cohort of small follicles, have been demonstrated to be lower among PM carriers compared with non-carriers of all ages (18 to 50 years). Overall, in women with a PM it seems that AMH is a good indicator of ovarian senescence, being even better than FSH concentrations [16, 20].

In short, FXPOI is a new clinical entity in which PM females (with 55 to 200 CGG repeats) present early ovarian dysfunction, with menopause occurring 5 years earlier than non-carrier family members.

DIAGNOSIS AND GENETIC COUNSELLING

FMRI testing is recommended in all women with confirmed ovarian failure: cessation of menstrual cycles during three-four months with elevated FSH ($>30\text{U/L}$) in two determinations carried out within a two-month period and with lower than normal AMH levels. Unfortunately, standardized diagnostic criteria for POI have not yet been established. Nonetheless, determination of estradiol (E2) and prolactine levels in serum may help in establishing the diagnosis [21]. The screening history should also focus on a family history of early menopause, autism and ID. However, in cases in which the diagnosis of ovarian failure is not clear (women with fertility concerns but normal or erratic cycles and women with occult forms of POI) but which present elevated FSH levels, *FMRI* testing is also recommended [20].

This recommendation is especially important in women following fertility treatments since carriers of a PM may have children with FXS [22].

In conclusion, *FMRI* testing is undoubtedly important in women with reduced fertility from the point of view of Genetic Counseling. Nonetheless, CGG studies are also indicated in other women, such as in the case of egg donors due to the high frequency of PMs and intermediate alleles in the population [23].

MECHANISMS PRODUCING FXPOI

Despite abundant literature, all authors agree that only a subset of PM carriers develops FXPOI, however, it remains unknown why some present POI while others do not.

The precise molecular mechanisms of how the *FMR1* premutation causes FXPOI are not well understood, although recent studies have attempted to provide some insight into these mechanisms. Among these reports, a significant positive association of repeat size has been demonstrated with POI and it has been found that women with fewer than 100 repeats have an increased risk of FXPOI, although this relationship is not linear [10, 24]. In this range, it seems that women with 80 to 100 repeats have the highest risk of presenting altered cycle traits and subfertility [20]. Carriers of both smaller and larger PM repeat sizes also present ovarian insufficiency, albeit with a lower frequency. Furthermore, a loss of normal ovarian function has also been demonstrated in women with intermediate alleles [11].

A few studies have described the CGG structure in women with FXPOI, that is, the number of AGG interruptions (See Figure 3 Chapter 1). AGG interruptions have been shown to decrease the risk of the expansion of a PM to a FM, and in one small study it was shown that women with POI and intermediate alleles had no AGG interruptions [13]. Nevertheless, a study by our group [25] did not confirm this observation.

Others studies have postulated the hypothesis that the *FMR1* PM may have a toxic RNA gain-of-function effect on ovarian follicle dynamics. In fact, studies on *FMR1* gene expression in males with the PM showed elevated *FMR1* mRNA levels, and this increased transcriptional activity appeared to be positively correlated with CGG repeat size and with the manifestation of FXTAS [26, 27]. Similarly, the effect of the CGG repeats on the mRNA levels was statistically significant in patients with POF, but not in those without POF (Figure 3) [10], demonstrating RNA gain-of-function toxicity in PM carrier women.

However, the relationship between molecular status and phenotype is much more complex in women, in whom the activation ratio of the X chromosome must be considered [9]. Nevertheless, levels of the *FMR1* transcript among female premutation carriers have been demonstrated to be increased, with the relationship between mRNA levels and repeat size being nonlinear, with a cut-off at 100 repeats [10]. It has also been suggested that due to skewed X inactivation mRNA levels tend to normalize in females with an increase in the number of CGG repeats [9].

Indeed, since the *FMR1* gene is located on the X chromosome, it has been proposed that X-chromosome inactivation may play a role in the possible development of FXPOI. However, the first study by Murray et al., [28] reported no significant effect on comparing FXPOI, *FMR1* repeat size and X-chromosome inactivation patterns, having been confirmed by other studies [20].

Regarding FMRP levels, some studies have described increased FMRP levels in the range of 80 to 89 repeats in carrier males [29]; however, similar studies have not been carried out in women [20].

In relation to the parental origin of the PM allele, Hundscheid and co-workers (2000) [30] suggested a role for imprinting on finding a significant effect between FXPOI and parental origin. Nonetheless, further studies have failed to confirm these findings [8].

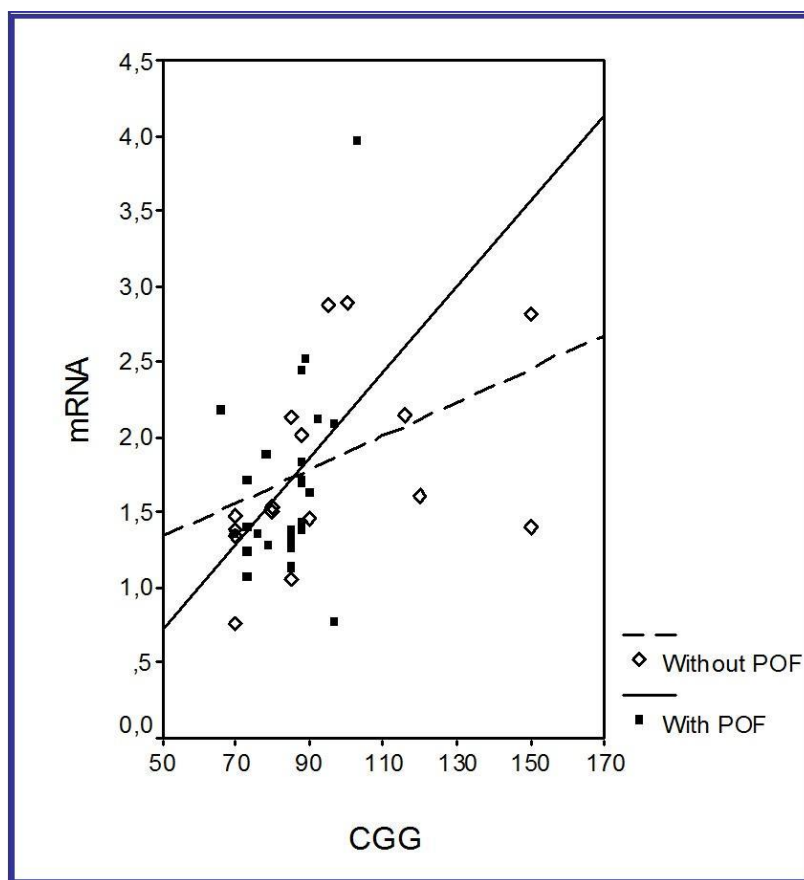


Figure 3. Relation between the CGG repeats and mRNA levels depending on the POF status of the patients. The effect of the CGG repeats on the mRNA levels was statistically significant ($p=0.0437$) in patients with POF, but not ($p=0.0724$) in patients without POF (Figure from Tejada et al., 2008 [10]).

Finally, FXPOI may not only depend on the PM allele but also on other modifier genes and environmental factors. Hunter et al., [31] showed significant familial aggregation of age at menopause, providing evidence of the presence of additional genes influencing and interacting with FXPOI. The identification of additional genes involved in ovarian function is therefore a new step in future studies. In this regard, Allen and co-workers (2014) [32] performed a study of a candidate gene using whole genome sequencing (WGS) to identify possible variants that may influence the onset of FXPOI. Although no conclusions could be drawn, they did find some potentially damaging variants that deserve further investigation.

In short, to date neither the exact etiology of FXPOI nor the cause of the variation in phenotype is understood, and thus, additional studies are clearly needed.

MOUSE MODELS

Non-invasive methods to study the mechanisms of the PM in ovarian function in humans are currently not available. This is why the development of model systems is necessary to enhance the knowledge of FXPOI. Only recently have mutation vertebrate (mouse and rat) and

invertebrate (*Drosophila melanogaster*) models been used to study ovarian function and the results published to date indicate the value of studying the etiology of FXPOI. The recent work by Sherman and colleagues (2014) [33] reviews and summarizes everything published so far.

Rodent models have basically shown that FMRP is highly expressed in oocytes in which this protein is important for folliculogenesis. The two PM mouse models studied to date [34,35] have showed evidence of ovarian dysfunction and, together, suggest that the long repeat in the transcript itself may have some pathological effect very different from any effect of the toxic protein.

Furthermore, ovarian morphology in young animals appears normal and the primordial follicle pool size does not differ from that of wild-type animals. However, there is a progressive premature decline in the levels of most follicle classes. Observations also include granulosa cell abnormalities and altered gene expression patterns. In this line, in a rat model, Ferder and collaborators (2013) [36] found changes in *FMR1* expression during follicle maturation, both at the protein and mRNA levels. Indeed, in all PM mouse models, increased expression of *FMR1* mRNA has been described in the ovary.

Premutation model systems in non-human primates and those based on induced pluripotent stem cells have shown particular promise and will complement current models, although definite conclusions remain to be established.

Fly models have shown to be very useful because of the relative ease of model construction compared with other model systems. For example, *Drosophila* germ line stem cells have been used as a model to show that FMRP can modulate the fate of stem cells [37]. This study suggested that both the reduction of FMRP and expression of PM CGG repeats could have detrimental effects on fly ovary and stem cell maintenance.

Among other ideas in progress, Sherman and co-workers (2014) [33] have also highlighted the possibility of testing the effect of genetic modifiers on the ovarian phenotype. This could be valuable for not only understanding the pathogenic mechanism, but may also shed light on human homolog genes of which may contribute to the variable penetrance of FXPOI.

In summary, all these studies mentioned above are very promising but definitive conclusions remain to be drawn. Further comparisons of all of these models are still needed to gain insight into the etiology of ovarian dysfunction.

CONCLUSION

FXPOI is a new clinical entity in which carrier PM females of the *FMR1* gene may present early ovarian dysfunction with menopause occurring at approximately 40 years of age. To date neither the exact etiology of FXPOI nor the cause of the variation in phenotype is understood. New technologies with next generation sequencing and studies in animal models are very promising, but definitive conclusions remain to be drawn, thereby requiring further investigations in this pathology.

REFERENCES

- [1] Hagerman RJ and Hagerman PJ.: *Fragile X Syndrome, Diagnosis, treatment and research*. Edited by Hagerman and Hagerman. The Johns Hopkins University Press 2002.
- [2] Fernández-Carvajal I, Walichiewicz P, et al., Screening for expanded alleles of the FMR1 gene in blood spots from newborns males in a Spanish population. *J. Mol. Diagn.* 2009;11(4):324-9.
- [3] Oberlé I, Rousseau F, Heitz D, et al.,: Instability of a 550-base pair DNA segment and abnormal methylation in Fragile X syndrome. *Science*. 1991;252:1097-102.
- [4] Rousseau F, Heitz D, Tarleton J, et al.,: A multicenter study on genotype-phenotype correlations in the Fragile X syndrome, using direct diagnosis with probe StB12.3: The first 2.253 cases. *Am. J. Hum. Genet.* 1994;55:225-237.
- [5] Nolin SL, Brown WT, Glicksman A, et al.,: Expansion of the Fragile X CGG repeat in females with Premutation or Intermediate alleles. *Am. J. Hum. Genet.* 2003;72:454-464.
- [6] Vianna-Morgante AM, Costa SS, Pares AS, et al.,: FRAXA premutation associated with premature ovarian failure. *Am. J. Med. Genet.* 1996;64:373-375.
- [7] Allingham-Hawkins DJ, Babul-Hirji R, Chitayat D, et al., Fragile X premutation is a significant risk factor for premature ovarian failure: the International Collaborative POF in Fragile X study-preliminary data. *Am. J. Med. Genet.* 1999;83:322-325.
- [8] Mallolas J, Duran M, Sanchez A, et al.,: Implications of the FMR1 gene in menopause: study of 147 Spanish women. *Menopause*. 2001;8(2):106-10.
- [9] García-Alegría E, Ibáñez B, Mínguez M, et al., Analysis of FMR1 gene expression in female premutation carriers using robust segmented linear regression models. *RNA*. 2007;13:756-62.
- [10] Tejada MI, García-Alegría E, Bilbao A, et al.,: Analysis of the molecular parameters that could predict the risk of manifesting POF in female PM carriers of fragile X syndrome. *Menopause*. 2008;15(5):945-9.
- [11] Wittenberger MD, Hagerman RJ, Sherman SL, et al., The FMR1 premutation and reproduction. *Fertil Steril*. 2007;87:456-65.
- [12] Murray A, Schoemaker MJ, Bennet CE, et al., Population-based estimates of the prevalence of FMR1 expansion mutations in women with early menopause and primary ovarian insufficiency. *Genetics in Medicine*. 2014;16:19-24.
- [13] Bodega B, Bione S, Dalpra L, et al.,: Influence of intermediate and uninterrupted FMR1 CGG expansions in premature ovarian failure manifestation. *Hum. Reprod.* 2006;21:952-7.
- [14] Gleicher N, Weghofer A and Barad DH.: A pilot study of premature ovarian senescence: I. Correlation of triple CGG repeats on the FMR1 gene to ovarian reserve parameters FSH and anti-Müllerian hormone. *Fertil Steril*. 2009;91:1700-1706.
- [15] Nelson LM. Menstrual cycles, the FMR1 Premutation an "Primary Ovarian Insufficiency". *The National Fragile X Foundation Quarterly*. 2007;26:22-23.
- [16] De Caro JJ, Dominguez C and Sherman SL.: Reproductive health of adolescent girls who carry the FMR1 Premutation: Expected phenotype based on current knowledge of Fragile X-Associated primary ovarian insufficiency. *Ann. NY Acad. Sci.* 2008;1135:99-111.
- [17] Nelson LM, Covington SN, and Rebar RW.: An update: spontaneous premature ovarian failure is not an early menopause. *Fertil Steril* 2005;83:1327-1332.

- [18] Nelson LM. Clinical practice. Primary ovarian insufficiency. *N. Engl. J. Med.* 2009;360:606-614.
- [19] Welt CK, Smith PC and Taylor AE.: Evidence of early ovarian aging in Fragile X premutation carriers. *J. Clin. Endocrinol. Metab.* 2004;89:4569-74.
- [20] Sullivan SD, Welt C and Sherman S.: FMR1 and the Continuum of Primary Ovarian Insufficiency. *Semin. Reprod. Med.* 2011;29:299-307.
- [21] Tejada MI.: Manifestaciones clínicas en las mujeres con premutación. Fallo Ovárico Prematuro (POF). En: *Síndrome X Frágil*. Libro de consulta para familias y profesionales, pags. 39-44. Edita: Real Patronato sobre discapacidad, Madrid 2006. [Clinical manifestations in women with premutation. Premature Ovarian Failure (POF). In: *Fragile X Syndrome*. Book of consultation for families and professionals, pages. 39-44. Edit: Royal Board on disability, Madrid 2006].
- [22] Santoro N. Research on the mechanisms of premature ovarian failure. *J. Soc. Gynecol. Investig.* 2001;8:S10-2.
- [23] Crawford DC, Acuña JM, Sherman SL. FMR1 and the fragile X syndrome: human genome epidemiology review. *Genet Med.* 2001;3:359-371.
- [24] Ennis S, Ward D, Murray A. Nonlinear association between CGG repeat number and age of menopause in FMR1 premutation carriers. *Eur. J. Hum. Genet.* 2006;14:253-5.
- [25] Tejada MI, Maortua H, Mendoza R, et al., Study of triple CGG repeats (FMR1 gene) in 41 patients with a reduced ovarian response to gonadotropin stimulation: no correlation found with four ovarian parameters. *Hum. Reprod.* 2013;28:241-2.
- [26] Tassone F, Hagerman RJ, Taylor AK, et al. Elevated Levels of FMR1 mRNA in Carrier Males: A New Mechanism of involvement in the Fragile-X Syndrome. *Am. J. Hum. Genet.* 2000;66:6–15.
- [27] Jacquemont S, Hagerman RJ, Leehey M, et al., Fragile X premutation tremor/ataxia syndrome: molecular, clinical, and neuroimaging correlates. *Am. J. Hum. Genet.* 2003;72:869-78.
- [28] Murray A, Ennis S, Mac Swiney F, et al., Reproductive and menstrual history of females with fragile X expansions. *Eur. J. Hum. Genet.* 2000;8:247-52.
- [29] Peprah E, He W, Allen E et al., Examination of FMR1 transcript and protein levels among 74 premutation carriers. *J. Hum. Genet.* 2010;55:66-68.
- [30] Hunscheid RD, Sistermans EA, Thomas CM, et al., Imprinting effect in premature ovarian failure confined to paternally inherited fragile X premutations. *Am. J. Hum. Genet.* 2000;66:413-28.
- [31] Hunter JE, Epstein MP, Tinker SW, et al., Fragile X-associated primary ovarian insufficiency: evidence for additional genetic contributions to severity. *Genet Epidemiol.* 2008;32:553-559.
- [32] Allen EG, Grus WE, Narayan S, et al., Approaches to identify genetic variants that influence the risk for onset of fragile X-associated primary ovarian insufficiency (FXPOI): a preliminary study. *Frontiers in Genetics.* 2014;5:1-8.
- [33] Sherman S, Curnow EC, Easley ChA, et al., Use of model systems to understand the etiology of fragile X-associated primary ovarian insufficiency (FXPOI). *Journal of Neurodevelopmental Disorders.* 2014; 6:26.
- [34] Lu C, Lin L, Tan H, et al., Fragile X premutation RNA is sufficient to cause primary ovarian insufficiency in mice. *Hum. Mol. Genet.* 2012, 21:5039–5047.

- [35] Ascano M Jr, Mukherjee N, Bandaru P, et al., FMRP targets distinct mRNA sequence elements to regulate protein expression. *Nature*. 2012, 492:382–386.
- [36] Ferder I, Parborell F, Sundblad V, et al., Expression of fragile X mental retardation protein and FMR1 mRNA during folliculogenesis in the rat. *Reprod Abstr. Ser.* 2013, 145:335–343.
- [37] Yang L, Duan R, Chen D, et al., Fragile X mental retardation protein modulates the fate of germline stem cells in *Drosophila*. *Hum. Mol. Genet.* 2007, 16:1814–20.

Chapter 89

FRAGILE X-ASSOCIATED TREMOR/ATAXIA SYNDROME

L. Rodriguez-Revenge

Biochemistry and Molecular Genetics Department. Hospital Clínic,
Barcelona, Spain
CIBER de Enfermedades Raras (CIBERER),
Barcelona, Spain
IDIBAPS (Institut d'Investigacions Biomèdiques August Pi i Sunyer),
Barcelona, Spain

ABSTRACT

Fragile X-associated tremor/ataxia syndrome (FXTAS) is a late-onset inherited neuropsychiatric degenerative disorder that occurs predominantly in male carriers of the *FMR1* premutation (55-200 CGG repeats). *FMR1* premutation is relatively frequent in the general population, affecting approximately 1 out of 800 males and 1 out of 250 females, and leading to symptoms of FXTAS in up to 1 in 3,000 men older than 50 years. Clinical symptoms in FXTAS patients usually begin with an action tremor. After that, different findings including ataxia (balance problems with frequent falling), and more variably, loss of sensation in the distal lower extremities and autonomic dysfunction (e.g., impotence, hypertension, and loss of bowel and bladder function), may occur, and gradually progress. Molecular mechanism leading to FXTAS is distinct from the *FMR1* silencing mechanism and/or a deficit in FMRP operating in fragile X syndrome. Individuals with *FMR1* premutation alleles have markedly elevated levels of expanded CGG-repeat *FMR1* mRNA, which is thought to have a toxic gain-of-function. Since 2001, when FXTAS was first described, the advancement of our understanding of the clinical phenotype as well as the molecular pathophysiology has occurred very quickly. The aim of this chapter is to present the most recent advances in the current knowledge of FXTAS.

Keywords: *FMR1*, FXTAS, ataxia, tremor, mRNA toxicity

INTRODUCTION

Fragile X-associated tremor/ataxia syndrome, FXTAS (OMIM# 300623) was identified in 2001 by Hagerman and co-workers as a late-onset neurodegenerative disorder [1]. It took more than 10 years after the *FMR1* gene identification to recognize FXTAS as a *FMR1* premutation associated phenotype. One explanation for this is that the movement disorder experienced by older carriers, who were thought to be clinically normal, was not associated with the fragile X syndrome (FXS) (a childhood disorder) affecting children. Mothers of FXS children, being seen in clinics, were often expressing concerns about their fathers (*FMR1* premutation carriers) who were experiencing problems with hand tremor and unsteady gait. When Hagerman and co-workers evaluated these male carriers they found that they were more likely to develop, from 50 years onwards, a stereotyped clinical picture characterized by unsteadiness while walking and action tremor in both hands. Both symptoms used to follow a progressive course, and were often accompanied by progressive cognitive and behavioral disturbances [1]. Although, FXTAS was firstly identified among older *FMR1* premutation male carriers, to date it is well established that it also affects females. Nevertheless, it has been suggested that it occurs less frequently and that the phenotype is milder with older age at onset [1, 2]. An explanation for this difference is the presence of a second normal allele and a random X-inactivation of the premutated one; however, there may be additional sex-specific effects that reduce penetrance among females [3].

FXTAS penetrance is not complete; meaning that not all *FMR1* premutation carriers develop FXTAS. It is largely dependent on carrier age, with symptoms typically seen in 30% of males by the age of 50 and in 50% of males older than 70 years. Moreover, there is significant variability in the progression of neurological dysfunction [2-5].

The description and characterization of FXTAS syndrome is of great interest to the population, because the prevalence of *FMR1* premutation in the general population is relatively high. In fact, it is estimated that 1 in 800 to 1,200 males and from 1 in 250 to 400 females are carriers of a *FMR1* premutation allele [6-8]. Although the contribution of FXTAS to the morbidity and mortality of the aging population requires further study, the disorder is likely the most common single-gene form of tremor and ataxia in the older adult population. Taking into consideration the prevalence of *FMR1* premutation in the general population, the prevalence of FXTAS might be estimated in ~1/3,000 males aged over 50 years of age (~1/10,000 males of all ages) [3].

FRAGILE X-ASSOCIATED TREMOR/ATAXIA SYNDROME

FXTAS Clinical and Cognitive Overview

Originally FXTAS was clinically described as a motor disorder with core features including kinetic tremor and cerebellar gait ataxia. Clinical symptoms of FXTAS syndrome appear in patients in their 50s and they usually begin with an action tremor. Additional clinical features include peripheral neuropathy (60%), impotence (80%), bowel and bladder dysfunction (30-55%), erectile dysfunction, loss of sensation in the distal lower extremities, hearing loss and dysphagia [reviewed in 9].

Over the last years, the clinical spectrum of the disease has expanded and new data have evidenced that progressive cognitive decline, with a gradual progression to dementia in some individuals, and neuropsychological problems are common among FXTAS patients [10-16]. Several studies comparing psychiatric symptoms in FXTAS with matched comparison groups have been performed. Overall, they describe that the neuropsychiatric phenotype of FXTAS is characterized primarily by poor performance on measures of executive function, working memory, information processing speed, and fine motor control. While attention and language function may also be affected in patients with FXTAS, there is limited evidence regarding the affection of visuospatial function. Secondary behavioral features may include depression, anxiety, and obsessive-compulsive symptoms. Finally all evidences concluded that this profile is consistent with a dysexecutive fronto-subcortical syndrome [10, 16, 17].

FXTAS clinical symptoms tends to progress from mild ataxia and/or tremor to a disabling condition impacting on motor daily activities, thinking, and social skills that seriously compromise patients' quality of life.

Because of the variety of clinical symptoms in FXTAS, genetic testing for FXTAS must be considered in any patient (male and female) with ataxia developing after the age of 50. Diagnostic testing consists of determining *FMR1* CGG repeat size by polymerase chain reaction (PCR) technique, which nowadays is a highly sensitive test performed in many laboratories.

FXTAS Neuroimaging and Neuropathological Profiles

Magnetic resonance imaging (MRI) of the brain was obtained in FXTAS patients searching for specific features. Individuals with FXTAS demonstrate moderate to severe generalized brain atrophy with ventricular enlargement, cerebellar atrophy, and subcortical and/or pontocerebellar white matter lesions [18, 19]. This pattern frequently includes hyperintensities in the cerebellar white matter and middle cerebellar peduncles on T2-weighted images, called the "MCP sign" [19]. These characteristic findings, initially described on conventional MRI, were classified into two categories -major and minor criteria-, and proposed, together with clinical findings, as diagnostic criteria for FXTAS (Table 1) [18].

Cerebellar abnormalities were confirmed by post-mortem histological studies that identified several neuropathological features, including Purkinje cell decreases and spongiform changes [20, 21]. Significant loss in whole cerebellar volume has been revealed in both male and female patients with FXTAS [22, 23].

Although the MCP sign is considered a major radiological feature of FXTAS, it has been also reported in patients with other forms of adult-onset cerebellar ataxia, thus resting specificity for FXTAS [24]. As more evidence among FXTAS individuals developing neurological features is being gained, the spectrum and variability of MRI features becomes broader and the potential usefulness of neuroimaging in providing insight in FXTAS becomes more evident [reviewed in 25]. Furthermore, with the advent of high field strength MRI as well as more sophisticated postprocessing tools the understanding of FXTAS neuropathological profile becomes wider. In this context, Hashimoto and co-workers [14] assessed focal changes of grey and white matter density in the brain of *FMR1* premutated carriers using voxel based morphometry method. In this study, patients with FXTAS demonstrated a distinct pattern of grey matter volume loss, involving multiple cortical and subcortical regions. This included

different parts of the cerebellum, as well as of the medial surface of the brain, including the dorsomedial prefrontal cortex, anterior cingulate and precuneus. Additional volume loss was seen in the lateral prefrontal cortex, orbitofrontal cortex, amygdala, and insula. More interestingly, there were significant correlations between grey matter loss in different brain regions, behavioral scales, and CGG repeats [14].

Table 1. Clinical criteria for FXTAS

Examination and Degree	Observation
Radiological:	
Major	MRI white matter lesions in MCPs and or brain stem
Minor	MRI white matter lesions in cerebral white matter
Minor	Moderate-to-severe generalized atrophy
Clinical:	
Major	Intention tremor
Major	Gait ataxia
Minor	Parkinsonism
Minor	Moderate-to-severe short-term memory deficiency
Minor	Executive function deficit

Inclusion criterion: CGG repeat number between 55 and 200.

Note. Data described by [18].

These radiologic features have been histologically confirmed in FXTAS post mortem cases. Early neuropathological, post mortem studies of the brain of patients with FXTAS revealed intranuclear inclusions in neurons and astrocytes throughout the cerebrum and brainstem, being most numerous in the hippocampal formation [20]. These inclusions appear as eosinophilic, hyaline, refractile, 2-5 μm diameter, round to ovoid bodies that show positive reactivity with antibodies against over 20 different proteins including ubiquitin, $\alpha\beta$ -crystallin, lamin A/C, hnRNP A2, myelin basic protein, DNA repair-ubiquitin-associated HR23B, and Sam68, among others [26-28]. The inclusions are PAS, silver, amyloid, and α -synuclein negative [20, 21]. Additionally, the *FMR1* mRNA but not the FMRP has been found contained within intranuclear inclusions [27, 29].

Although Purkinje cells rarely have inclusions, there is neurodegeneration in the cerebellum, with marked Purkinje cell loss, axonal swelling, and gliosis. Apart from reduced Purkinje cell number, additional neuropathological features present in FXTAS include, axonal torpedoes, and prominent cortical and subcortical white matter pathology [20, 21].

Recently, a broad distribution of intranuclear inclusions in non-central nervous system has been observed in *FMR1* premutation carriers with FXTAS [28]. Inclusions were found in somatic organs such as the endocrine organs, gastrointestinal tract, heart, and kidney [28]. Although these findings are consistent with the expanding range of co-morbid medical features reported in FXTAS (see chapter 5), intranuclear inclusions are not the cause of these conditions.

Nowadays FXTAS is considered as a new class of inclusion disorder. However there are several unclear aspects such as the cellular processes underling inclusion formation and whether the inclusions are themselves toxic or simply reflect underlying cellular dysfunction.

Molecular Genetics Overview

FXTAS is an allelic disorder to the FXS, and therefore should be considered as a distinct neurodegenerative disorder. Although both disorders are caused by expansions of the CGG repeat element found in the *FMR1* gene, the molecular mechanism leading to FXTAS is distinct from the *FMR1* silencing mechanism and/or a deficit in FMRP operating in FXS. In premutated patients the *FMR1* gene is rarely silenced and FMRP levels are generally normal or only slightly lowered [3]. The only known molecular abnormality among *FMR1* premutation carriers is the presence of markedly elevated levels (~2-8 fold) of *FMR1* mRNA.

The increased transcriptional activity of the *FMR1* gene seems to be positively correlated with the size of the CGG repeat. That is, CGG repeats in the upper range (100-200 CGG) result in average 5-8 fold elevation, whereas CGGs in the lower range (50-100 CGG) result in an average 2-fold elevation [30-33]. Although the precise mechanism for this overexpression is unknown, several possible mechanisms have been postulated. A feedback mechanism suggests that the cell attempts to compensate for reduced levels of FMRP by increasing the amount of available *FMR1* transcript [reviewed in 34, 35]. Alternatively, it is likely that the increasing length of the CGG repeat near the *FMR1* promoter proportionally opens the chromatin, allowing more ready access to transcription factors [36]. The presence of these elevated levels of abnormal (expanded CGG repeat) *FMR1* mRNA led to propose an RNA “toxic gain-of-function” model for FXTAS, in which the mRNA itself, with the abnormal CGG repeat tract, is causative of the neurological disorder [1, 3, 18, 20]. Animal and cell-based studies have demonstrated that the sole expression of expanded CGG repeats is necessary and sufficient to cause pathology similar to human FXTAS [37-40], and thus indicate that the expanded CGG repeats in RNA are the likely cause of the neurodegeneration in FXTAS.

In particular, the hypothesis of a RNA toxic gain-of-function suggests that CGG repeat *FMR1* mRNA recruits RNA-binding proteins and other proteins through direct RNA–protein interactions, but probably also through indirect protein–protein interactions. The sequestration of these proteins is also thought to prevent normal function leading to downstream alterations [34, 41].

Nowadays, several proteins have been identified that interact with the CGG repeat and might be potential mediators of downstream cellular dysregulation [reviewed in 42]. Table 2 summarizes candidate protein mediators of the CGG-repeat mRNA toxicity. The protein components of FXTAS inclusions fell into eight major functional categories, including: histone family; intermediate filament; microtubule; myelin-associated proteins; RNA-binding proteins; stress-related proteins; chaperones and ubiquitin-proteasome-related proteins [reviewed in 34]. Although most of them have been found to localize with ubiquitin-positive inclusions in CGG-expressing *Drosophila*, knock-in (KI) mouse model and FXTAS patients, they are not found to be sequestered by expanded CGG repeats and consequently they are not expected to lose their functions in FXTAS patients [43]. The lack of a principal protein species in the inclusions is not surprising, since there is no known abnormal protein product associated with FXTAS. In particular, the protein product of the *FMR1* gene, FMRP, is not structurally abnormal, as the expanded CGG repeat is located in a non-coding portion (5'-UTR, 5'-untranslated region) of the gene.

A striking characteristic of CGG expanded repeats is that they form dynamic intranuclear RNA aggregates that enlarge with time, resulting in the formation of giant inclusions. Continuous enlargement of CGG RNA aggregates suggests that these repeats may constantly

recruit proteins, implying a founding RNA-protein interaction event that would subsequently trap other proteins through indirect RNA-protein or protein-protein interactions [43].

Table 2. Candidate proteins involved in mediating the RNA-triggered pathogenesis

Protein	Function	Evidences	Reference
Sam68	Splicing modulator	- its ablation in mouse leads to motor dysfunction - is partially sequestered and functionally inactivated in FXTAS patients	[43, 44]
Pur α	Single-stranded cytoplasmatic DNA and RNA binding protein	- mediates neurodegeneration in <i>Drosophila</i> and mouse - binds CGG repeats in both mammalian and <i>Drosophila</i> brains - found to be present in the inclusions of FXTAS brain tissues - overexpression can alleviate neurodegeneration in FXTAS fly model	[45]
hnRNP A1 hnRNP A2/B1	RNA binding proteins	- assists mRNA dendritic transport in <i>drosophila</i> and cultured rat neurons - binds CGG repeats in both mammalian and <i>Drosophila</i> brains - found to be present in the inclusions of human FXTAS brain tissues - overexpression suppresses the phenotype of the CGG transgenic fly	[27, 46, 47]
DROSHA-DGCR8	microRNA nuclear processing complex	- binds to expanded CGG repeats - microRNA dysregulation in human FXTAS brain tissue - overexpression of DGCR8 rescues neuronal cell death	[48]
ubiquitin	post-translational modification	- present in the inclusions of human FXTAS brain samples	[27]
α B-crystallin	heat shock protein	- present in the inclusions of human FXTAS brain samples	[27]
Lamin A/C	intermediate filament proteins	- present in the inclusions of human FXTAS brain samples	[27]
MBP	myelin basic protein	- present in the inclusions of human FXTAS brain samples	[27]

Proteins recruited through protein-protein interactions have not been listed (e.g., CUGBP1, Rm62, Hsp70).

The current hypothesis contemplates that the CGG-repeat expansion drive to cellular dysregulation that cause eventual cell death. Mitochondrial abnormalities [49, 50], altered calcium regulation [51], altered lamin A nuclear architecture [52], and reduced telomere length [53], have been described as potential downstream pathways that could mediate cellular dysregulation [reviewed in 54].

Although FXATS is defined as a late-onset disorder, recent findings prompt to a process that begins in early development. Abnormalities of neuronal morphology and network activity as well as alerted calcium regulation have found in late embryonic and neonatal premutation mouse model [51, 55-57]. This observation suggests that there is an early developmental impairment that precedes cellular loss of viability, which somehow might explain the broad clinical phenotype associated with the *FMR1* premutation (See chapter 5).

As exposed in this chapter, there are several evidences that support the RNA-toxic-gain of function model for FXTAS. However, at least two additional mechanisms of toxicity have been described [reviewed in 58]. The first one compromises the translation in all reading frames of expanded CGG repeats in the absence of an ATG initiation codon (non-AUG initiated translation, RAN), leading to the production of homopolymeric protein. In particular, Todd and co-workers [59] demonstrated the existence of one particular CGG RAN translation product (polyglycine), which directly modulates CGG-associated pathology in two distinct model systems. Therefore on the basis of their observations it is suggested that a protein gain-of-function may also occur in cells of patients with FXTAS. Secondly, an antisense *FMR1* mRNA and a number of long noncoding RNAs have been identified within/near the *FMR1* gene [60-62]. Although the relevance of these transcripts is not fully understood they might have important functions and/or modulate certain aspects of FXTAS [reviewed in 58].

Overall, these observations raise the possibility of different or additive/synergistic molecular mechanisms contributing to the pathogenesis of FXTAS.

Even though the understanding of molecular basis of FXTAS has evolved over the last years and many different aspects of its pathogenesis have been unraveled, we still do not know how to link all of them. Furthermore, there is the puzzling issue of the incomplete penetrance. Additional genetic and/or environmental factors might play a key role and might be related with the emergence of FXTAS clinical symptoms. In this scenario, further research is needed since the better understanding off all these aspects might help to prevent or delay the onset of the syndrome.

FXTAS Treatment

Currently, there is no specific targeted treatment for patients with FXTAS. The present medical management of patients with FXTAS is limited to medications that targets specific symptoms such as tremor, ataxia, mood changes, depression, cognitive decline and/or dementia [reviewed in 9, 54]. Although neurological rehabilitative therapies have not been studied specifically in FXTAS they should be considered in treatment. Furthermore, occupational and physical therapy might also be beneficial and improve strength and stability in those with tremor, gait and balance deficits.

RNA toxicity initiates brain changes well before the onset of FXTAS. Typically, the age of onset of FXTAS is in the early 60s, and it presents with onset of tremor. However, as described before in this chapter different findings suggest that asymptomatic changes associated with the pathological changes of FXTAS exist before the onset of tremor or ataxia. Additional genetic and/or environmental factors might be necessary to precipitate FXTAS symptoms. In this regard, chemotherapy, prolonged surgery with general anesthesia or exposure to toxins, genetic predisposition to mitochondrial dysfunction, coexistence with

another disease such as multiple sclerosis or Alzheimer's disease have been described as factors that can precipitate FXTAS clinical manifestations [reviewed in 54].

On the basis of these observations, FXTAS treatment should envision two approaches; a direct one addressed to eliminate the expanded CGG-repeat mRNA and a prophylactic one, protecting neurological cells from neurotoxic chemicals.

CONCLUSION

Since 2001, when FXTAS was first described, the advancement of our understanding of the clinical phenotype as well as the molecular pathophysiology has occurred very quickly. Nowadays, FXTAS clinical description includes a much broader symptoms spectrum than the initial core features of intention tremor and gait ataxia. Furthermore, since the observation of increased expression of *FMR1* mRNA, a RNA toxicity hypothesis and other possible pathological mechanisms have been proposed with multiple studies supporting them. Although a lot has been achieved in this short period of time, further work is necessary to link and fully understand all the molecular mechanisms that might play a role in FXTAS pathology. The better understanding of these processes should also shed light on therapeutic approaches that will combat not only neurodegeneration but also the rest of symptoms. Finally, as the prevalence of *FMR1* premutated alleles is relatively high in general population, FXTAS may represent one of the more common monogenic causes of tremor, ataxia, and dementia. For this reason, it is probably that many carriers with FXTAS are being seen by a clinical specialist without awareness of the underlying genetic basis for the symptoms. The early diagnosis of those patients not only benefits themselves but also the rest of the family that should be advised for the FXS.

REFERENCES

- [1] Hagerman RJ, Leehey M, Heinrichs W, *et al.* Intention tremor, parkinsonism, and generalized brain atrophy in male carriers of fragile X. *Neurology*. 2001; 57: 127-130.
- [2] Jacquemont S, Farzin F, Hall D, *et al.* Aging in individuals with the FMR1 mutation. *Am. J. Ment. Retard*. 2004; 109: 154-164.
- [3] Hagerman PJ, Hagerman RJ. The fragile-X premutation: a maturing perspective. *Am. J. Hum. Genet*. 2004; 74: 805-816.
- [4] Jacquemont S, Hagerman RJ, Leehey MA, *et al.* Penetrance of the fragile X-associated tremor/ataxia syndrome in a premutation carrier population. *JAMA*. 2004; 291: 460-469.
- [5] Rodriguez-Revenga L, Madrigal I, Pagonabarraga J, *et al.* Penetrance of FMR1 premutation associated pathologies in fragile X syndrome families. *Eur. J. Hum. Genet*. 2009; 17: 1359-1362.
- [6] Rousseau F, Rouillard P, Morel ML *et al.* Prevalence of carriers of premutation-size alleles of the FMRI gene--and implications for the population genetics of the fragile X syndrome. *Am. J. Hum. Genet*. 1995; 57: 1006-1018.
- [7] Dombrowski C, Lévesque S, Morel ML, *et al.* Premutation and intermediate-size FMR1 alleles in 10572 males from the general population: loss of an AGG interruption is a late

- event in the generation of fragile X syndrome alleles. *Hum. Mol. Genet.* 2002; 11: 371-378.
- [8] Rifé M, Badenas C, Mallolas J, *et al.* Incidence of fragile X in 5,000 consecutive newborn males. *Genet Test.* 2003; 7: 339-343.
- [9] Hall DA, O'Keefe JA. Clinical neurogenetics: fragile x-associated tremor/ataxia syndrome. *Neurol. Clin.* 2013; 31: 1073-1084.
- [10] Bacalman S, Farzin F, Bourgeois JA, *et al.* Psychiatric phenotype of the fragile X-associated tremor/ataxia syndrome (FXTAS) in males: newly described fronto-subcortical dementia. *J. Clin. Psychiatry.* 2006; 67: 87-94.
- [11] Bourgeois JA, Cogswell JB, Hessel D, *et al.* Cognitive, anxiety and mood disorders in the fragile X-associated tremor/ataxia syndrome. *Gen. Hosp. Psychiatry.* 2007; 29: 349-56.
- [12] Bourgeois JA, Seritan AL, Casillas EM, *et al.* Lifetime prevalence of mood and anxiety disorders in fragile X premutation carriers. *J. Clin. Psychiatry.* 2011; 72: 175-82.
- [13] Adams PE, Adams JS, Nguyen DV, *et al.* Psychological symptoms correlate with reduced hippocampal volume in fragile X premutation carriers. *Am. J. Med. Genet. B Neuropsychiatr Genet.* 2010; 153B: 775-785.
- [14] Hashimoto R1, Backer KC, Tassone F, *et al.* An fMRI study of the prefrontal activity during the performance of a working memory task in premutation carriers of the fragile X mental retardation 1 gene with and without fragile X-associated tremor/ataxia syndrome (FXTAS). *J. Psychiatr. Res.* 2011; 45: 36-43.
- [15] Hessel D, Tassone F, Loesch DZ, *et al.* Abnormal elevation of FMR1 mRNA is associated with psychological symptoms in individuals with the fragile X premutation. *Am. J. Med. Genet. B Neuropsychiatr Genet.* 2005; 139B: 115-121.
- [16] Birch RC, Cornish KM, Hocking DR, Trollor JN. Understanding the Neuropsychiatric Phenotype of Fragile X-Associated Tremor Ataxia Syndrome: a Systematic Review. *Neuropsychol .Rev.* 2014.
- [17] Brega AG, Goodrich G, Bennett RE, *et al.* The primary cognitive deficit among males with fragile X-associated tremor/ataxia syndrome (FXTAS) is a dysexecutive syndrome. *Clin. Exp. Neuropsychol.* 2008; 30: 853-869.
- [18] Jacquemont S, Hagerman RJ, Leehey M, *et al.* Fragile X premutation tremor/ataxia syndrome: molecular, clinical, and neuroimaging correlates. *Am. J. Hum. Genet.* 2003; 72: 869-878.
- [19] Brunberg JA, Jacquemont S, Hagerman RJ, *et al.* Fragile X premutation carriers: characteristic MR imaging findings of adult male patients with progressive cerebellar and cognitive dysfunction. *AJNR Am. J. Neuroradiol.* 2002; 23: 1757-1766.
- [20] Greco CM, Hagerman RJ, Tassone F, *et al.* Neuronal intranuclear inclusions in a new cerebellar tremor/ataxia syndrome among fragile X carriers. *Brain.* 2002; 125: 1760-1771.
- [21] Greco CM, Berman RF, Martin RM, *et al.* Neuropathology of fragile X-associated tremor/ataxia syndrome (FXTAS). *Brain.* 2006; 129: 243-255.
- [22] Cohen S, Masyn K, Adams J, *et al.* Molecular and imaging correlates of the fragile X-associated tremor/ataxia syndrome. *Neurology.* 2006; 67: 1426-1431.
- [23] Adams JS, Adams PE, Nguyen D, *et al.* Volumetric brain changes in females with fragile X-associated tremor/ataxia syndrome (FXTAS). *Neurology.* 2007; 69: 851-859.
- [24] Okamoto K, Tokiguchi S, Furusawa T, *et al.* MR features of diseases involving bilateral middle cerebellar peduncles. *AJNR Am. J. Neuroradiol.* 2003; 24: 1946-1954.

- [25] Rodriguez-Revenga L, Gómez-Ansón B, Granell Moreno E, *et al.* Neuroimaging in Fragile X-Associated, *Neuroimaging-Clinical Applications*, Prof. Peter Bright (Ed.), 2012.
- [26] Bergink S, Severijnen LA, Wijgers N, *et al.* The DNA repair-ubiquitin-associated HR23 proteins are constituents of neuronal inclusions in specific neurodegenerative disorders without hampering DNA repair. *Neurobiol. Dis.* 2006; 23: 708-716.
- [27] Iwahashi CK, Yasui DH, An HJ, *et al.* Protein composition of the intranuclear inclusions of FXTAS. *Brain.* 2006; 129: 256-271.
- [28] Hunsaker MR, Greco CM, Spath MA, *et al.* Widespread non-central nervous system organ pathology in fragile X premutation carriers with fragile X-associated tremor/ataxia syndrome and CGG knock-in mice. *Acta Neuropathol.* 2011; 122: 467-479.
- [29] Tassone F, Iwahashi C, Hagerman PJ. FMR1 RNA within the intranuclear inclusions of fragile X-associated tremor/ataxia syndrome (FXTAS). *RNA Biol.* 2004; 1: 103-105.
- [30] Kenneson A, Zhang F, Hagedorn CH, Warren ST. Reduced FMRP and increased FMR1 transcription is proportionally associated with CGG repeat number in intermediate-length and premutation carriers. *Hum. Mol. Genet.* 2001; 10: 1449-1454.
- [31] Oostra BA, Willemsen R. A fragile balance: FMR1 expression levels. *Hum. Mol. Genet.* 2003; 12: R249-R257.
- [32] Tassone F, Hagerman RJ, Taylor AK, *et al.* Elevated levels of FMR1 mRNA in carrier males: a new mechanism of involvement in the fragile-X syndrome. *Am. J. Hum. Genet.* 2000; 66: 6-15.
- [33] Tassone F, Hagerman RJ, Taylor AK, *et al.* Clinical involvement and protein expression in individuals with the FMR1 premutation. *Am. J. Med Genet.* 2000; 91: 144-152.
- [34] Galloway JN, Nelson DL. Evidence for RNA-mediated toxicity in the fragile X-associated tremor/ataxia syndrome. *Future Neurol.* 2009; 4: 785.
- [35] Tassone F, Hagerman PJ. Expression of the FMR1 gene. *Cytogenet Genome Res.* 2003; 100: 124-128.
- [36] Tan H, Li H, Jin P. RNA-mediated pathogenesis in fragile X-associated disorders. *Neurosci. Lett.* 2009; 466: 103-108.
- [37] Jin P, Zarnescu DC, Zhang F, *et al.* RNA-mediated neurodegeneration caused by the fragile X premutation rCGG repeats in *Drosophila*. *Neuron.* 2003; 39: 739-747.
- [38] Arocena DG, Iwahashi CK, Won N, *et al.* Induction of inclusion formation and disruption of lamin A/C structure by premutation CGG-repeat RNA in human cultured neural cells. *Hum. Mol. Genet.* 2005; 14: 3661-3671.
- [39] Hashem V, Galloway JN, Mori M, *et al.* Ectopic expression of CGG containing mRNA is neurotoxic in mammals. *Hum. Mol. Genet.* 2009; 18: 2443-2451.
- [40] Hukema RK, Buijsen RAM, RaskeC *et al.* Induced expression of expanded CGG RNA causes mitochondrial dysfunction in vivo. *Cell Cycle.* 2014; 16: 2600-2608.
- [41] Li Y, Jin P. RNA-mediated neurodegeneration in fragile X-associated tremor/ataxia syndrome. *Brain Res.* 2012; 1462: 112-117.
- [42] Hagerman PJ. Current Gaps in Understanding the Molecular Basis of FXTAS. *Tremor Other Hyperkinet Mov* (N Y). 2012; 2.
- [43] Sellier C, Rau F, Liu Y, *et al.* Sam68 sequestration and partial loss of function are associated with splicing alterations in FXTAS patients. *EMBO J.* 2010; 29: 1248-1261.
- [44] Lukong KE, Richard S. Motor coordination defects in mice deficient for the Sam68 RNA-binding protein. *Behav. Brain Res.* 2008; 189: 357-363.

- [45] Jin P, Duan R, Qurashi A, *et al.* Pur alpha binds to rCGG repeats and modulates repeat-mediated neurodegeneration in a *Drosophila* model of fragile X tremor/ataxia syndrome. *Neuron*. 2007; 55: 556-564.
- [46] Sofola OA, Jin P, Qin Y *et al.* RNA-binding proteins hnRNP A2/B1 and CUGBP1 suppress fragile X CGG premutation repeat-induced neurodegeneration in a *Drosophila* model of FXTAS. *Neuron*. 2007; 55: 565-571.
- [47] Muslimov IA, Patel MV, Rose A, Tiedge H. Spatial code recognition in neuronal RNA targeting: role of RNA-hnRNP A2 interactions. *J. Cell Biol.* 2011; 194: 441-457.
- [48] Sellier C, Freyermuth F, Tabet R, *et al.* Sequestration of DROSHA and DGCR8 by expanded CGG RNA repeats alters microRNA processing in fragile X-associated tremor/ataxia syndrome. *Cell Rep.* 2013; 3: 869-880.
- [49] Napoli E, Ross-Inta C, Wong S, *et al.* Altered zinc transport disrupts mitochondrial protein processing/import in fragile X-associated tremor/ataxia syndrome. *Hum. Mol. Genet.* 2011; 20: 3079-3092.
- [50] Ross-Inta C, Omanska-Klusek A, Wong S, *et al.* Evidence of mitochondrial dysfunction in fragile X-associated tremor/ataxia syndrome. *Biochem. J.* 2010; 429: 545-552.
- [51] Cao Z, Hulsizer S, Tassone F, *et al.* Clustered burst firing in FMR1 premutation hippocampal neurons: amelioration with allopregnanolone. *Hum. Mol. Genet.* 2012; 21: 2923-2935.
- [52] Garcia-Arocena D, Yang JE, Brouwer JR, *et al.* Fibroblast phenotype in male carriers of FMR1 premutation alleles. *Hum. Mol. Genet.* 2010; 19: 299-312.
- [53] Jenkins EC, Tassone F, Ye L, *et al.* Reduced telomere length in older men with premutation alleles of the fragile X mental retardation 1 gene. *Am. J. Med. Genet. A.* 2008; 146A: 1543-1546.
- [54] Hagerman R, Hagerman P. Advances in clinical and molecular understanding of the FMR1 premutation and fragile X-associated tremor/ataxia syndrome. *Lancet Neurol.* 2013; 12: 786-798.
- [55] Chen Y, Tassone F, Berman RF, *et al.* Murine hippocampal neurons expressing Fmr1 gene premutations show early developmental deficits and late degeneration. *Hum. Mol. Genet.* 2010; 19: 196-208.
- [56] Cunningham CL, Martínez Cerdeño V, Navarro Porras E, *et al.* Premutation CGG-repeat expansion of the Fmr1 gene impairs mouse neocortical development. *Hum. Mol. Genet.* 2011; 20: 64-79.
- [57] Qin M, Entezam A, Usdin K, *et al.* A mouse model of the fragile X premutation: effects on behavior, dendrite morphology, and regional rates of cerebral protein synthesis. *Neurobiol Dis.* 2011; 42: 85-98.
- [58] Sellier C, Usdin K, Pastori C, *et al.* The multiple molecular facets of fragile X-associated tremor/ataxia syndrome. *J. Neurodev. Disord.* 2014; 6: 23.
- [59] Todd PK, Oh SY, Krans A, *et al.* CGG repeat-associated translation mediates neurodegeneration in fragile X tremor ataxia syndrome. *Neuron*. 2013; 78: 440-455.
- [60] Ladd PD, Smith LE, Rabaia NA, *et al.* An antisense transcript spanning the CGG repeat region of FMR1 is upregulated in premutation carriers but silenced in full mutation individuals. *Hum. Mol. Genet.* 2007; 16: 3174-3187.
- [61] Khalil AM, Faghihi MA, Modarresi F, *et al.* A novel RNA transcript with antiapoptotic function is silenced in fragile X syndrome. *PLoS One*. 2008; 3: e1486.

- [62] Pastori C, Peschansky VJ, Barbouth D, *et al.* Comprehensive analysis of the transcriptional landscape of the human FMR1 gene reveals two new long noncoding RNAs differentially expressed in Fragile X syndrome and Fragile X-associated tremor/ataxia syndrome. *Hum. Genet.* 2014; 133: 59-67.

Chapter 90

PSYCHIATRIC ASPECTS IN FRAGILE X SYNDROME AND RELATED PHENOTYPES

E. Mur¹ and M. Milà^{2-4,*}

¹CSM Martí i Julià. Instituto de Neuropsiquiatria y Adicciones (INAD),
Parc de Salut Mar. Barcelona, Spain

²Servicio de Bioquímica i Genética Molecular,
Hospital Clínic, Barcelona. Spain

³CIBER de Enfermedades Raras (CIBERER) Barcelona, Spain

⁴IDIBAPS (Institut de Investigacions
Biomèdiques August Pi I Sunyer) Barcelona, Spain

ABSTRACT

This chapter describes the psychopathological alterations of the different phenotypes associated with the *FMR1* gene. Fragile X patients present enormous functional impairment with a behavioral phenotype of hyperactivity and impaired attention, marked anxiety, with poor eye contact, affective lability, aggression, self-injurious behavior and autistic features. Since its first description the basic formulation of the fragile X behavioral phenotype has remained intact to the present day, with substantial confirmation of these basic findings in subsequent studies. Autism is one of the most recognized and severe behavioral abnormalities observed in males with fragile X syndrome (FXS) and this syndrome is the leading known monogenic cause of autism, accounting for approximately 5% of autism cases. Approximately 30–50% of FXS individuals meet full Diagnostic and Statistical Manual of Mental Disorders DSM-IV-TR criteria for autism with 60–74% fulfilling criteria for an autism spectrum disorder (ASD). Attention deficit and hyperactivity disorder (ADHD) is the most common diagnosable condition in FXS patients, with most males meeting formal criteria at some point in their lives. In children with FXS, ADHD is reported to be characterized by more inattentiveness, restlessness, fidgetiness and impulsivity. However, affective symptoms can be severe and disruptive, and are a common target for psychopharmacologic intervention. Several clinical subgroups present a higher risk for presenting anxiety outcomes, including children with FXS. Males with FXS display a broad range of anxiety symptoms, but these symptoms often do not fit into the established categories of major anxiety disorders employed by the DSM.

Regarding individuals carrying a premutation there is a growing body of literature suggesting that neuropsychiatric features of fragile X-associated tremor/ataxia syndrome (FXTAS) follow a fronto-subcortical pattern with primary impairments in executive function and increased vulnerability to mood and anxiety disorders. Increased rates of psychiatric symptoms may represent early markers of neurodegenerative diseases and have also been reported among *FMR1* premutation carriers without FXTAS. On the other hand, several studies have reported an excess of intermediate *FMR1* alleles in patients with cognitive and/or behavioral phenotypes. Numerous studies have investigated neuropsychological phenotypes among premutation allele carriers in women, but no definitive profile has been achieved. An increased risk of anxiety and mood disorders among premutation allele carriers has not been established, although they seem to be more common in these subjects than in controls.

Keywords: ASD, *FMR1*, hyperactive disorders, mood disorders, anxiety disorders

INTRODUCTION

The fragile X Syndrome (FXS) is associated with a complex but relatively consistent psychiatric phenotype and provides the enticing possibility of understanding a complex neuropsychiatric disorder at the synaptic, molecular, and genetic levels. This chapter describes the psychopathological alterations of the different phenotypes associated with the different status of the *FMR1* gene: FXS (>200 CGG), premutated individuals (55-200 CGG) and intermediate alleles (45-54 CGG). In 1984, Fryns [1] studied a large population of fragile X males, describing a behavioral phenotype of hyperactivity and impaired attention, marked anxiety with poor eye contact, affective lability, aggression, self-injurious behavior (especially the characteristic hand biting), and autistic features reported as repetitive, perseverative and stereotypic behaviors. This basic formulation of the fragile X behavioral phenotype has remained intact to the present day, with substantial confirmation of these basic findings in subsequent studies.

FRAGILE X SYNDROME

Fragile X patients suffer maladaptive behaviors and emotional disturbance with an enormous functional impairment; symptoms are a frequent reason for families to seek treatment and can lead to institutionalization in more severe cases.

Autism Disorders

Autistic disorder is the most debilitating subgroup of a larger category known as pervasive developmental disorders (American Psychiatric Association) characterized by impairment in social interaction and verbal and non-verbal communication, and restricted, repetitive and stereotypic patterns of behavior, interests and activities. Although there is considerable variability in individual symptoms, core deficits in social communication and restricted and repetitive behaviors are hallmarks of the disorder [2]. FXS is the leading known monogenic

cause of autism, accounting for approximately 5% of autism cases [reviewed in 3] and autism is one of the most recognized and severe behavioral abnormalities observed in males with FXS [4-8]. In individuals with the full mutation, approximately 30–50% meet full DSM-IV-TR criteria for autism with 60–74% fulfilling criteria for an autism spectrum disorder (ASD) [9-13]. Over 90% of individuals with FXS display some form of atypical behavior characteristic of autism, including social interaction (e.g., avoidance of eye contact, social withdrawal, and social anxiety) and repetitive and stereotyped behaviors [14]. Co-morbid FXS and autism are indicative of worse developmental outcomes [15-17], and greater impairment in cognition and adaptive behavior skills are more severe aberrant behavior than FXS without autism [18]. It has been suggested that autistic behaviors increase slowly but significantly over time, as do associated social avoidance behaviors [19].

The main dilemma arises when considering the pathogenesis of clinical autism spectrum in patients diagnosed with the FXS. Some authors consider that the clinical autism spectrum in these patients is a consequence of pathophysiological alterations resulting from the mutation of the *FMR1* gene, while other authors consider that this clinical autism spectrum in these patients has an etiopathological mechanism similar to idiopathic autism. These different points of view are generated due to the fact that diagnostic criteria for autism spectrum disorders are purely clinical.

Likewise, Harris (2011) [20] has proposed a focus on brain-behavior relationships targeting advancement of behavioral phenotyping in neurogenetic disorders precluding the application of DSM-IV diagnostic behavioral criteria to identify disorders such as FXS. He proposed that FXS is a neural model and phenocopy of autism and should not be considered a genetic model for autism. On the other hand, a number of investigators have reported findings indicating highly similar profiles between individuals with FXS and autism versus idiopathic autism and apply categorical or dimensional ratings of autism in FXS [9,16,21-23]. These authors agree with the concept that the autism phenomenon represents a range of behaviors, and perhaps also of other neurologic features [24]. While there is clear consensus regarding the shared phenomenology between FXS and idiopathic autism, there is great debate regarding diagnostic issues. Two of the primary debates on FXS center around questions of whether autism in FXS represents a continuum, with only those most severely affected meeting criteria for autism, and whether autism in FXS is the same as or different from idiopathic autism [9]. Although the literature on the co-morbidity of FXS and autism is extensive, few published studies have longitudinally examined early indicators of autism in infants. Study of the emerging characteristics in FXS is critical for understanding if early features in infants with FXS are associated with later autistic behaviors as reported in idiopathic autism. To date, the studies that have been conducted lend evidence to the fact that indicators of autism are present as early as 12 months of age in males with FXS, and replicate findings in idiopathic autism that implicate difficulties, with disengagement of attention and shifts in behavior at 6 to 12 months of age in the later development of autism [25-27]. Clearly, brain-behavior relationships in idiopathic autism and FXS are complex, and there is currently insufficient evidence to resolve these debates. In summary, however, despite all the controversies generated in the diagnosis of clinical ASD, at present the *FMR1* gene is considered the most common monogenic cause of ASD, and therefore, the molecular diagnosis of FXS should be ruled out in all ASD [28].

Hyperactivity Disorders

Attention deficit and hyperactivity (ADHD) is the most common diagnosable condition in FXS patients, with most males meeting formal criteria at some point in their lives. This condition is typically not stable over time in any given individual [1]. Using DSM-V, several of the ADHD symptoms must be present in individuals prior to the age of 12 years, compared to 7 years as the age of onset in DSM-IV. This change is supported by substantial research published since 1994 that found no clinical differences between children identified at 7 years versus later in terms of the course, severity, outcome, or treatment response. DSM-5 includes no exclusion criteria for people with ASD, since symptoms of both disorders co-occur. The higher inattentiveness, restlessness, fidgetiness and impulsivity in ADHD in children with FXS is suggestive of the ADHD inattentive sub-type compared to ADHD in the general population; these features do not necessarily improve with age [29]. The signature of the FXS is strong unsustained attention, but poor capacity to switch between tasks and weaker inhibitory control [reviewed in 30]. Very young children with FXS are often noted to be physically hypoactive, with somewhat impaired attention. Preschool children can display dramatic increases in activity levels, leading to markedly disruptive behavior. As children grow, hyperactivity declines with increasing body mass, while problems with attention continue throughout life. This can be seen as similar to the course of ADHD in the normal population, though the degree of hyperactivity in FXS is impressive. There is also evidence that the attention deficit seen in males with FXS has a specific profile [31], which is distinct from other causes of developmental disorders, suggesting that the attention problems seen in the course of FXS may represent more than nonspecific immaturity. There is evidence that ADHD symptoms in FXS respond to stimulants [32,33].

Mood Disorders

There has been considerable controversy regarding the behavior phenotype of FXS and the nature of behavioral and emotional problems associated with it.

As described by Backes and colleagues (2000) [34], males with FXS rarely meet formal criteria for a diagnosis of a major mood disorder as defined in DSM-V. Diagnoses such as major depression or bipolar disorder require periods of abnormal mood that are sustained, whereas individuals with FXS typically exhibit labile mood, irritability, self-injurious behavior, and aggressive outbursts of a more fleeting and episodic nature, not meeting the conventional duration criteria. These episodes are typically adaptive, related to environmental stressors and are less frequent in familial or more structured settings. However, affective symptoms can be severe and disruptive, and psychopharmacologic intervention is needed.

Selective serotonin reuptake inhibitors are a commonly employed treatment strategy for affective symptoms, along with other antidepressants, anticonvulsants, and atypical antipsychotics in more severe cases [33]. There have been no clinical trials of any size, open or controlled, of antidepressants or anticonvulsants for the treatment of affective symptoms of FXS. Curiously, the study of valproic acid by Torrioli and collaborators (2010) [35] focused exclusively on ADHD symptoms in boys and they considered that this treatment could be considered as an alternative to treating symptoms with stimulants the efficacy of which needs

to be confirmed by further studies in patients with FXS. Behavioral problems in FXS improve from childhood to young adulthood [36].

Anxiety Disorders

Children with neurodevelopmental disorders, such as FXS, have a higher risk of presenting anxiety. Patients with FXS display a broad range of anxiety symptoms, but these symptoms often do not fit into the established categories of major anxiety disorders employed by the DSM. Cordeiro and coworkers (2011) [37] examined the prevalence of anxiety disorders in a FXS population using DSM-IV-TR criteria. They found that 83% of participants (n=97, ages 5.5–33.3 years) met criteria for any anxiety disorder, and 58% exhibited multiple anxiety disorders. In addition to the high proportion of anxiety disorders in FXS, an important part of individuals with FXS display autistic symptoms as mentioned previously.

Symptoms of anxiety and autism overlap and interrelate within FXS, although these disorders have been distinguished through behavioral and physiological profiles. Kaufmann, Budimirovic and colleagues have published a series of papers characterizing anxiety and autism as distinct social interaction disorders in FXS [38-39]. These authors propose that social anxiety emerges from a combination of lower non-verbal abilities and moderate social withdrawal, whereas autism (either alone or in conjunction with social anxiety) is characterized by a more complex constellation of severe social withdrawal and lower adaptive socialization or verbal skills.

Negative affect is one of the most commonly studied predictors of problem behaviors in non-clinical [40-41] and clinical [42] populations. Negative affect is composed of several specific dimensions of temperament; including fear, approach, soothability, sadness, anger, discomfort, and motor activity [43] and predicts anxiety in preschool boys with FXS [44]. Multilevel models indicate associations between elevated anxiety and higher fear and sadness, lower soothability, and steeper longitudinal increases in approaches in the FXS population.

A minority of males with FXS fulfill formal criteria for the diagnosis of obsessive-compulsive disorder (OCD), while “compulsive symptoms” have been noted in several studies in a large majority of subjects with FXS. In most cases of FXS, individuals exhibit symptoms strongly reminiscent of obsessions and compulsions, but which do not meet the precise psychiatric definitions for these symptoms. Often, pleasure is derived from repetitive and “compulsive” behaviors, in contrast to the ego-dystonic nature of true obsessions and compulsions. Hoarding, counting, and the need for symmetry are all typical symptoms of OCD frequently seen in FXS. Similarly, younger children with FXS meet the criteria for separation anxiety disorder in a small minority of cases [34], while symptoms of separation anxiety, social phobia, panic, and agoraphobia are seen clinically at a much higher rate.

The psychiatric drugs currently available can provide significant symptomatic relief of the hyperactivity, anxiety disorders, and affective disturbances often seen in the course of FXS. However, patients with this syndrome may be especially susceptible to the psychiatric side effects of these medications, requiring particular care in their prescription.

PREMUTATION CARRIERS

For many years, premutation carriers of the *FMR1* gene were considered asymptomatic, but various clinical disorders have been associated in both men and women, being Fragile X tremor–ataxia syndrome (FXTAS) one of the most severe. Individuals with carrier premutation are those that have alleles between 55-200 CGG. Some of the clinical features have been related to the CGG number; in general, more length in the CGG track implies more severe clinical manifestations.

There is a growing body of evidence suggesting that *FMR1* premutation carriers may have increased vulnerability of presenting psychiatric disorders [45]. However, the presence of neuropsychological and behavioral impairment among *FMR1* premutation carriers remains controversial, as these features were initially thought to be associated with the stress of raising children with FXS [reviewed in 46].

Altered neurobehavioral profiles including variation of phenotypes associated with mood and anxiety may be expected among younger premutation carriers.

An increased risk of anxiety and mood disorders among premutation carriers has not been established. Some studies have reported a lack of phenotype [47-48], while others have described repeat length associations with psychiatric symptoms [49-50]. Hessel and coworkers (2005)[51] found that the *FMR1* transcript level, but not repeat length or Fragile X Mental retardation Protein (FMRP) levels, was significantly associated with increased severity of psychiatric symptoms in males, independently of FXTAS status. These results suggest that premutation carriers may be at risk of presenting emotional morbidity; however, phenotypic differences were subtle and of a small CGG effect size.

The largest and most recent study of life-time mood and anxiety in the premutation population was completed by Bourgeois and colleagues (2009) [52]. In this study, the prevalence of anxiety disorders in carriers with and without FXTAS was compared with a very large age-matched national dataset. In terms of all anxiety disorders, only those with FXTAS demonstrated a higher prevalence. Upon separation, this was similarly true for panic disorder, post-traumatic stress disorder and specific phobia. Generalized anxiety disorder and OCD failed to demonstrate any difference between carriers and controls. Only social phobia was found to have higher levels in premutation carriers without FXTAS compared to controls. Chronic anxiety has also been associated with radiological signs on MRI; specifically, the higher the anxiety score the smaller the size of the hippocampus in women with the premutation [53].

Rodriguez-Revenga and collaborators (2008) [54] examined psychiatric and depressive symptoms in 34 *FMR1* premutation carrier mothers of children with FXS in comparison with two control groups (39 mothers with a non-FXS intellectual disability child and 39 mothers from the general population). Both groups of mothers with a child with intellectual disability showed greater susceptibility to psychological problems than the control group without a mentally retarded child, but *FMR1* premutated mothers evidenced a higher tendency to depression. These results suggest that, despite the stress of caring for a child with mental retardation, the premutation by itself could be responsible for some psychiatric traits.

In a screening study of individuals from families with FXS, roughly 14% of boys and 5% of girls with the premutation were found to also have an ASD [10]. Even among those carriers not diagnosed with ASD, related psychological traits are more common among carriers

compared to controls without the premutation. A recent study examined a broad range of pragmatic language skills as well as related behavioral features of the broad autism phenotype among women with the premutation compared with mothers of children with autism, and mothers of typically developing children with no family history of FXS, autism, or language impairment [55]. In this study, conversational samples from a semistructured videotaped interview were used to assess pragmatic language using the Pragmatic Rating Scale (PRS) [56]. This study replicated previous findings in the autism parent group and also showed that women with the premutation exhibited similarly elevated rates of pragmatic language problems relative to the controls. The presence of broad autism phenotype traits was associated with greater expression of autism symptoms in their children with FXS. Other studies have also found increased rates of both social aloofness [49] and a rigid perfectionism [57] among carrier women.

Given its relative rarity in the general population, psychosis has been challenging to study in premutation carriers. Initial linkage analysis failed to show a clear relationship of schizophrenia to the *FMR1* gene [58]. Prevalence studies have found the overall rate of psychotic disorders to be low [49]. There have, however, been several case reports of combined psychotic illnesses and the premutation, including schizoaffective disorder [59] and with combined schizophrenia and schizoid personality disorder [60]. Interestingly, as opposed to frank psychotic disorders, multiple studies have found an increased prevalence of schizotypal personality traits in the carrier population [49,61].

Attention regulation difficulties have been proposed to be a problem in people with the premutation. Notably, when compared with their control siblings, premutation carriers had significantly more issues related to attention than their noncarrier siblings [62]. Inattention and impulsivity amongst *FMR1* carriers can be problematic through adulthood [63], although hyperactivity was not noted to be increased in prevalence.

Dysthymia and bipolar disorder have generally failed to demonstrate significant levels in carriers compared to controls [52].

FRAGILE X ASSOCIATED TREMOR ATAXIA SYNDROME (FXTAS)

FXTAS is an X-linked neurodegenerative disorder affecting up to 45.5% of males and 16.5% of females carrying a premutation in the *FMR1* gene [64-65].

There is a growing body of literature suggesting that the neuropsychiatric features of FXTAS follow a fronto-subcortical pattern with primary impairments in executive function and increased vulnerability to mood and anxiety disorders [51,53, 66-69]

Increased rates of psychiatric symptoms may represent early markers of neurodegenerative diseases such as Parkinsons Disease (PD) [70] and Huntington's Disease [71], which have also been reported among *FMR1* premutation carriers without FXTAS, including obsessive-compulsive symptoms [51], social phobia [72], depression [50], schizotypal features [73] and abnormalities in social cognition [74].

Three studies have examined the psychiatric features of FXTAS: (1) Bacalman et al. (2006) [67]; (2) Adams et al. (2010) [53]; and (3) Hashimoto et al. (2011) [69]. They showed no significant association between FXTAS and different psychiatric diagnoses. The studies compared the prevalence of neuropsychiatric symptoms between individuals with FXTAS and

matched controls with normal *FMR1* alleles (cohort 1 and 3), and between asymptomatic *FMR1* premutation carriers in cohort 2.

Symptoms of depression were significantly elevated among males with FXTAS when measured using the informant-rated NPI in cohort 1; however, self-reported symptoms measured using the SCL-90-R in cohort 3 were comparable to controls. Women with FXTAS (cohort 2) tended to report higher rates of depression than controls with normal *FMR1* alleles; however, this difference did not withstand correction for multiple comparisons.

Anxiety was significantly elevated among older males with FXTAS compared to controls in cohort 2; however, in cohort 1 the group difference (50% FXTAS group vs. 0% control group) was not significant, possibly due to the small sample size ($n=14$). Male carriers with FXTAS exhibited elevated rates of anxiety in cohort 3 (Cohen's $d=0.69$) and comparable scores in cohort 2. Females with FXTAS (cohort 2) exhibited higher rates of anxiety and obsessive-compulsive symptoms compared to controls with normal *FMR1* alleles and asymptomatic *FMR1* premutation carriers.

Males with FXTAS (cohorts 2 and 3) reported significantly higher rates of obsessive compulsive symptoms compared to asymptomatic carriers (Cohen's $d=0.92$, cohort 7), but not controls (cohort 2). Symptoms of anxiety among asymptomatic male and female *FMR1* premutation carriers were compared to controls with normal alleles in cohort 2 and no group differences were detected.

Additional informant-rated behavioral disturbances found to be significantly elevated among males with FXTAS (cohort 1) compared to controls included apathy, irritability, disinhibition, and agitation/aggression (Cohen's $d=3.53$) [67].

Increased obsessive-compulsive and depressive symptoms (but not anxiety) were also associated with decreased left amygdala volume among males with FXTAS in cohort 3 (all $ps<0.02$) [69].

It is difficult to determine the psychiatric features most commonly associated with FXTAS based on this limited number of studies; however, the evidence available suggests that FXTAS may be associated with increased rates of depression, anxiety and obsessive-compulsive symptoms. This is supported by a previous systematic review that included case reports and case series, suggesting elevated rates of mood and anxiety disorders among those with FXTAS [52].

Emerging evidence suggests that a subset of carriers without FXTAS may exhibit cognitive deficits, psychiatric symptoms, and changes in brain structure and function which may be modulated by *FMR1* gene expansion [51,53,69,74-75]. Specifically, abnormalities in structural and functional connectivity have been described in the middle cerebellar peduncles, hippocampus, amygdala and prefrontal cortex, raising the possibility that these early indicators of fronto-subcortical and cortico-cerebellar dysfunction may be detectable prior to the onset of clinical symptoms. The capacity to identify the subjects most at-risk or closest to onset of illness would be invaluable for recruitment into clinical trials as treatments for FXTAS become available. In the absence of longitudinal follow-up of these cohorts, however, it is unclear whether cognitive, psychiatric and radiological changes observed in some asymptomatic carriers represent an independent developmental phenotype associated with the *FMR1* premutation, or serve as early indicators of later progression to FXTAS.

Intermediate *FMR1* alleles (alleles within the 45–54 CGG repeat range are described as an ‘intermediate or gray zone’). During the last few years, several studies have reported an excess of intermediate *FMR1* alleles in patients with cognitive and/or behavioral phenotypes [76-80]

while other studies do not support this relationship. Madrigal and colleagues (2011) [81] did not find evidence of an association between intermediate alleles and behavioral or cognitive phenotypes, suggesting that intermediate alleles are not clearly implicated in these pathologies.

A frequent concern in FXS screening is the genetic counseling to intermediate allele carriers. Currently, the follow-up of individuals with no AGG interruptions is indicated, despite the low risk of expansion in the next generation. Intermediate alleles should not be considered a risk factor for ID or behavioral phenotypes. Nonetheless, these alleles should be characterized to provide accurate genetic counseling [81].

Table 1. Percentage of affected individuals for the psychiatric aspects

Psychiatric features	FXS	PM with FXTAS	PM without FXTAS	References
Autism traits	>90%			[14]
Autism criteria DSM-IV	30-50%			[11-13]
Autism spectrum disorders	60-74%		14% Males 5% females	[11-12,57]
ADHD	30% Females 60% Males		30% Females 14% Males	[62]
Affective disorders	22% Females 12% Males	65%	55.7% 34.2%	[49,62,72]
Anxiety	83%	52%	35-40%	[37,49,72,]

CONCLUSION

FMR1 is the first monogenic cause of autism disorder. Approximately 30–50% of FXS meet full DSM-IV-TR criteria for autism, 60–74% fulfilling criteria for an ASD and over 90% of individuals display some form of atypical behavior characteristic of autism. Molecular diagnosis of FXS should be ruled out in all ASD. Increased rates of psychiatric symptoms may represent early markers of neurodegenerative diseases which have been reported among *FMR1* premutation carriers with and without FXTAS; including obsessive-compulsive symptoms, social phobia, depression, schizotypal features, and abnormalities in social cognition. Altered neurobehavioral profiles including a variation of phenotypes associated with mood and anxiety may be expected among younger premutation carriers. The trend is to present higher rates of anxiety and mood disorders in premutation carriers, but no study has found significant differences between carriers and the control population.

REFERENCES

- [1] Fryns JP. The fragile X syndrome. A study of 83 families. *Clin. Genet.* 1984;26(6):497–528.

- [2] Tager-Flusberg H, Rogers S, Cooper J, Landa R, Lord C, Paul R, Rice M, Stoel-Gammon C, Wetherby A, Yoder P. Defining spoken language benchmarks and selecting measures of expressive language development for young children with autism spectrum disorders. *J Speech Lang Hear Res.* 2009 Jun;52(3):643-52. Hagerman RJ, Hall DA, Coffey S, Leehey M, Bourgeois J, Gould J, Zhang L, Seritan A, Berry-Kravis E, Olichney J, Miller JW, Fong AL, Carpenter R, Bodine C, Gane LW, Rainin E, Hagerman H, Hagerman PJ. Treatment of fragile X-associated tremor ataxia syndrome (FXTAS) and related neurological problems. *Clin. Interv Aging.* 2008;3(2):251-62. Review.
- [3] Hagerman RJ, Jackson AW III, Levitas A, Rimland B, Barden M, 1986. An analysis of autism in fifty males with the fragile X syndrome. *Am. J. Med. Genet* 23:359-374.
- [4] Baumgardner TL, Reiss AL, Freund LS, Abrams MT. Specification of the neurobehavioral phenotype in males with fragile X syndrome. *Pediatrics* 1995;95(5):744-752.
- [5] Bailey DB Jr, Mesibov GB, Hatton DD, Clark RD, Roberts JE, Mayhew L. Autistic behavior in young boys with fragile X syndrome. *J. Autism Dev. Disord.* 1998 Dec;28(6):499-508.
- [6] Kaufman WE, Reiss AL. Molecular and cellular genetics of fragile X syndrome. *Am. J. Med. Genet.* 1999;88:11-24.
- [7] Hagerman RJ. The physical and behavioral phenotype. In Hagerman RJ, Hagerman PJ, editors. *Fragile X syndrome: Diagnosis, treatment, and research.* Baltimore, MD: The Johns Hopkins University Press. 2002;pp 3-109
- [8] Kaufmann WE, Cortell R, Kau AS, Bukelis I, Tierney E, Gray RM, Cox C, Capone GT, Stanard P. Autism spectrum disorder in fragile X syndrome: communication, social interaction, and specific behaviors. *Am. J. Med. Genet A.* 2004 Sep 1;129A(3):225-34.
- [9] Clifford S, Dissanayake C, Bui QM, Huggins R, Taylor AK, Loesch DZ. Autism spectrum phenotype in males and females with fragile X full mutation and premutation. *J. Autism Dev. Disord* 2007, 37:738-747.
- [10] Hall SS, Lightbody AA, Reiss AL. Compulsive, self-injurious, and autistic behavior in children and adolescents with fragile X syndrome. *Am. J. Ment. Retard.* 2008 Jan;113(1):44-53.
- [11] Harris SW, Hessel D, Goodlin-Jones B, Ferranti J, Bacalman S, Barbato I, Tassone F, Hagerman PJ, Herman H, Hagerman RJ. Autism profiles of males with fragile X syndrome. *Am. J. Ment. Retard.* 2008 Nov;113(6):427-38.
- [12] Bailey DB, Raspa M, Holiday D, Bishop E, Olmsted M. Functional skills of individuals with fragile x syndrome: a lifespan cross-sectional analysis. *Am. J. Intellect Dev. Disabil.* 2009 Jul;114(4):289-303.
- [13] Hernandez RN, Feinberg RL, Vaurio R, Passanante NM, Thompson RE, Kaufmann WE. Autism spectrum disorder in fragile X syndrome: a longitudinal evaluation. *Am. J. Med. Genet A.* 2009 Jun;149A(6):1125-37.
- [14] Bailey DB Jr, Hatton DD, Mesibov G, Ament N, Skinner M. Early development, temperament, and functional impairment in autism and fragile X syndrome. *J. Autism Dev. Disord.* 2000 Feb;30(1):49-59.
- [15] Rogers SJ, Wehner DE, Hagerman R. The behavioral phenotype in fragile X: symptoms of autism in very young children with fragile X syndrome, idiopathic autism, and other developmental disorders. *J. Dev. Behav. Pediatr.* 2001 Dec;22(6):409-17.

- [16] Kau AS, Tierney E, Bukelis I, Stump MH, Kates WR, Trescher WH, Kaufmann WE. Social behavior profile in young males with fragile X syndrome: characteristics and specificity. *Am. J. Med. Genet A*. 2004 Apr 1;126A(1):9-17.
- [17] Hatton DD, Sideris J, Skinner M, Mankowski J, Bailey DB Jr, Roberts J, Mirrett P. Autistic behavior in children with fragile X syndrome: prevalence, stability, and the impact of FMRP. *Am. J. Med. Genet A*. 2006 Sep 1;140A(17):1804-13.
- [18] Roberts JE, Weisenfeld LA, Hatton DD, Heath M, Kaufmann WE. Social approach and autistic behavior in children with fragile X syndrome. *J. Autism Dev. Disord.* 2007 Oct;37(9):1748-60. Epub 2006 Dec 19.
- [19] Harris JC. Brain and behavior in fragile x syndrome and idiopathic autism. *Arch. Gen. Psychiatry*. 2011 Mar;68(3):230-1.
- [20] Bailey DB Jr. Newborn screening for fragile X syndrome. *Ment. Retard Dev. Disabil Res. Rev.* 2004;10(1):3-10. Review.
- [21] Lewis P, Abbeduto L, Murphy M, Richmond E, Giles N, Bruno L, Schroeder S. Cognitive, language and social-cognitive skills of individuals with fragile X syndrome with and without autism. *J. Intellect Disabil Res.* 2006 Jul;50(Pt 7):532-45.
- [22] Hoeft F, Walter E, Lightbody AA, Hazlett HC, Chang C, Piven J, Reiss AL. Neuroanatomical differences in toddler boys with fragile x syndrome and idiopathic autism. *Arch. Gen. Psychiatry*. 2011 Mar;68(3):295-305. doi: 10.1001/archgenpsychiatry.2010.153. Epub 2010 Nov 1.
- [23] Lord C, Petkova E, Hus V, Gan W, Lu F, Martin DM, Ousley O, Guy L, Bernier R, Gerdtz J, Algermissen M, Whitaker A, Sutcliffe JS, Warren Z, Klin A, Saulnier C, Hanson E, Hundley R, Piggot J, Fombonne E, Steiman M, Miles J, Kanne SM, Goin-Kochel RP, Peters SU, Cook EH, Guter S, Tjernagel J, Green-Snyder LA, Bishop S, Esler A, Gotham K, Luyster R, Miller F, Olson J, Richler J, Risi S. A multisite study of the clinical diagnosis of different autism spectrum disorders. *Arch. Gen. Psychiatry*. 2012 Mar;69(3):306-13. Epub 2011 Nov 7.
- [24] Zwaigenbaum L, Bryson S, Rogers T, Roberts W, Brian J, Szatmari P. Behavioral manifestations of autism in the first year of life. *Int. J. Dev. Neurosci.* 2005 Apr-May;23(2-3):143-52.
- [25] Baranek GT, Roberts JE, David FJ, Sideris J, Mirrett PL, Hatton DD, Bailey DB Jr. Developmental trajectories and correlates of sensory processing in young boys with fragile X syndrome. *Phys. Occup. Ther Pediatr.* 2008;28(1):79-98.
- [26] Roberts TP, Cannon KM, Tavabi K, Blaskey L, Khan SY, Monroe JF, Qasmieh S, Levy SE, Edgar JC. Auditory magnetic mismatch field latency: a biomarker for language impairment in autism. *Biol. Psychiatry*. 2011 Aug 1;70(3):263-9. Epub 2011 Mar 9.
- [27] Shen Y, Dies KA, Holm IA, Bridgemohan C, Sobeih MM, Caronna EB, Miller KJ, Frazier JA, Silverstein I, Picker J, Weissman L, Raffalli P, Jeste S, Demmer LA, Peters HK, Brewster SJ, Kowalczyk SJ, Rosen-Sheidley B, McGowan C, Duda AW 3rd, Lincoln SA, Lowe KR, Schonwald A, Robbins M, Hisama F, Wolff R, Becker R, Nasir R, Urion DK, Milunsky JM, Rappaport L, Gusella JF, Walsh CA, Wu BL, Miller DT; Autism Consortium Clinical Genetics/DNA Diagnostics Collaboration. Clinical genetic testing for patients with autism spectrum disorders. *Pediatrics*. 2010 Apr;125(4):e727-35.
- [28] Turk J, Cornish K. Face recognition and emotion perception in boys with fragile-X syndrome. *J. Intellect Disabil. Res.* 1998 Dec;42 (Pt 6):490-9.

- [29] Cornish KM, Turk J, Wilding J, Sudhalter V, Munir F, Kooy F, Hagerman R. Annotation: Deconstructing the attention deficit in fragile X syndrome: a developmental neuropsychological approach. *J. Child Psychol. Psychiatry.* 2004 Sep;45(6):1042-53. Review.
- [30] Munir F, Cornish KM, Wilding J. A neuropsychological profile of attention deficits in young males with fragile X syndrome. *Neuropsychologia.* 2000;38(9):1261-70.
- [31] Berry-Kravis E, Potanos K. Psychopharmacology in fragile X syndrome-present and future. *Ment. Retard Dev. Disabil Res. Rev.* 2004;10(1):42-8.
- [32] Tsiouris JA, Brown WT. Neuropsychiatric symptoms of fragile X syndrome: pathophysiology and pharmacotherapy. *CNS Drugs.* 2004;18(11):687-703.
- [33] Backes M, Genç B, Schreck J, Doerfler W, Lehmkuhl G, von Gontard A. Cognitive and behavioral profile of fragile X boys: correlations to molecular data. *Am. J. Med. Genet.* 2000 Nov 13;95(2):150-6.
- [34] Torrioli M, Vernacotola S, Setini C, Bevilacqua F, Martinelli D, Snape M, Hutchison JA, Di Raimo FR, Tabolacci E, Neri G. Treatment with valproic acid ameliorates ADHD symptoms in fragile X syndrome boys. *Am. J. Med. Genet A.* 2010 Jun;152A(6):1420-7.
- [35] Einfeld S, Tonge B, Turner G. Longitudinal course of behavioral and emotional problems in fragile X syndrome. *Am. J. Med. Genet.* 1999 Dec 22;87(5):436-9.
- [36] Cordeiro L, Ballinger E, Hagerman R, Hessler D. Clinical assessment of DSM-IV anxiety disorders in fragile X syndrome: prevalence and characterization. *J. Neurodev Disord.* 2011 Mar;3(1):57-67.
- [37] Budimirovic DB, Bukelis I, Cox C, Gray RM, Tierney E, Kaufmann WE. Autism spectrum disorder in Fragile X syndrome: differential contribution of adaptive socialization and social withdrawal. *Am. J. Med. Genet A.* 2006 Sep 1;140A(17):1814-26.
- [38] Budimirovic DB, Kaufmann WE. What can we learn about autism from studying fragile X syndrome? *Dev. Neurosci.* 2011;33(5):379-94.. Review.
- [39] Calkins SD, Blandon AY, Williford AP, Keane SP. Biological, behavioral, and relational levels of resilience in the context of risk for early childhood behavior problems. *Dev. Psychopathol.* 2007 Summer;19(3):675-700.
- [40] Lemery KS, Essex MJ, Smider NA. Revealing the relation between temperament and behavior problem symptoms by eliminating measurement confounding: expert ratings and factor analyses. *Child Dev.* 2002 May-Jun;73(3):867-82.
- [41] Hutman T, Rozga A, DeLaurentis AD, Barnwell JM, Sugar CA, Sigman M. Response to distress in infants at risk for autism: a prospective longitudinal study. *J. Child Psychol. Psychiatry.* 2010 Sep;51(9):1010-20.
- [42] Putnam SP, Stifter CA. Behavioral approach-inhibition in toddlers: prediction from infancy, positive and negative affective components, and relations with behavior problems. *Child Dev.* 2005 Jan-Feb;76(1):212-26.
- [43] Tonnsen BL, Malone PS, Hatton DD, Roberts JE. Early negative affect predicts anxiety, not autism, in preschool boys with fragile X syndrome. *J. Abnorm Child Psychol.* 2013 Feb;41(2):267-80.
- [44] Hunter JE, Leslie M, Novak G, Hamilton D, Shubeck L, Charen K, Abramowitz A, Epstein MP, Lori A, Binder E, Cubells JF, Sherman SL. Depression and anxiety symptoms among women who carry the FMR1 premutation: impact of raising a child

- with fragile X syndrome is moderated by CRHR1 polymorphisms. *Am. J. Med. Genet B Neuropsychiatr Genet.* 2012. Epub 2012 May 9.
- [45] Hunter JE, Abramowitz A, Rusin M, Sherman SL. Is there evidence for neuropsychological and neurobehavioral phenotypes among adults without FXTAS who carry the FMR1 premutation? A review of current literature. *Genet Med.* 2009 Feb;11(2):79-89. Review.
- [46] Reiss AL, Freund L, Abrams MT, Boehm C, Kazazian H. Neurobehavioral effects of the fragile X premutation in adult women: a controlled study. *Am. J. Hum. Genet* 1993, 52:884–894.
- [47] Sobesky WE, Taylor AK, Pennington BF, Bennetto L, Porter D, Riddle J, Hagerman RJ. Molecular/clinical correlations in females with fragile X. *Am. J. Med. Genet.* 1996 Aug 9;64(2):340-5.
- [48] Franke P, Leboyer M, Gansicke M, Weiffenbach O, Biancalana V, Cornillet-Lefebvre P, Maier W. Genotype-phenotype relationship in female carriers of the premutation and full mutation of FMR-1. *Psychiatry Res* 1998, 80:113–127.
- [49] Johnston C, Eliez S, Dyer-Friedman J, Hessl D, Glaser B, Blasey C, Taylor A, Reiss A. Neurobehavioral phenotype in carriers of the fragile X premutation. *Am. J. Med. Genet.* 2001 Nov 1;103(4):314-9.
- [50] Hessl D, Tassone F, Loesch DZ, Berry-Kravis E, Leehey MA, Gane LW, Barbato I, Rice C, Gould E, Hall DA, Grigsby J, Wegelin JA, Harris S, Lewin F, Weinberg D, Hagerman PJ, Hagerman RJ. Abnormal elevation of FMR1 mRNA is associated with psychological symptoms in individuals with the fragile X premutation. *Am. J. Med. Genet B Neuropsychiatr Genet.* 2005 Nov 5;139B(1):115-21.
- [51] Bourgeois JA, Coffey SM, Rivera SM, Hessl D, Gane LW, Tassone F, Greco C, Finucane B, Nelson L, Berry-Kravis E, Grigsby J, Hagerman PJ, Hagerman RJ. A review of fragile X premutation disorders: Expanding the psychiatric perspective. *J. Clin. Psychiatry* 2009, 70:852–862.
- [52] Adams PE, Adams JS, Nguyen DV, Hessl D, Brunberg JA, Tassone F, Zhang W, Koldewyn K, Rivera SM, Grigsby J, Zhang L, Decarli C, Hagerman PJ, Hagerman RJ. Psychological symptoms correlate with reduced hippocampal volume in fragile X premutation carriers. *Am. J. Med. Genet B Neuropsychiatr Genet.* 2010 Apr 5;153B(3):775-85.
- [53] Rodriguez-Revenga L, Madrigal I, Alegret M, Santos M, Milà M. Evidence of depressive symptoms in fragile-X syndrome premutated females. *Psychiatr Genet.* 2008 Aug;18(4):153-5.
- [54] Losh M, Klusek J, Martin GE, Sideris J, Parlier M, Piven J. Defining genetically meaningful language and personality traits in relatives of individuals with fragile X syndrome and relatives of individuals with autism. *Am. J. Med. Genet B* 2012, 159B:660–668.
- [55] Landa R, Piven J, Wzorek MM, Gayle JO, Chase GA, Folstein SE: Social language use in parents of autistic individuals. *Psychol. Med.* 1992, 22:245–254
- [56] Hessl D, Rivera SM, Reiss AL: The neuroanatomy and neuroendocrinology of fragile X syndrome. *Ment. Retard Dev. D R* 2004, 10:17–24
- [57] Ashworth A, Abusaad I, Walsh C, Nanko S, Murray RM, Asherson P, Collier DA. Linkage analysis of the fragile X gene FMR-1 and schizophrenia: no evidence for linkage

- but report of a family with schizophrenia and an unstable triplet repeat. *Psychiatr Genet* 1996, 6:81–86.
- [58] Al-Semaan Y, Malla AK, Lazosky A: Schizoaffective disorder in a fragile-X carrier. *Aust. N Z J. Psychiatry* 1999, 33:436–440.
- [59] Khin NA, Tarleton J, Raghu B, Park SK: Clinical description of an adult male with psychosis who showed FMR1 gene methylation mosaicism. *Am. J. Med. Genet* 1998, 81:222–224.
- [60] Thompson NM, Rogeness GA, McClure E, Clayton R, Johnson C: Influence of depression on cognitive functioning in Fragile X females. *Psychiatry Res* 1996, 64:97–104.
- [61] Bailey DB Jr, Raspa M, Olmsted M, Holiday DB. Co-occurring conditions associated with FMR1 gene variations: findings from a national parent survey. *Am. J. Med. Genet A*. 2008 Aug 15;146A(16):2060-9.
- [62] Hunter JE, Allen EG, Abramowitz A, Rusin M, Leslie M, Novak G, Sherman SL. No evidence for a difference in neuropsychological profile among carriers and noncarriers of the FMR1 premutation in adults under the age of 50. *Am. J. Hum. Genet.* 2008, 83:692–702.
- [63] Hagerman RJ, Hagerman PJ. Fragile X syndrome: a model of gene-brain-behaviour relationships. *Rev. Neurol.* 2001 Oct;33 Suppl 1:S51-7. Review. Spanish.
- [64] Rodriguez-Revena L, Madrigal I, Pagonabarraga J, Xunclà M, Badenas C, Kulisevsky J, Gomez B, Milà M. Penetrance of FMR1 premutation associated pathologies in fragile X syndrome families. *Eur. J. Hum. Genet.* 2009 Oct;17(10):1359-62.
- [65] Hagerman RJ, Leavitt BR, Farzin F, Jacquemont S, Greco CM, Brunberg JA, Tassone F, Hessel D, Harris SW, Zhang L, Jardini T, Gane LW, Ferranti J, Ruiz L, Leehey MA, Grigsby J, Hagerman PJ. Fragile-X-associated tremor/ataxia syndrome (FXTAS) in females with the FMR1 premutation. *Am. J. Hum. Genet.* 2004 May;74(5):1051-6.
- [66] Bacalman S, Farzin F, Bourgeois JA, Cogswell J, Goodlin-Jones BL, Gane LW, Grigsby J, Leehey MA, Tassone F, Hagerman RJ. Psychiatric phenotype of the fragile X-associated tremor/ataxia syndrome (FXTAS) in males: newly described fronto-subcortical dementia. *J. Clin. Psychiatry.* 2006 Jan;67(1):87-94.
- [67] Bourgeois JA, Cogswell JB, Hessel D, Zhang L, Ono MY, Tassone F, Farzin F, Brunberg JA, Grigsby J, Hagerman RJ. Cognitive, anxiety and mood disorders in the fragile X-associated tremor/ataxia syndrome. *Gen. Hosp. Psychiatry.* 2007 Jul-Aug;29(4):349-56.
- [68] Hashimoto R, Backer KC, Tassone F, Hagerman RJ, Rivera SM. An fMRI study of the prefrontal activity during the performance of a working memory task in premutation carriers of the fragile X mental retardation 1 gene with and without fragile X-associated tremor/ataxia syndrome (FXTAS). *J. Psychiatr Res.* 2011 Jan;45(1):36-43.
- [69] Shiba M, Bower JH, Maraganore DM, McDonnell SK, Peterson BJ, Ahlskog JE, Schaid DJ, Rocca WA. Anxiety disorders and depressive disorders preceding Parkinson's disease: a case-control study. *Mov. Disord.* 2000 Jul;15(4):669-77.
- [70] Duff K, Paulsen JS, Beglinger LJ, Langbehn DR, Stout JC; Predict-HD Investigators of the Huntington Study Group. Psychiatric symptoms in Huntington's disease before diagnosis: the predict-HD study. *Biol. Psychiatry.* 2007 Dec 15;62(12):1341-6. Epub 2007 May 3.
- [71] Bourgeois JA, Seritan AL, Casillas EM, Hessel D, Schneider A, Yang Y, Kaur I, Cogswell JB, Nguyen DV, Hagerman RJ. Lifetime prevalence of mood and anxiety disorders in

- fragile X premutation carriers. *J. Clin. Psychiatry*. 2011 Feb;72(2):175-82. Epub 2010 Aug 24.
- [72] Sobesky WE, Hull CE, Hagerman RJ. Symptoms of schizotypal personality disorder in fragile X women. *J. Am. Acad. Child Adolesc. Psychiatry*. 1994 Feb;33(2):247-55.
- [73] Cornish K, Kogan C, Turk J, Manly T, James N, Mills A, Dalton A. The emerging fragile X premutation phenotype: evidence from the domain of social cognition. *Brain Cogn*. 2005 Feb;57(1):53-60.
- [74] Hocking DR, Kogan CS, Cornish KM. Selective spatial processing deficits in an at-risk subgroup of the fragile X premutation. *Brain Cogn*. 2012 Jun;79(1):39-44. doi: 10.1016/j.bandc.2012.02.005.
- [75] Mitchell RJ, Holden JJ, Zhang C et al: FMR1 alleles in Tasmania: a screening study of the special educational needs population. *Clin. Genet*. 2005; 67: 38–46.
- [76] Crawford DC, Meadows KL, Newman JL et al: Prevalence and phenotype consequence of FRAXA and FRAXE alleles in a large, ethnically diverse, special education-needs population. *Am. J. Hum. Genet* 1999; 64: 495–507.
- [77] Patsalis PC, Sismani C, Hettinger JA et al: Frequencies of ‘grey-zone’ and premutation-size FMR1 CGG-repeat alleles in patients with developmental disability in Cyprus and Canada. *Am. J. Med. Genet* 1999; 84: 195–197.
- [78] Mornet E, Chateau C, Simon-Bouy B, Serre JL: The intermediate alleles of the fragile X CGG repeat in patients with mental retardation. *Clin. Genet* 1998; 53: 200–201.
- [79] Otsuka S, Sakamoto Y, Siomi H et al: Fragile X carrier screening and FMR1 allele distribution in the Japanese population. *Brain Dev*. 2010; 32: 110–114.
- [80] Madrigal I, Xunclà M, Tejada MI, Martínez F, Fernández-Carvajal I, Pérez-Jurado LA, Rodríguez-Revenga L, Milà M. Intermediate FMR1 alleles and cognitive and/or behavioural phenotypes. *Eur. J. Hum. Genet*. 2011 Aug;19(8):921-3.

Chapter 91

CLINICAL FEATURES ASSOCIATED WITH *FMR1* PREMUTATION CARRIERS

M. I. Alvarez-Mora^{1,2} and L. Rodriguez-Revenge^{1,2,3,*}

¹Centre for Biomedical Network Research of Rare Disease (CIBERER), Barcelona, Spain

²Biochemistry and Molecular Genetics Department, Hospital Clinic, Barcelona, Spain

³IDIBAPS (Institut d'Investigacions Biomèdiques August Pi i Sunyer). Barcelona, Spain

ABSTRACT

The expansion of the CGG trinucleotide located within the 5'UTR of the *FMR1* gene is involved in a growing number of diseases; the most well-established are Fragile X Syndrome (FXS), Fragile X Tremor/Ataxia Syndrome (FXTAS) and Fragile X Primary Ovarian Insufficiency (FXPOI). Whereas full mutation alleles (>200CGGs) are responsible for the FXS, smaller expansions called premutation alleles (55-200CGGs) are associated with FXTAS and FXPOI. Numerous evidence have currently been reported suggesting that premutation alleles give rise to an increased risk for carriers of these alleles in relation to additional medical, psychiatric and cognitive features which occur at a greater frequency than what would be expected for the general population. In this chapter, we review the clinical features including peripheral neuropathy, immune-mediated disorders, migraines and neurocognitive involvement which have been suggested to be associated with premutation alleles. In addition, the current understanding of the pathogenic molecular mechanisms that give rise to the spectrum of *FMR1* premutation associated disorders is also reviewed. Although further research is needed in order to shed light on the factors underlying the common incomplete penetrance applicable to all phenotypes associated with the premutation, it is likely that a combination of environmental and genetic factors with differences in intrinsic susceptibility may modulate the appearance and the severity of these disorders.

Keywords: *FMR1* premutation, fibromyalgia, thyroid disease, peripheral neuropathy

INTRODUCTION

In the last years, there has been intense interest in identifying and characterizing the Fragile X premutation-associated phenotypes from the perspective not only of basic science but also of public health given its high prevalence affecting 1:250 females and 1:800 males among the general population [1]. Historically, carriers of *FMR1* premutation (PM) alleles were considered to be clinically unaffected, since the gene is generally not methylated and these individuals do not present intellectual disabilities (ID). The significance of these alleles was generally thought to be their propensity for expansion to the full mutation range (>200 CGG repeats) during maternal transmission resulting in transcriptional silencing of *FMR1* gene, absence of the encoding fragile X mental retardation 1 protein (FMRP) and manifestation of fragile X syndrome (FXS) [2,3]. Although, the features of this expansion are not fully understood, it is currently accepted that they depend on the size of the maternal allele and also on the number of the AGG interruptions within the CGG-repeat track [4]. The AGG interruptions are likely to have stabilizing effects during transmission by decreasing the risk of DNA polymerase slippage during DNA replication [5]. In PM male carriers there is a relative stable transmission, and thus, the risk of expansion to a full mutation is negligible [reviewed in 6].

Despite the belief that PM carriers do not present signs of clinical involvement, prior to the discovery of the gene *FMR1* Cronister and collaborators (1991) [7] reported, higher rates of premature ovarian failure (POF) among women heterozygous for X-chromosome fragility. Later, Allingham-Hawkins and colleagues (1999) [8] established PM alleles as a significant risk factor for POF based on the study of 760 women in which 16% of PM carriers were affected with POF whereas none of the full mutation carriers and just one (0.4%) of the controls presented with POF [8]. Currently, the association between ovarian deficiency and PM female carriers, namely Fragile X-Primary ovarian Insufficiency (FXPOI), is well-established, presenting an incidence of around 20% in these women and estimated at around 1% among the general population. Chapter 2 describes the features of FXPOI. Ten years after, the first description of a neurodegenerative disorder named Fragile X Tremor Ataxia Syndrome (FXTAS) associated to older adults PM carriers was made by Hagerman and colleagues (2001) [9]. This syndrome is characterized by white matter changes and global brain atrophy, presenting with core features of intention tremor and gait ataxia. Details of FXTAS are described in Chapter 3.

Over the past 10-15 years, an increasingly broad spectrum of clinical manifestations has been related to individuals who are carriers of PM alleles (Table 1). Domains of clinical involvement seen in some, but not all carriers of PM encompass the presence of medical, emotional and cognitive manifestations which have been widely reported to occur more frequently among PM carriers than in the general population. Some of these features have recently been classified as being ‘definitely related’, ‘probably related’, ‘possibly related’ or ‘not likely related’ to the molecular changes associated with an *FMR1* expansion based on clinical and previously reported data [reviewed in 10].

Coffey and collaborators (2008) [11] reported the first evidence of an expanded clinical phenotype of women with the PM. In this study, PM female carriers without core features of FXTAS showed significantly more complaints of chronic muscle pain, persistent paraesthesias in the extremities, and a history of tremor than controls. Furthermore, a significantly greater

presence of medical co-morbidity was detected in females with definite or probable FXTAS, with an increased prevalence of thyroid disease, hypertension, seizures, peripheral neuropathy, fibromyalgia compared with controls [11]. In addition, some of the comorbidities associated with FXTAS beyond central nervous system involvement, specifically peripheral neuropathy [11,12] and neuroendocrine dysfunction [11,13-15] have also been associated with PM carriers without FXTAS. Moreover, there is increasing evidence that young PM carriers present increased rates of neurodevelopmental phenotypes such as attention-deficit hyperactivity disorder (ADHD), autism spectrum disorder (ASD) and seizures [reviewed in 16]. Furthermore, increased rates of psychiatric involvement, particularly depression and anxiety have also been associated with FXTAS among adult PM carriers [reviewed in 17]. Neuropsychiatric aspects are discussed in Chapter 4.

Table 1. Clinical manifestations associated with some *FMR1* premutation carriers

	Cohort studied*	References
Immune mediated disorders		
Fibromyalgia	PM females	[11-15]
Thyroid disease	PM females	[11,13,15]
Irritable bowel syndrome	PM females	[15]
Neurodevelopmental Phenotypes		
Working memory deficiencies	PM males	[27,28]
Language dysfluencies	PM females	[29]
Spatiotemporal processing impairment	Young-adult PM carriers	[32]
Arithmetic weaknesses	PM females	[30,31]
Developmental Delay	PM carriers	[75]
Reproductive features		
Ovarian Insufficiency (FXPOI)	PM females	[reviewed in 10]
Obstetric and perinatal difficulties	PM females	
Estrogen-deficiency related conditions	FXPOI PM females	
Autonomic dysfunction		
Impotence	FXTAS PM males	[56]
Hypertension	FXTAS PM both	[11,67]
Bowel and bladder incontinence	FXTAS PM both	[18,68]
Neurocognitive and psychiatric Involvement		
Depression	PM females	[69-71]
Anxiety disorders	PM females	
Mood disorders	PM females	
Seizures	PM male children	[72]
ASD	PM male children	[72,73]
ADHD	PM females	[74]
Other clinical manifestations		
Migraine	PM carriers both	[25]
Peripheral neuropathy	PM carriers both	[11,76]

CLINICAL FEATURES ASSOCIATED WITH *FMR1* PREMUTATION CARRIERS

Neuropathy

Peripheral neuropathy is characterized by a damaging of peripheral nerves which carry information to and from the brain as well as to and from the spinal cord to the rest of the body.

Depending on the type of nerve affected it may promote impaired sensation, movement, gland or organ function. The association between neuropathy and *FMR1* alleles was first reported in PM carriers with FXTAS [9,18]. Thereafter, Berry-Kravis and collaborators (2007) [12] reported the first evidence that signs of neuropathy on clinical examination are associated with PM carrier status based on neurological examination data from 207 unrelated individuals. Results from this study revealed that the degree of clinical involvement strongly correlated with the CGG repeat length in males since these individuals presented significantly higher mean scores in the neuropathy screening scale score ($P=0.0014$), the vibration score ($P=0.0015$), and the reflex score ($P=0.0014$) than sex-matched controls suggesting that PM male carriers present higher impairment of both distal vibratory sense and reflexes [12]. The lack of significant differences among PM female carriers presumably reflects the broad variation in clinical involvement among carriers as a result of variation in the X-chromosome activation ratio as well as the decreased penetrance of the clinical manifestation due to the protective effect of the second non-mutated X chromosome [12]. Afterwards, Coffey and collaborators (2008) [11] demonstrated that females with PM alleles also presented significantly higher signs of neuropathy based on findings from 128 PM female carriers without FXTAS. In this study it was shown that these women presented a significant rate of numbness and tingling and muscle pain in the extremities, albeit evidence of neuropathy is increased in PM female carriers presenting with FXTAS [11]. It has been suggested that neuropathic symptoms in PM female carriers are manifested together with the emergence of FXTAS disease [reviewed in 10].

Immune-Mediated Disorders

Numerous reports support elevated rates of immune-mediated disorders (IMD) in PM female carriers, particularly regarding hypothyroidism and fibromyalgia [11, 13-15]. In contrast, IMDs have not been evidenced among males with the premutation, likely due to the relative rarity of these disorders among the general male population. In a recent study, Winarni and colleagues (2012) [15] examined the relative likelihood of large clinical manifestations among 344 female carriers of PM alleles and 72 controls including autoimmune thyroid disorder, multiple sclerosis, Sjögren syndrome, rheumatoid arthritis, systemic lupus erythematosus, Raynaud's phenomenon, irritable bowel syndrome and optic neuritis. The results of this study evidenced that among women over 40 years of age 46.54% of PM females without FXTAS experienced one or more of the IMDs surveyed, and the prevalence increased to about 72.73% for those with FXTAS compared to 31.58% for the control group [15]. With respect to FXPOI, both groups of PM females carriers present higher odds ratios of IMDs compared to controls, and similarly, when considering FXTAS symptoms, the odds ratio of IMDs among PM female carriers presenting with FXPOI is about 2.4-fold higher when compared to those without FXPOI. Moreover, these authors found that an autoimmune thyroid disorder was the most common IMD followed by fibromyalgia and irritable bowel syndrome [15].

Increased penetrance for both thyroid disease and fibromyalgia has been broadly reported among PM female carriers [13], although the penetrance of these disorders is highly variable among the general population since it increases with age. For thyroid disease, it has been suggested that the association with PM alleles may be more relevant in older women [10] due to the lack of statistical significance when considering women between 18 and 50 years of age

[19]. Nonetheless, in the general population the penetrance estimated for thyroid disorders is of around 10% [20] and around 2-4% for fibromyalgia whereas it has been estimated to be around 15.9% and 24.4%, respectively, among PM female carriers [13]. Conversely, in 700 unrelated Spanish patients with fibromyalgia the frequency of PM alleles did not significantly differ from the estimated rate in the general population [21]. In contrast, another study found a higher incidence of PM alleles among a Spanish female fibromyalgia cohort. Indeed, the incidence of PM alleles was 1 of 88, being 1 of 250 in females in the general population [22]. These controversial results are likely to be caused by a sample size effect since the data reported by Martorell and colleagues (2012) [22] were based on the screening of 353 females whereas the data presented by Rodriguez-Revenga and coworkers (2013) [21] were based on 700 samples. Nonetheless, it has been shown that the pathophysiology of fibromyalgia involves hyperexcitability of central neurons through several synaptic and neurotransmitter/neurochemical mechanisms [reviewed in 23] suggesting that it could arise through an alteration of pain neurotransmitter mechanisms among PM female carriers [14].

Finally, it has been recently demonstrated that individual carriers of PM alleles present an immune dysregulation and decreased immune responses when compared with healthy controls [24]. Moreover, it has been found that PM carriers present a reduction in the levels of cytokine production which is negatively associated with CGG repeat length, mainly with IL-12 production [24]. Furthermore, these women have also been shown to present a decrease in the relative levels of the surface marker CD25 in T cells suggesting potential differences in the activation of T-cells that regulate immune response [24].

Migraines

Au and collaborators (2013) [25] have recently reported that PM carriers show increased rates in the prevalence of migraines based on physical and medical examination of 315 PM carriers (203 females and 112 males) and 154 controls (83 females and 71 males). Migraine is a neurologic disorder characterized by light and sound sensitivity and pulsatile pain, which is thought to have a polygenic and multifactorial etiology. The prevalence of migraine among the general population is estimated around 27.3% in women and 9.7% in men [26], reaching up to 54.2% and 26.79% among female and male PM carriers, respectively, both resulting in statistical significant differences [25]. In addition, this significance was obtained considering both those affected with and also those without FXTAS, adjusted for age. However, the risk of migraine headaches was not correlated with either CGG repeats or *FMR1* mRNA expression [25].

Neurocognitive Features

The expanded range of clinical involvement associated with PM carriers also includes an alteration of various cognitive domains including executive function, working memory and arithmetic skills which become apparent even in young individuals, with a usually more progressive course in PM individuals than in the general population [reviewed in 10]. However, neurocognitive deficits are reportedly more frequent in male than in female carriers of PM alleles. Results reported by Kogan and collaborators (2008) [27] revealed that the CGG

expansion confers a significant risk for working memory difficulties based on a controlled study of 40 PM male carriers without manifest symptoms of FXTAS. In addition, Cornish and colleagues (2009) [28] reported neuropsychological measures in PM males regarding core subcomponents of working memory such as verbal memory, visual-spatial memory, and central executive memory revealing that PM males present specific vulnerability in executive control of memory including tasks requiring simultaneous manipulation and storage of new information, regardless of the presence of FXTAS symptoms. These authors revealed an impairment of the central executive working memory among PM male carriers without FXTAS, which was significantly correlated with larger CGG repeat expansions, whereas FXTAS patients demonstrated a more general impairment in terms of phonological working memory in addition to central executive working memory [28]. Moreover, language dysfluencies associated with deficits in organization and planning have been evidenced among PM female carriers [29]. Past research has demonstrated that language dysfluencies are an indicator of executive functioning deficits which are characteristic of other neurodegenerative disorders such as Parkinson and Alzheimer diseases.

Regarding arithmetic skills it has been suggested that PM female carriers show weaknesses in mathematical tasks [30] and this has recently been supported by other groups [31]. Furthermore, it has been demonstrated that PM carriers from 19 to 45 years of age show impairment in spatiotemporal processing which may underlie the impairments observed in arithmetic skills among these individuals since the representations of space and time provide the foundation for an understanding of numbers [32].

Although further research is needed, it has been suggested that determining whether cognitive impairments are detectable in PM carriers without FXTAS should be prudent since it may not only be an early indicator of cognitive decline in PM carriers [29] but could also be used as a biomarker of disease progression if these features precede motor impairment [32].

CURRENT UNDERSTANDING OF THE MOLECULAR MECHANISMS UNDERLYING PREMUTATION-ASSOCIATED PATHOLOGIES

Premutation-associated phenotypes have been mainly attributed to a pathogenic mechanism involving a gain-of-function toxicity of the expanded *FMR1* mRNA, a process entirely distinct from the FMRP deficiency responsible for the FXS phenotype [reviewed in 33]. This observation was based on the restriction of these clinical phenotypes to the premutation range in which the molecular signature is a 2-8 fold increase in the expression of the PM mRNA and, paradoxically, a slight reduction in Fragile X Mental Retardation Protein (FMRP) levels [34]. The mechanisms underlying the increased transcriptional activity of the PM alleles remain unclear, however, it has been reported that these alleles use differential transcriptional start sites leading to different expression compared to non-expanded *FMR1* alleles [35]. On other hand, it has also been suggested that the reduction of FMRP in PM carriers may promote an added contribution to the clinical involvement observed in both children and adults related to phenotypes associated with reduced cognition and disturbed behavior [36]. However, the FMRP deficiency cannot be a driving factor in the PM-associated disorders since FXTAS and FXPOI are not experienced by full mutation carriers. Despite wide reports that the levels of FMRP expression are only slightly decreased in patients with FXTAS, most of these

measurements have been performed using lymphocytes or whole blood samples rather than brain tissue, in which changes are likely to be more robust [reviewed in 37]. The efficiency of FMRP translation is associated with the length of the CGG expansion in light of impairment in translation of larger CGG tracks by interference with ribosomal scanning through the 5'UTR, thereby preventing appropriate loading of the expanded *FMR1* mRNA into polyribosomal complexes [38-40].

It is currently considered that the dual mechanism of involvement, the excess of PM mRNA expression and the decrease in FMRP translation, is a double hit which may promote phenotypic features of FXS and PM associated disorders [reviewed in 33, 41]. Notwithstanding, the reduction of FMRP synthesis, the phenotype of PM carriers, is different from carriers of full mutation alleles since milder protein deficiency among PM carriers usually leads to mild developmental problems with these individuals having higher IQs and less severe behavioral problems than those with FXS [reviewed in 42]. A FMRP deficit has been correlated with lowered activity of the amygdala among PM male carriers compared to controls on functional magnetic resonance imaging whereas increased levels of expanded mRNA have been strongly associated with obsessive-compulsive symptoms and psychoticism in PM male carriers [43,44]. It has also been described that RNA toxicity leads to the up-regulation of the heat shock proteins Hsp70 and α B-crystallin, which may stimulate immune dysregulation [15]. Regarding migraines, their association with mitochondrial dysfunction is well established. Interestingly, there is evidence pointing to a deregulation of mitochondrial function among PM carriers [45-47], therefore the increased prevalence of migraines in PM carriers may be the result of RNA toxicity leading to mitochondrial deregulation [25]. Likewise, RNA toxicity has also been proposed to shed light on the high rate of thyroid dysfunction among females with the PM by causing a direct effect on the hypothalamic-pituitary-adrenal axis or on the thyroid gland. Additionally, RNA toxicity has also been proposed to promote an autoimmune mechanism or apoptosis in thyroid cells [11]. However, data reported by Cunningham and coworkers (2011) [48] show that the presence of PM alleles promotes migration defects in the neocortex and altered expression of neuronal lineage markers among embryonic PM mice. These results support the hypothesis that the role of the RNA toxicity may be restricted to the initial triggering events since many features of the neuronal and astrocytic cellular phenotype observed in FXTAS patients are already present in the neonatal period, suggesting that the clinical involvement among children carriers of PM alleles may be manifestations of this early, non-degenerative process [reviewed in 33].

The sequestration hypothesis of RNA toxicity was first proposed and established for myotonic dystrophy type 1 caused by an expansion of a CTG repeat in the 3'UTR of *DMPK* gene [reviewed in 49]. Particularly, in Fragile X PM the model hypothesized that expanded CGG repeats form hairpin loops which are sticky and recruit an excess of specific RNA-binding proteins, resulting in a functional insufficiency of the sequestered proteins and leading to cell dysfunction and death [50]. Although the initial triggering events in these disorders are based on RNA toxicity, this model does not provide evidence regarding cell sickening and death. In this way, there are several candidate downstream pathways, although alterations of both mitochondrial function and calcium regulation are emerging as core mediators of cellular deregulation and dysfunction [reviewed in 31, 39]. The increased expression of the expanded *FMR1* mRNA is thought to be the main cause of clinical involvement in PM carriers since the *FMR1* mRNA and the sequestered proteins form aggregates leading to intranuclear inclusions present in several tissues. Interestingly, the expanded mRNA of *FMR1* is detected within the

inclusion whereas FMRP is not present [51]. Furthermore, proteomic analysis of these inclusions revealed a large number of proteins presented in the aggregates, including the RNA binding protein hnRNP A2/B1, the nuclear envelope protein lamin A/C, the small heat shock protein α B-crystallin [52], the splicing factor Sam68 [53] and part of the microRNA processor complex DGCR8 [50]. Indeed, it has recently been demonstrated that the sequestration of DGCR8 promotes the deregulation of microRNAs biogenesis, suggesting a central role of DGCR8 as an inductor of RNA toxicity by leading to cell dysfunction and cell loss [50].

Intranuclear inclusions, which represent the neuropathological hallmark of FXTAS [52], have been also detected among PM carriers through different cell types including the central and peripheral nervous system and other tissue including the adrenal glands, the testes, pancreas and heart [54-58]. Recently, Hunsaker and colleagues (2011) [58] reported autopsy findings from ten PM carriers with FXTAS in which intranuclear inclusions were detected throughout multiple tissues including the hypothalamic-pituitary-adrenal axis, pineal gland, cardiac conduction system, peripheral nerves and autonomic ganglia, the thyroid gland, the digestive system, the testes and pancreas. The broad distribution of these inclusions suggests that many organ systems may be affected by RNA toxicity. Nevertheless, it is necessary to study the processes underlying inclusion formation in depth to address whether they themselves are toxic or reflect cellular dysfunction [58].

Furthermore, additional mechanisms have been proposed as possible triggering events in the PM-associated disorders, although most evidence support CGG-repeat mediated protein sequestration [reviewed in 33,37]. These mechanisms include a RNA-mediated protein aggregation model whereby the CGGs contained within the *FMR1* mRNA might promote a conformational transition in proteins with prion-like domains that may initiate a cascade of protein aggregation similar to what occurs in amyloid plaque formation in Alzheimer's disease. Moreover, the production of a toxic polyglycine peptide has also been proposed as an alternative toxicity model by a non-AUG-initiated (RAN) translation [59]. In addition, a specific splicing isoform is detected exclusively with transcripts of PM alleles suggesting a possible role of antisense transcripts generated at the *FMR1* locus [60].

Otherwise very little attention has been focused on the other end of the spectrum, the so-called "low-normal" numbers of CGG repeats, up to 23 trinucleotide repeats. In this framework, data based on genotype-phenotype correlations have recently been reported suggesting that this range of CGG repeats may have substantial implications for cognitive functioning, cancer, and the odds of having children with neurodevelopmental or neuropsychiatric conditions [61]. Despite these range of CGG number not being associated with altered FMRP synthesis, Chen and co-workers (2003) [62] reported that the efficiency of FMRP translation was based on the number of CGG repeats, conferring the greatest efficiency of protein synthesis to the allele of 30 repeats. Thus, inefficient translation may be related to the clinical manifestations associated with the low numbers of CGG repeats. Likewise, Ramocki and Zoghbi (2008) [63] suggested that imbalances in homeostatic controls of multiple genes including *FMR1* may partially promote the appearance of neurodevelopmental and neuropsychiatric disorders.

At this point, the difficulty in understanding PM-associated pathologies lies in the inability of prediction of which PM carriers will develop any of these phenotypes. The incomplete penetrance across the phenotypic spectrum is likely to be associated with a combination of genetic and environmental factors which may confer specific vulnerability to PM carriers (Figure 1).

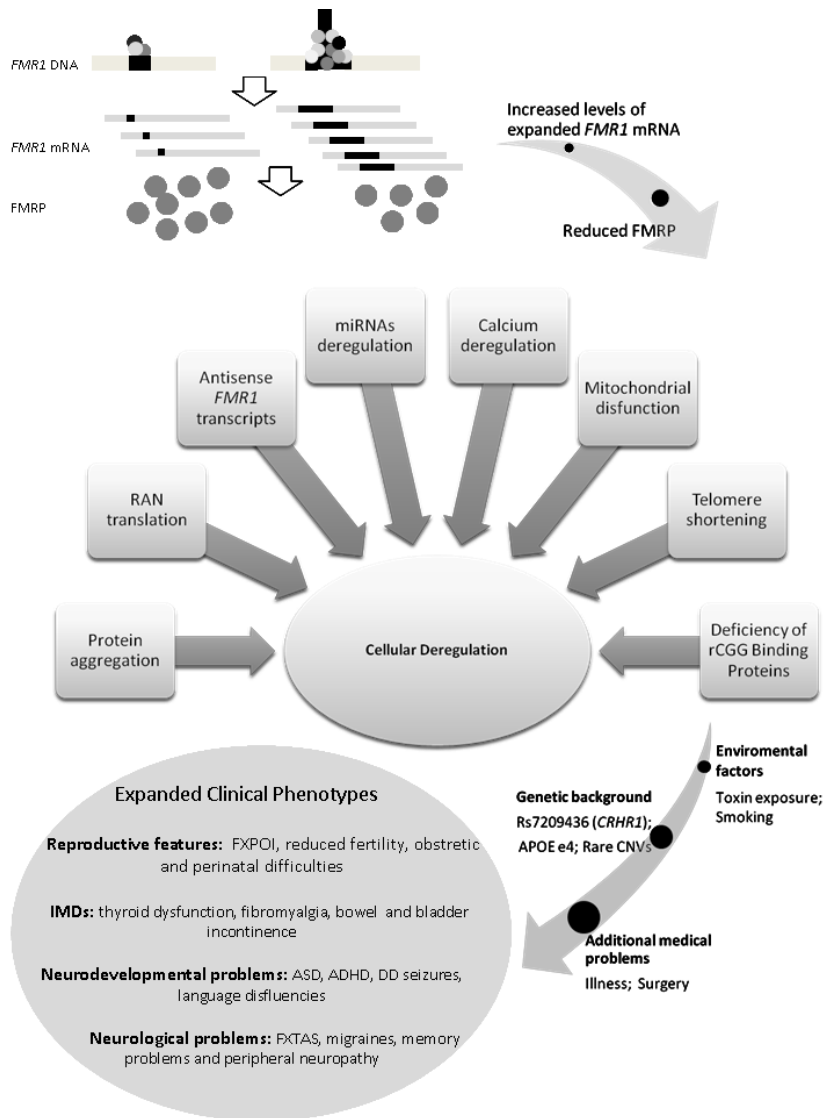


Figure 1. Diagram of the potential players contributing to the clinical involvement associated to PM carriers.

Genetic factors that may contribute to the PM-associated disorders include CGG repeat length, expression levels of the expanded *FMR1* mRNA, aberrant translation of the repeat sequence as well as genomic changes in other regions of the genome. Within this framework, specific polymorphisms of the *CRHR1* gene have been associated with female clinical involvement (rs7209436), particularly with depression and anxiety mainly as this gene regulates the expression and release of ACTH from the anterior pituitary gland which, in turn, stimulates the release of cortisol from the adrenal cortex [64]. Furthermore, risk factors for other neurodegenerative disorders such as allele $\epsilon 4$ of the *APOE* gene may also influence the risk of FXTAS as a higher frequency of these alleles has been reported in PM carriers with compared to those without FXTAS [65]. Moreover, it has recently been suggested that

individuals who are carriers of PM alleles presenting with ID, seizures or ASD are likely to have a second hit since PM carriers show a significant enrichment ($P=2.27\text{e-}07$) of CNVs compared to controls [66]. Furthermore, these authors found an association between the presence of rare CNVs (not detected in 8000 controls) among PM carriers with either autistic traits or neurological involvement, suggesting that they may have a possible role in this phenotypic variability [66]. Regarding environmental factors, it has been suggested that smoking, prolonged surgery with anesthesia, drug and alcohol abuse or the stress of carrying a child with FXS are also puzzling factors which may act as additional determinants for the phenotypic variability among PM individuals [reviewed in 41,42]. Finally, further longitudinal studies are required to determine the context in which any of the PM-associated phenotypes are developed and what protective factors might reduce the risks of more negative outcomes [reviewed in 10].

CONCLUSION

Overall, it is currently accepted that PM alleles led to multiple distinct clinical features which are present at a greater frequency among PM carriers than what would be expected in the general population. However, the association of the PM with the phenotypes reviewed in this chapter is less well established than its association with FXTAS and FXPOI. Although further research is needed in order to shed light on the factors underlying the common incomplete penetrance applicable to all phenotypes associated with the PM, a combination of environmental and genetic factors with differences in intrinsic susceptibility likely modulate the appearance and the severity of these disorders.

REFERENCES

- [1] Hagerman PJ. The fragile X prevalence paradox. *J Med. Genet.* 2008 Aug;45(8):498-9. Erratum in: *J. Med. Genet.* 2008 Nov;45(11):768.
- [2] Rousseau F, Heitz D, Biancalana V, Blumenfeld S, Kretz C, Boué J, Tommerup N, Van Der Hagen C, DeLozier-Blanchet C, Croquette MF, et al. Direct diagnosis by DNA analysis of the fragile X syndrome of mental retardation. *N. Engl. J. Med.* 1991 Dec 12;325(24):1673-81.
- [3] Nolin SL, Lewis FA 3rd, Ye LL, Houck GE Jr, Glicksman AE, Limprasert P, Li SY, Zhong N, Ashley AE, Feingold E, Sherman SL, Brown WT. Familial transmission of the FMR1 CGG repeat. *Am. J. Hum. Genet.* 1996 Dec;59(6):1252-61.
- [4] Yrigollen CM, Tassone F, Durbin-Johnson B, Tassone F. The role of AGG interruptions in the transcription of PM alleles. *PLoS One.* 2011;6(7):e21728.
- [5] Gacy AM, Goellner G, Juranić N, Macura S, McMurray CT. Trinucleotide repeats that expand in human disease form hairpin structures in vitro. *Cell.* 1995 May 19;81(4):533-40.
- [6] Brouwer JR, Willemsen R, Oostra BA. Microsatellite repeat instability and neurological disease. *Bioessays.* 2009 Jan;31(1):71-83. doi: 10.1002/bies.080122.

- [7] Cronister A, Schreiner R, Wittenberger M, Amiri K, Harris K, Hagerman RJ. Heterozygous fragile X female: historical, physical, cognitive, and cytogenetic features. *Am. J. Med. Genet.* 1991 Feb-Mar;38(2-3):269-74.
- [8] Allingham-Hawkins DJ, Babul-Hirji R, Chitayat D, Holden JJ, Yang KT, Lee C, Hudson R, Gorwill H, Nolin SL, Glicksman A, Jenkins EC, Brown WT, Howard-Peebles PN, Becchi C, Cummings E, Fallon L, Seitz S, Black SH, Vianna-Morgante AM, Costa SS, Otto PA, Mingroni-Netto RC, Murray A, Webb J, Vieri F, et al. Fragile X premutation is a significant risk factor for premature ovarian failure: the International Collaborative POF in Fragile X study--preliminary data. *Am. J. Med. Genet.* 1999 Apr 2;83(4):322-5.
- [9] Hagerman RJ, Leehey M, Heinrichs W, Tassone F, Wilson R, Hills J, Grigsby J, Gage B, Hagerman PJ. Intention tremor, parkinsonism, and generalized brain atrophy in male carriers of fragile X. *Neurology.* 2001 Jul 10;57(1):127-30.
- [10] Wheeler AC, Bailey DB Jr, Berry-Kravis E, Greenberg J, Losh M, Mailick M, Milà M, Olichney JM, Rodriguez-Revenga L, Sherman S, Smith L, Summers S, Yang JC, Hagerman R. Associated features in females with an PM. *J Neurodev Disord.* 2014;6(1):30. Review.
- [11] Coffey SM, Cook K, Tartaglia N, Tassone F, Nguyen DV, Pan R, Bronsky HE, Yuhas J, Borodyanskaya M, Grigsby J, Doerflinger M, Hagerman PJ, Hagerman RJ. Expanded clinical phenotype of women with the PM. *Am. J. Med. Genet. A.* 2008 Apr 15;146A(8):1009-16.
- [12] Berry-Kravis E, Goetz CG, Leehey MA, Hagerman RJ, Zhang L, Li L, Nguyen D, Hall DA, Tartaglia N, Cogswell J, Tassone F, Hagerman PJ. Neuropathic features in fragile X premutation carriers. *Am. J. Med. Genet. A.* 2007 Jan 1;143A(1):19-26.
- [13] Rodriguez-Revenga L, Madrigal I, Pagonabarraga J, Xunclà M, Badenas C, Kulisevsky J, Gomez B, Milà M. Penetrance of FMR1 premutation associated pathologies in fragile X syndrome families. *Eur. J. Hum. Genet.* 2009 Oct;17(10):1359-62.
- [14] Leehey MA, Legg W, Tassone F, Hagerman R. Fibromyalgia in fragile X mental retardation 1 gene premutation carriers. *Rheumatology (Oxford).* 2011 Dec;50(12):2233-6.
- [15] Winarni TI, Chonchaiya W, Sumekar TA, Ashwood P, Morales GM, Tassone F, Nguyen DV, Faradz SM, Van de Water J, Cook K, Hamlin A, Mu Y, Hagerman PJ, Hagerman RJ. Immune-mediated disorders among women carriers of fragile X premutation alleles. *Am. J. Med. Genet. A.* 2012 Oct;158A(10):2473-81.
- [16] Besterman AD, Wilke SA, Mulligan TE, Allison SC, Hagerman R, Seritan AL, Bourgeois JA. Towards an Understanding of Neuropsychiatric Manifestations in Fragile X Premutation Carriers. *Future Neurol.* 2014 Mar;9(2):227-239.
- [17] Hunter JE, Abramowitz A, Rusin M, Sherman SL. Is there evidence for neuropsychological and neurobehavioral phenotypes among adults without FXTAS who carry the PM? A review of current literature. *Genet Med.* 2009 Feb;11(2):79-89.
- [18] Jacquemont S, Hagerman RJ, Leehey M, Grigsby J, Zhang L, Brunberg JA, Greco C, Des Portes V, Jardini T, Levine R, Berry-Kravis E, Brown WT, Schaeffer S, Kissel J, Tassone F, Hagerman PJ. Fragile X premutation tremor/ataxia syndrome: molecular, clinical, and neuroimaging correlates. *Am. J. Hum. Genet.* 2003 Apr;72(4):869-78.
- [19] Hunter JE, Rohr JK, Sherman SL. Co-occurring diagnoses among FMR1 premutation allele carriers. *Clin. Genet.* 2010 Apr;77(4):374-81.

- [20] Canaris GJ, Manowitz NR, Mayor G, Ridgway EC. The Colorado thyroid disease prevalence study. *Arch. Intern Med.* 2000 Feb 28;160(4):526-34.
- [21] Rodriguez-Revenga L, Madrigal I, Blanch-Rubió J, Elurbe DM, Docampo E, Collado A, Vidal J, Carbonell J, Estivill X, Mila M. Screening for the presence of PM alleles in women with fibromyalgia. *Gene.* 2013 Jan 10;512(2):305-8.
- [22] Martorell L, Tondo M, Garcia-Fructuoso F, Naudo M, Alegre C, Gamez J, Genovés J, Poo P. Screening for the presence of FMR1 premutation alleles in a Spanish population with fibromyalgia. *Clin. Rheumatol.* 2012 Nov;31(11):1611-5.
- [23] Yunus MB. Role of central sensitization in symptoms beyond muscle pain, and the evaluation of a patient with widespread pain. *Best Pract. Res. Clin. Rheumatol.* 2007 Jun;21(3):481-97. Review.
- [24] Careaga M, Rose D, Tassone F, Berman RF, Hagerman R, Ashwood P. Immune dysregulation as a cause of autoinflammation in fragile X premutation carriers: link between FMRI CGG repeat number and decreased cytokine responses. *PLoS One.* 2014 Apr 9;9(4):e94475.
- [25] Au J, Akins RS, Berkowitz-Sutherland L, Tang HT, Chen Y, Boyd A, Tassone F, Nguyen DV, Hagerman R. Prevalence and risk of migraine headaches in adult fragile X premutation carriers. *Clin Genet.* 2013 Dec;84(6):546-51.
- [26] Lipton RB, Stewart WF, Diamond S, Diamond ML, Reed M. Prevalence and burden of migraine in the United States: data from the American Migraine Study II. *Headache* 2001; 41 (7): 646–657.
- [27] Kogan CS, Turk J, Hagerman RJ, Cornish KM. Impact of the Fragile X mental retardation 1 (FMR1) gene premutation on neuropsychiatric functioning in adult males without fragile X-associated Tremor/Ataxia syndrome: a controlled study. *Am. J. Med. Genet. B Neuropsychiatr Genet.* 2008 Sep 5;147B(6):859-72.
- [28] Cornish KM, Kogan CS, Li L, Turk J, Jacquemont S, Hagerman RJ. Lifespan changes in working memory in fragile X premutation males. *Brain Cogn.* 2009 Apr;69(3):551-8.
- [29] Sterling AM, Mailick M, Greenberg J, Warren SF, Brady N. Language dysfluencies in females with the PM. *Brain Cogn.* 2013 Jun;82(1):84-9.
- [30] Lachiewicz AM, Dawson DV, Spiridigliozzi GA, McConkie-Rosell A. Arithmetic difficulties in females with the fragile X premutation. *Am. J. Med. Genet. A.* 2006 Apr 1;140(7):665-72.
- [31] Semenza C, Bonollo S, Polli R, Busana C, Pignatti R, Iuculano T, Maria Laverda A, Priftis K, Murgia A. Genetics and mathematics: FMR1 premutation female carriers. *Neuropsychologia.* 2012 Dec;50(14):3757-63
- [32] Wong LM, Goodrich-Hunsaker NJ, McLennan Y, Tassone F, Harvey D, Rivera SM, Simon TJ. Young adult male carriers of the fragile X premutation exhibit genetically modulated impairments in visuospatial tasks controlled for psychomotor speed. *J. Neurodev Disord.* 2012 Nov 13;4(1):26.
- [33] Hagerman P. Fragile X-associated tremor/ataxia syndrome (FXTAS): pathology and mechanisms. *Acta Neuropathol.* 2013 Jul;126(1):1-19.
- [34] Tassone F, Beilina A, Carosi C, Albertosi S, Bagni C, Li L, Glover K, Bentley D, Hagerman PJ. Elevated FMR1 mRNA in premutation carriers is due to increased transcription. *RNA.* 2007 Apr;13(4):555-62.

- [35] Tassone F, De Rubeis S, Carosi C, La Fata G, Serpa G, Raske C, Willemsen R, Hagerman PJ, Bagni C. Differential usage of transcriptional start sites and polyadenylation sites in FMR1 premutation alleles. *Nucleic Acids Res.* 2011 Aug;39(14):6172-85.
- [36] Tassone F, Hagerman RJ, Taylor AK, Mills JB, Harris SW, Gane LW, Hagerman PJ (2000) Clinical involvement and protein expression in individuals with the FMR1 premutation. *Am. J. Med. Genet.* 91(2):144–152
- [37] Renoux AJ, Todd PK. Neurodegeneration the RNA way. *Prog. Neurobiol.* 2012 May;97(2):173-89.
- [38] Primerano B, Tassone F, Hagerman RJ, Hagerman P, Amaldi F, Bagni C. Reduced FMR1 mRNA translation efficiency in fragile X patients with premutations. *RNA.* 2002 Dec;8(12):1482-8.
- [39] Ludwig AL, Raske C, Tassone F, Garcia-Arocena D, Hershey JW, Hagerman PJ. Translation of the FMR1 mRNA is not influenced by AGG interruptions. *Nucleic Acids Res.* 2009 Nov;37(20):6896-904.
- [40] Ludwig AL, Hershey JW, Hagerman PJ. Initiation of translation of the FMR1 mRNA Occurs predominantly through 5'-end-dependent ribosomal scanning. *J. Mol. Biol.* 2011 Mar 18;407(1):21-34.
- [41] Hagerman R, Hagerman P. Advances in clinical and molecular understanding of the FMR1 premutation and fragile X-associated tremor/ataxia syndrome. *Lancet Neurol.* 2013 Aug;12(8):786-98.
- [42] Polussa J, Schneider A, Hagerman R. (2014) Molecular Advances Leading to Treatment Implications for Fragile X Premutation Carriers. *Brain Disord Ther* 3:2.
- [43] Hessler D, Tassone F, Loesch DZ, Berry-Kravis E, Leehey MA, Gane LW, Barbato I, Rice C, Gould E, Hall DA, Grigsby J, Wegelin JA, Harris S, Lewin F, Weinberg D, Hagerman PJ, Hagerman RJ. Abnormal elevation of FMR1 mRNA is associated with psychological symptoms in individuals with the fragile X premutation. *Am. J. Med. Genet. B Neuropsychiatr Genet.* 2005; 139B (1):115–121.
- [44] Hoem G, Raske CR, Garcia-Arocena D, Tassone F, Sanchez E, Ludwig AL, Iwahashi CK, Kumar M, Yang JE, Hagerman PJ. CGG-repeat length threshold for FMR1 RNA pathogenesis in a cellular model for FXTAS. *Hum. Mol. Genet.* 2011; 20 (11):2161–2170.
- [45] Ross-Inta C, Omanska-Klusek A, Wong S, Barrow C, Garcia-Arocena D, Iwahashi C, Berry-Kravis E, Hagerman RJ, Hagerman PJ, Giulivi C. Evidence of mitochondrial dysfunction in fragile X-associated tremor/ataxia syndrome. *Biochem. J.* 2010 Aug 1;429(3):545-52.
- [46] Napoli E, Ross-Inta C, Wong S, Omanska-Klusek A, Barrow C, Iwahashi C, Garcia-Arocena D, Sakaguchi D, Berry-Kravis E, Hagerman R, Hagerman PJ, Giulivi C. Altered zinc transport disrupts mitochondrial protein processing/import in fragile X-associated tremor/ataxia syndrome. *Hum. Mol. Genet.* 2011 Aug 1;20(15):3079-92.
- [47] Mateu-Huertas E, Rodriguez-Revenga L, Alvarez-Mora MI, Madrigal I, Willemsen R, Milà M, Martí E, Estivill X. Blood expression profiles of fragile X premutation carriers identify candidate genes involved in neurodegenerative and infertility phenotypes. *Neurobiol. Dis.* 2014 May;65:43-54.
- [48] Cunningham CL, Martínez Cerdeño V, Navarro Porras E, Prakash AN, Angelastro JM, Willemsen R, Hagerman PJ, Pessah IN, Berman RF, Noctor SC. Premutation CGG-

- repeat expansion of the Fmr1 gene impairs mouse neocortical development. *Hum. Mol. Genet.* 2011 Jan 1;20(1):64-79.
- [49] Todd PK, Paulson HL. RNA-mediated neurodegeneration in repeat expansion disorders. *Ann. Neurol.* 2010 Mar;67(3):291-300. Review.
- [50] Sellier C, Freyermuth F, Tabet R, Tran T, He F, Ruffenach F, Alunni V, Moine H, Thibault C, Page A, Tassone F, Willemsen R, Disney MD, Hagerman PJ, Todd PK, Charlet-Berguerand N. Sequestration of DROSHA and DGCR8 by expanded CGG RNA repeats alters microRNA processing in fragile X-associated tremor/ataxia syndrome. *Cell Rep.* 2013 Mar 28;3(3):869-80.
- [51] Tassone F, Iwahashi C, Hagerman PJ. FMR1 RNA within the intranuclear inclusions of fragile X-associated tremor/ataxia syndrome (FXTAS). *RNA Biol.* 2004 Jul;1(2):103-5.
- [52] Iwahashi CK, Yasui DH, An HJ, Greco CM, Tassone F, Nannan K, Babineau B, Lebrilla CB, Hagerman RJ, Hagerman PJ. Protein composition of the intranuclear inclusions of FXTAS. *Brain.* 2006 Jan;129(Pt 1):256-71.
- [53] Sellier C, Rau F, Liu Y, Tassone F, Hukema RK, Gattoni R, Schneider A, Richard S, Willemsen R, Elliott DJ, Hagerman PJ, Charlet-Berguerand N. Sam68 sequestration and partial loss of function are associated with splicing alterations in FXTAS patients. *EMBO J.* 2010 Apr 7;29(7):1248-61.
- [54] Louis E, Moskowitz C, Friez M, Amaya M, Vonsattel JP. Parkinsonism, dysautonomia, and intranuclear inclusions in a fragile X carrier: a clinical-pathological study. *Mov. Disord.* 2006 Mar;21(3):420-5.
- [55] Greco CM, Berman RF, Martin RM, Tassone F, Schwartz PH, Chang A, Trapp BD, Iwahashi C, Brunberg J, Grigsby J, Hessel D, Becker EJ, Papazian J, Leehey MA, Hagerman RJ, Hagerman PJ. Neuropathology of fragile X-associated tremor/ataxia syndrome (FXTAS). *Brain.* 2006 Jan;129(Pt 1):243-55.
- [56] Greco CM, Soontrapornchai K, Wirojanan J, Gould JE, Hagerman PJ, Hagerman RJ. Testicular and pituitary inclusion formation in fragile X associated tremor/ataxia syndrome. *J. Urol.* 2007 Apr;177(4):1434-7.
- [57] Gokden M, Al-Hinti JT, Harik SI. Peripheral nervous system pathology in fragile X tremor/ataxia syndrome (FXTAS). *Neuropathology.* 2009 Jun;29(3):280-4.
- [58] Hunsaker MR, Greco CM, Spath MA, Smits AP, Navarro CS, Tassone F, Kros JM, Severijnen LA, Berry-Kravis EM, Berman RF, Hagerman PJ, Willemsen R, Hagerman RJ, Hukema RK. Widespread non-central nervous system organ pathology in fragile X premutation carriers with fragile X-associated tremor/ataxia syndrome and CGG knock-in mice. *Acta Neuropathol.* 2011 Oct;122(4):467-79.
- [59] Todd PK, Oh SY, Krans A, He F, Sellier C, Frazer M, Renoux AJ, Chen KC, Scaglione KM, Basrur V, Elenitoba-Johnson K, Vonsattel JP, Louis ED, Sutton MA, Taylor JP, Mills RE, Charlet-Berguerand N, Paulson HL. CGG repeat-associated translation mediates neurodegeneration in fragile X tremor ataxia syndrome. *Neuron.* 2013 May 8;78(3):440-55.
- [60] Ladd PD, Smith LE, Rabaia NA, Moore JM, Georges SA, Hansen RS, Hagerman RJ, Tassone F, Tapscott SJ, Filippova GN. An antisense transcript spanning the CGG repeat region of FMR1 is upregulated in premutation carriers but silenced in full mutation individuals. *Hum. Mol. Genet.* 2007 Dec 15;16(24):3174-87.

- [61] Mailick MR, Hong J, Rathouz P, Baker MW, Greenberg JS, Smith L, Maenner M. Low-normal FMR1 CGG repeat length: phenotypic associations. *Front Genet.* 2014 Sep 9;5:309.
- [62] Chen LS, Tassone F, Sahota P, Hagerman PJ. The (CGG)_n repeat element within the 5' untranslated region of the FMR1 message provides both positive and negative cis effects on in vivo translation of a downstream reporter. *Hum. Mol. Genet.* 2003 Dec 1;12(23):3067-74. Epub 2003 Sep 30.
- [63] Ramocki MB, Zoghbi HY. Failure of neuronal homeostasis results in common neuropsychiatric phenotypes. *Nature.* 2008 Oct 16;455(7215):912-8. Review.
- [64] Hunter JE, Leslie M, Novak G, Hamilton D, Shubeck L, Charen K, Abramowitz A, Epstein MP, Lori A, Binder E, Cubells JF, Sherman SL. Depression and anxiety symptoms among women who carry the PM: impact of raising a child with fragile X syndrome is moderated by CRHR1 polymorphisms. *Am. J. Med. Genet. B Neuropsychiatr Genet.* 2012 Jul;159B(5):549-59.
- [65] Silva F, Rodriguez-Revenga L, Madrigal I, Alvarez-Mora MI, Oliva R, Milà M. High apolipoprotein E4 allele frequency in FXTAS patients. *Genet Med.* 2013 Aug;15(8):639-42.
- [66] Lozano R, Hagerman RJ, Duyzend M, Budimirovic DB, Eichler EE, Tassone F. Genomic studies in fragile X premutation carriers. *J. Neurodev Disord.* 2014;6(1):27.
- [67] Hamlin A, Liu Y, Nguyen DV, Tassone F, Zhang L, Hagerman RJ. Sleep apnea in fragile X premutation carriers with and without FXTAS. *Am. J. Med. Genet. B Neuropsychiatr Genet.* 2011;156B (8):923–928.
- [68] Leehey MA. Fragile X-associated tremor/ataxia syndrome: clinical phenotype, diagnosis, and treatment. *J. Investig. Med.* 2009 Dec;57(8):830-6. Review.
- [69] Rodriguez-Revenga L, Madrigal I, Alegret M, Santos M, Milà M. Evidence of depressive symptoms in fragile-X syndrome premutated females. *Psychiatr Genet.* 2008 Aug;18(4):153-5.
- [70] Bourgeois JA, Seritan AL, Casillas EM, Hessel D, Schneider A, Yang Y, Kaur I, Cogswell JB, Nguyen DV, Hagerman RJ. Lifetime prevalence of mood and anxiety disorders in fragile X premutation carriers. *J. Clin. Psychiatry.* 2011 Feb;72(2):175-82.
- [71] Hunter JE, Leslie M, Novak G, Hamilton D, Shubeck L, Charen K, Abramowitz A, Epstein MP, Lori A, Binder E, Cubells JF, Sherman SL. Depression and anxiety symptoms among women who carry the FMR1 premutation: impact of raising a child with fragile X syndrome is moderated by CRHR1 polymorphisms. *Am. J. Med. Genet. B Neuropsychiatr Genet.* 2012 Jul;159B(5):549-59.
- [72] Chonchaiya W, Au J, Schneider A, Hessel D, Harris SW, Laird M, Mu Y, Tassone F, Nguyen DV, Hagerman RJ. Increased prevalence of seizures in boys who were probands with the FMR1 premutation and co-morbid autism spectrum disorder. *Hum. Genet.* 2012; 131 (4):581–589.
- [73] Farzin F, Perry H, Hessel D, Loesch D, Cohen J, Bacalman S, Gane L, Tassone F, Hagerman P, Hagerman R. Autism spectrum disorders and attention-deficit/hyperactivity disorder in boys with the fragile X premutation. *J. Dev. Behav. Pediatr.* 2006; 27 (2 Suppl):S137–144.
- [74] Kraan CM, Hocking DR, Georgiou-Karistianis N, Metcalfe SA, Archibald AD, Fielding J, Trollor J, Bradshaw JL, Cohen J, Cornish KM. Impaired response inhibition is associated with self-reported symptoms of depression, anxiety, and ADHD in female

- FMR1 premutation carriers. *Am. J. Med. Genet. B Neuropsychiatr Genet.* 2014 Jan;165(1):41-51.
- [75] Bailey DB Jr, Raspa M, Olmsted M, Holiday DB. Co-occurring conditions associated with FMR1 gene variations: findings from a national parent survey. *Am. J. Med. Genet. A.* 2008; 146A (16): 2060–2069.
- [76] Hagerman RJ, Coffey SM, Maselli R, Soontarapornchai K, Brunberg JA, Leehey MA, Zhang L, Gane LW, Fenton-Farrell G, Tassone F, Hagerman PJ. Neuropathy as a presenting feature in fragile X-associated tremor/ataxia syndrome. *Am. J. Med. Genet. A.* 2007 Oct 1;143A(19):2256-60.

Chapter 92

GENETIC COUNSELING OF *FMR1*

I. Madrigal

Biochemistry and Molecular Genetics Department,
Hospital Clínic and IDIBAPS
Centre for Biomedical Research on Rare Diseases (CIBERER),
ISCIH, Barcelona, Spain
IDIBAPS (Institut d'Investigacions Biomèdiques August Pi i Sunyer),
Barcelona, Spain

ABSTRACT

Fragile X syndrome is the most common form of inherited intellectual disability, with an estimated incidence of 1 in 4,000 males and 1 in 6,000 females. Each diagnosis of an *FMR1* mutation has far reaching clinical and reproductive implications for the extended family. Until now genetic counseling was offered based on the expansion risk in premutation carrier women, but the description of *FMR1*-associated disorders has increased the complexity of the genetic counseling for FXS families, especially *FMR1* premutation carriers. Male individuals carrying full mutated alleles present with intellectual disability while the penetrance is incomplete in females (30-50%). Premutation allele carriers are intellectually unaffected, but several *FMR1* premutation-related disorders have been described. The most prevalent are fragile X-associated primary ovarian insufficiency and fragile X-associated tremor/ataxia syndrome, but behavioral features such as impaired executive function, social deficits or anxiety have also been related to several *FMR1* premutation carriers. Premutation women have 50% of risk of transmitting a premutation/full mutation allele to their offspring, depending on the CGG expansion repeat and the presence of AGG interruptions. On the contrary, premutation men carriers will only transmit the premutation allele to their daughters. Some issues such as risk assessment for intermediate alleles and the clinical prognosis for females with full mutations still remain challenging. Genetic counselors must have an updated and solid understanding of this genetic condition and the *FMR1*-associated disorders in order to cover all the counseling aspects of these disorders.

Keywords: *FMR1*, FXS, premutation alleles, full mutation alleles, genetic counseling, intermediate alleles

INTRODUCTION

The World Health Organization defines Genetic Counseling as the process through which trained genetic counselors inform about the genetic aspects of a particular genetic disease to those who are at an increased risk of developing the disorder or of passing it on to their unborn offspring. A genetic counselor must provide accurate information to the patient and their families about the clinical consequences of the conditions, the inheritance of illnesses and their risks of recurrence, management and/or treatment, quality of life, life expectancy and social aspects. These professionals must then work in a wide variety of fields such as general genetics, prenatal care or family planning. The counseling process consists of two stages: pre and the post-test genetic counseling. In pre-test genetic counseling individuals are informed about the purpose of the test, the clinical consequences of the conditions, including the phenotypic features, inheritance patterns, the reliability and limitations of the test and the possible psychological impacts to the counselees and their relatives.

In post-test genetic counseling, after disclosure of the test results, the counselor should focus on the emotional impact on the counselee and the relatives involved. The counselees must be informed about implications to them and their close relatives, including management and/or treatment options, quality of life, life expectancy and available reproductive choices. Counselors must ensure the privacy and confidentiality of the results.

GENETIC COUNSELING IN *FMR1*-ASSOCIATED DISORDERS

Fragile X syndrome (FXS, #300624) is the most common form of inherited intellectual disability (ID), with an estimated incidence of 1 in 4,000 males and 1 in 6,000 females [1]. The molecular basis of this syndrome is mainly the expansion of an unstable CGG repeat in the 5' untranslated region of the fragile X syndrome gene (*FMR1*). The polymorphic CGG repeat of the *FMR1* gene is distributed in the population in four allelic classes according to repeat length (Table 1). Alleles ranging from 6 to 44 CGG repeats are the most common in the general population and have a stable transmission to the next generation. Alleles within the 45–54 CGG repeat range are called intermediate alleles and most are stable. Individuals carrying intermediate alleles are phenotypically normal. Alleles about 55–200 repeats are called premutation alleles. These alleles are associated with a significant elevation of *FMR1* mRNA levels [2-3] and are highly unstable during maternal transmission. Premutation carrier individuals do not present with ID, but they have a risk of developing fragile X-associated tremor/ataxia syndrome (FXTAS), premature ovarian insufficiency (FXPOI) or other psychological symptoms [4-7]. Finally, expansion of the repeat region to more than 200 CGG trinucleotide sequences is called full mutation. *FMR1* is an X-linked gene regulated by methylation of the promoter during X inactivation in somatic cells of females. This leads to gene silencing and insufficient synthesis of the *FMR1* protein (FMRP). Expansions over 200 CGGs lead to hypermethylation of the CpG island resulting in non-expression of the *FMR1* gene and absence of the FMRP. The lack of FMRP is the direct cause of the FXS phenotype [8]. Besides CGG repeat expansions, it has been reported that almost 1% of individuals with FXS have a partial or full deletion, or point mutation of the *FMR1* gene [9-12].

Until now genetic counseling was offered based on the expansion risk in premutation carrier women, but the description of *FMR1*-associated disorders has increased the complexity of the genetic counseling for FXS families, particularly to *FMR1* premutation carriers. Furthermore, some issues such as risk assessment for intermediate alleles and the clinical prognosis for females with full mutations still remain challenging.

Most of the families who first receive a diagnosis of FXS have no prior knowledge of *FMR1*-premutation disorders, and it is essential that individuals under study receive pre and post genetic counseling, including information about possible outcomes and the implications of carrying full, premutation or intermediate alleles. Genetic counseling of FXS requires attention to a wide range of clinical manifestations including developmental, neurodegenerative, and reproductive symptoms that may vary in age of onset and severity. Counselors must have a solid understanding of this genetic condition, including the trinucleotide repeat instability and the phenotypic variability. Table 1 shows associated *FMR1* allele phenotypes.

Table 1. *FMR1* alleles and associated phenotypes

Allele	CGG repeat range	Phenotype
Normal	5-44	Normal
Intermediate	45-54	Normal
Premutation	55–200	Normal, FXTAS, FXPOI and others
Full Mutation	>200 (methylated)	ID in 100% males and reduced penetrance in females (30-50%)

ASSESSMENT RISK BASED ON CGGs EXPANSION

Normal Range and Intermediate Alleles

Alleles ranging from 6 to 44 CGG repeats, the most common in the general population, and most of the intermediate alleles have stable transmission to the next generation. Intermediate alleles may show some instability, including expansion to a full mutation in two generations [13-15].

Premutation Alleles

Premutation alleles are highly unstable during maternal transmission and may expand to a higher CGG repeat size or even to a full mutation in only one generation. Even expansion of paternal premutation alleles is possible; the expansion from a premutation allele to a full mutation has only been observed through female meioses. Thus, all the daughters from a premutation male will inherit the premutation allele and will not manifest FXS. Table 2 shows the risk of expansion of the different allele types. The expansion risk of a *FMR1* allele depends both on CGG repeat size and the presence of AGG interruptions. In 1994, Eichler *et al.* suggested that AGGs interspersed within the *FMR1* repeat region may be linked to repeat stability and the risk of expansion [16]. In the general population, almost 95% of alleles have

one or two AGG interruptions, which the most common allele pattern is two AGGs at positions 10 or 11 and 20 or 21 repeats. In contrast, alleles in FXS families contain no or few AGGs at the 5' end and they contain long stretches of uninterrupted CGGs at the 3' end [16,17]. Presumably maternal alleles with no AGG interruptions confer increased risk for unstable transmissions and thus, the inclusion of AGG genotype studies would be of benefit in clinical practice. AGG testing is expected to reassure premutation carriers with AGG interruptions while alerting premutation carriers with alleles without AGG interruptions who would have the highest risk for instability.

Table 2. Risk of expansion of *FMR1* alleles

Allele	CGG repeat range	Risk of expansion to a full mutation for females*
Normal	5-44	0%
Intermediate	45-54	0%
Premutation	55-59	3.70%
	60-69	5.30%
	70-79	31.10%
	80-89	57.80%
	90-99	80.10%
	>100	94-100%
Full Mutation	>200 (methylated)	

*[68-70].

Full Mutation Alleles

Full mutated males never transmit the full mutated allele to their daughters, only an allele in the premutation range. Full mutated females have a 50% of transmission risk of the full mutated allele to their offspring; all male descendants carrying the full mutation allele will be affected while a smaller percentage of women will present the syndrome.

GENETIC COUNSELING BASED ON *FMR1* ALLELES

Intermediate Alleles

Intermediate alleles present a 45–54 CGG repeat range [18] and have a frequency in the general population of 1/35 to 1/57 females [19,20]. Intermediate allele carriers do not manifest ID, FXTAS or FXPOI, although some groups have suggested associations between intermediate alleles and some disorders such as Parkinson's disease [21], primary ovarian insufficiency [22], and autism and cognitive disabilities [23,24]. However, these findings have not consistently been supported by other studies [25-27]. A frequent concern in FXS screening is the genetic counseling to intermediate allele carriers. Currently, the follow-up of individuals with no AGG interruptions is indicated, despite the low risk of expansion in the next generation.

Premutation Alleles

Individuals carrying premutation alleles are intellectually unaffected, but in the last decades several premutation-related disorders have been described. Among them the most prevalent are FXPOI and FXTAS [28,29]. In addition, behavioral features such as impaired executive function, social deficits or anxiety have been related to some *FMR1* premutation carriers [30]. Premutation alleles contain trinucleotide expansions in the range of 55 to 200 CGG repeats. In Western populations premutation frequencies in females range from 1/151 to 1/259, but there is some evidence of ethnic and racial variability. In Israel, for example, the premutation frequency is around 1/113 and in Taiwan 1/837 [20,31-33]. Some authors state that the different reproductive aptitudes of female premutation carriers and the differences in mean maternal age may contribute to this variation in premutation frequencies [34]. In males the rate ranges from 1/468 to 1/813 [20,35].

Men carrying the premutation allele must mainly be counseled about the risk of FXTAS. This syndrome presents with incomplete penetrance even among premutation carriers with identical CGG-repeat lengths. Approximately 50% of premutation carrier males will develop neurodegenerative symptoms of FXTAS after the age of 50 [36]. The first clinical signs of the syndrome typically appear when patients are in their 50s and 60s, with a mean tremor onset at approximately 60 years and ataxia onset at 62 years [37]. The risk and severity of the disorder appear to be related to the CGG repeat size, with higher risk in larger repeats [38]; nevertheless, a biomarker that predicts the apparition of FXTAS or the protective factors in asymptomatic carriers has still not been identified. Other clinical signs associated with premutation male carriers are neuroendocrine dysfunction, including testosterone deficiency [39], hypertension [40] or bowel and urinary incontinence [41].

Genetic counseling of premutation carrier women must be addressed to the risk for premature ovarian insufficiency, besides the risk for expansion and the FXTAS syndrome. In women, FXTAS is much less common than in men, as many as 16% will develop FXTAS symptoms, it presents a later age of onset and is milder in presentation. The risk for FXPOI in these women increases around 20% with an age of onset before the age of 40 years [28]. Other pathologies associated with females premutation carriers are psychiatric disorders such as depression, anxiety or mood disorders [42-44], migraine [45], immune-mediated disorders, particularly hypothyroidism (15.9%) [36], and fibromyalgia (25%) [7,46,47]. The latter two in particular are even more common among women with FXTAS, with a frequency of 43% and 50%, respectively [7].

Full Mutation Alleles

Full mutations are responsible for FXS, a spectrum of clinical features which includes physical, cognitive and behavioral aspects. All males carrying the full mutation will present with mild to severe intellectual disability and exhibit a variety of maladaptive behaviors overlapping those described for autism spectrum disorders [48,49]. On the contrary, only 50%-70% of women manifest FXS symptoms, albeit in a milder form than men, and some appear to be completely unaffected or exhibit minor neurobehavioral features [50-52]. Despite a common genetic etiology, the clinical presentation of FXS is variable and this variability is related to residual levels of the FMRP due to the presence of CGG expansion size mosaicism, different

methylation levels of the full mutation allele and X-inactivation that leads to a differential pattern of FMRP expression within tissues [51].

On the other hand, there are full mutation carrier males with a standardized IQ score of 70 or higher who are known as high functioning males. In normal males, the 5' CpG island containing the CGG trinucleotide repeat is not methylated, whereas a CGG triplet repeat expands to more than ~200 repeats, this site is methylated and the *FMR1* is transcriptionally silenced [53]. Many mildly affected individuals show mosaic methylation at the *FMR1* promoter.

Regarding their offspring, their daughters will inherit an allele in the premutation range. High functioning males do not present with intellectual disability but recently, FXTAS has been described in a high functioning male [54] and in the case of an unmethylated mosaic (premutation-full mutation) male [55]. The latter suggests that the definition of FXTAS should also include those cases with an expanded allele the size and lack of methylation of which leads to RNA toxicity.

***FMR1* Point Mutations/Deletions**

It has been suggested that patients with a clinical FXS-like phenotype but not carrying the *FMR1* gene full mutation should be screened for *FMR1* mutations [56,57]. It is estimated that *FMR1* point mutations or deletions are responsible for up 1% of FXS cases. Nevertheless, the prevalence of mutations in the *FMR1* coding region is still not well known since the standard FXS protocols only comprise the study of the CGG repeats size. These mutations can be *de novo* or inherited from a carrier mother. A male carrying a point mutation or deletion will always have carrier daughters, who might be affected depending on the X-chromosome inactivation. Carrier females have 50% of risk of transmitting the mutated allele. Within this 50%, all males will be affected and females will be variably affected, depending on the X-inactivation.

REPRODUCTIVE OPTIONS FOR PREMUTATION/FULL MUTATION CARRIERS

Individuals at risk for passing on FXS mutations to their offspring have a variety of pre- and postconception options available. Regardless of the option they choose, women must be advised of any contraindication of these procedures [58]. When offering the reproductive options to a couple at risk, a fertility issue has to be considered since the risk of POF has significant reproductive implications. First of all, the onset of POF is difficult to predict. Secondly, the subtle endocrine perturbations (elevated follicle-stimulating hormone levels) observed in these women decreases the efficiency of ovarian stimulation required for preimplantational diagnosis and finally, there is evidence that premutation carriers may have hormonal changes suggestive of early ovarian aging despite regular menstrual cycles (See Chapter 5).

Prenatal Diagnosis for FXS

Prenatal diagnosis for FXS is possible using chorionic villi or amniotic fluid cells and it depends on several factors, including the physician performing the procedure and patient preference. When the establishment of the sizing of the expansion is sufficient to determine the genotype, the prenatal diagnosis on chorionic villi allows performing a therapeutic abortion at early stages in the pregnancy. On the contrary, the methylation pattern of a full mutation is established after the 14th week of pregnancy, thus the use of a methylation-sensitive method (i.e Southern Blot) is not suitable for early prenatal diagnosis on DNA from chorionic villi [59]. Moreover, the risk of miscarriage may be lower on amniotic fluid cell sampling.

Table 3. Genetic counseling in *FMR1*

Progenitor allele and gender	% expected offspring	Offspring outcome	
		Males	Females
Normal male	No risk	Normal	Normal
Normal female	No risk	Normal	Normal
IA carrier male	Females IA range	Normal	Normal
IA carrier female	50% NA range 50% IA range	Normal	Normal
PM carrier male	Females PM range	Normal	Normal, FXTAS, FXPOI and others
PM carrier female	50% NA range 50%* PM/ FM*	Normal Normal, FXTAS, and others 100% ID	Normal Normal, FXTAS, FXPOI and others 30-50% ID
FM carrier male	females PM range	Normal	Normal, FXTAS, FXPOI and others
FM carrier female	50% NA range 50% FM	Normal 100% ID	Normal 30-50% ID

NA: normal allele, IA: intermediate allele; PM: premutation allele; FM: full mutation allele.

*percentage varies according to the expansion risk indicated in Table 2.

Prenatal diagnosis should be offered to women with premutation or full mutations, but it is not intended for the pregnant partner of a premutation male carrier. These males, nevertheless, should receive genetic counseling about phenotypic risk to their daughters, who will inherit the premutation allele. Neither is prenatal testing intended for intermediate allele carriers. As indicated before, there are no reports of intermediate alleles expanding to full mutations in a single generation. The risk for a female premutation carrier of transmitting the expansion is 50%, and the risk for the maternal premutation to expand to full mutation is proportional to its size and AGG content. Table 3 summarizes the possible outcomes depending on the carrier state of each parent.

The identification of a female fetus with the fragile X full mutation entails a significant challenge for professionals. The probability that a full mutation carrier fetus is affected is around 50-70%, but at present there are no biomarkers to predict this affectation. The diagnosis of a FXS patient requires extending the study to the mother in order to confirm her premutation carrier status.

Preimplantational Diagnosis for FXS

Preimplantation genetic diagnosis (PGD) is a technique based on the genetic analysis of an embryo obtained through in vitro fecundation. The most common option is to perform the genetic study on a day-3 embryo which is then transferred to the uterus. This approach increases the risk of embryo viability due to the biopsy. Further, the possibility of embryo mosaicism has been described at the cleavage stage and self-correction of aneuploidies between the cleavage and blastocyst stages [60]. Polar body biopsies can be used to avoid any misdiagnosis due to embryo mosaicism, although only maternal genetic information is obtained. The third option is to biopsy trophoctoderm cells from blastocysts (5-6 day embryo), which is less invasive and has higher concordance between inner cell mass and trophoctoderm cells [61]. PGD has several advantages over prenatal diagnosis. The diagnosis is performed in the embryo, avoiding the parental stress and the emotional trauma of a termination of pregnancy.

Offspring Renouncement and Adoption

These two options are drastic and are usually rejected by the couples. Therefore they are currently seldom observed as the first choice option by couples at risk that wish to prevent the birth of an affected child.

Donor Germline Cell

Another possible option is egg or sperm donation for female and male mutation carriers, respectively. Based on our experience, around 10% of FXS families chose for a germline cell donation [62]. At present, it is a very good option for premutation carrier females who have been involved in previous prenatal diagnosis and termination of pregnancies. Moreover, given the relatively high prevalence of FXS in the general population, potential gamete donors should be tested for this syndrome.

***FMR1* TESTING GUIDELINES**

Several scientific societies such as the European Molecular genetics Quality Network (EMQN), the American College of Medical Genetics (ACMG), the National Society of genetic Counselors (NSGC) and the American College of Obstetrics and Gynecology (ACOG) have published best practice guidelines for the molecular genetic testing and diagnosis of FXS and other fragile X-associated disorders [18,63-66]. The best practice guidelines for genetic analysis and reporting in FXS, FXPOI, and FXTAS are listed in Table 4. Appropriate *FMR1* molecular testing is very important for optimal genetic counseling in the fragile X-associated disorders due to the particular pattern and transmission of the CGG repeat. Although not being endorsed by current guidelines, screening women without known risk factors to FXS is increasingly being offered. In 2005 Musci and Caughey demonstrated the efficacy and cost effectiveness of prenatal population-based fragile X carrier screening [67].

Table 4. *FMR1* mutation testing recommendations***FMR1* mutation testing is recommended for*:**

Individuals of either sex with intellectual disability, developmental delay or autism
 Individuals who have a family history of FXS
 Women with a family history of *FMR1*-related disorders, including FXPOI
 Women with reproductive or fertility problems associated with elevated levels of follicle stimulating hormone (FSH)
 Individuals with late onset tremor or cerebellar ataxia of unknown origin
 Given the relatively high prevalence of FXS in the general population, potential gamete donors should be tested for this syndrome

*Sherman, 2005; ACOG, 20010.

CONCLUSION

When first described, FXS was a rare X-linked disease responsible for intellectual disabilities in males and females in a lesser percentage. Research into *FMR1* has brought to light the molecular and phenotypic complexity of *FMR1*-related diseases. The first issue to consider is that a FXS patient will always have a premutation carrier mother who should receive genetic counseling. Premutation carrier females have a 50% risk of transmitting a premutation/full mutated allele to their offspring, unlike carrier males who will only transmit the premutation allele. The expansion risk in females depends on the CGG repeats number and the AGG interruptions. Lastly, premutation carriers are at risk not only of having FXS offspring (females) but also of FXPOI (females), FXTAS and other disorders. The complexity of the issues surrounding genetic testing and the management of *FMR1*-associated disorders has increased, and investigation related to this gene covers many aspects, including molecular factors, epigenetic factors, emotional issues or targeted pharmaceuticals. As the knowledge regarding disease causing mechanisms evolves, genetic counselors must have an updated and solid understanding of this genetic condition and the *FMR1*-associated disorders in order to cover all the counseling aspects of these syndromes.

REFERENCES

- [1] Hagerman RJ. The physical and behavioural phenotype, in: RJ Hagerman, P Hagerman (Eds.), *Fragile-X Syndrome: Diagnosis, Treatment and Research*, The Johns Hopkins University Press, Baltimore, 2002, pp. 3–109.
- [2] Tassone F, Hagerman RJ, Loesch DZ, Lachiewicz A, Taylor AK, Hagerman PJ. Fragile X males with unmethylated, full mutation trinucleotide repeat expansions have elevated levels of *FMR1* messenger RNA. *Am. J. Med. Genet.* 2000a;94:232–236.
- [3] Tassone F, Hagerman RJ, Taylor AK, Gane LW, Godfrey TE, Hagerman PJ. Elevated levels of *FMR1* mRNA in carrier males: A new mechanism of involvement in the fragile-X syndrome. *Am. J. Hum. Genet.* 2000b;66:6-15.
- [4] Hagerman RJ, Leehey M, Heinrichs W, Tassone F, Wilson R, Hills J, Grigsby J, Gage B, Hagerman PJ. Intention tremor, parkinsonism, and generalized brain atrophy in male carriers of fragile X. *Neurology* 2001;57:127-130.

- [5] Jacquemont S, Hagerman RJ, Leehey M, Grigsby J, Zhang L, Brunberg JA, Greco C, Des Portes V, Jardini T, Levine R, Berry-Kravis E, Brown WT, Schaeffer S, Kissel J, Tassone F, Hagerman PJ. Fragile X premutation tremor/ataxia syndrome: molecular, clinical, and neuroimaging correlates. *Am. J. Hum. Genet.* 2003;72:869-878.
- [6] Hessler D, Tassone F, Loesch DZ, Berry-Kravis E, Leehey MA, Gane LW, Barbato I, Rice C, Gould E, Hall DA, Grigsby J, Wegelin JA, Harris S, Lewin F, Weinberg D, Hagerman PJ, Hagerman RJ. Abnormal elevation of FMR1 mRNA is associated with psychological symptoms in individuals with the fragile X premutation. *Am. J. Med. Genet. B Neuropsychiatr. Genet.* 2005;139B:115-21.
- [7] Coffey SM, Cook K, Tartaglia N, Tassone F, Nguyen DV, Pan R, Bronsky HE, Yuhass J, Borodyanskaya M, Grigsby J, Doerflinger M, Hagerman PJ, Hagerman RJ. Expanded clinical phenotype of women with the FMR1 premutation. *Am. J. Med. Genet. A.* 2008;146A:1009-16.
- [8] de Vries BB, Mohkamsing S, van den Ouweland AM, Mol E, Gelsema K, van Rijn M, Tibben A, Halley DJ, Duivenvoorden HJ, Oostra BA, Niermeijer MF. Screening with the FMR1 protein test among mentally retarded males. *Hum. Genet.* 1998;103:520-522.
- [9] De Boulle K, Verkerk AJ, Reyniers E, Vits L, Hendrickx J, Van Roy B, Van den Bos F, de Graaff E, Oostra BA, Willems PJ. A point mutation in the FMR-1 gene associated with fragile X mental retardation. *Nat. Genet.* 1993;3:31-5.
- [10] Lugenbeel KA, Peier AM, Carson NL, Chudley AE, Nelson DL. Intragenic loss of function mutations demonstrate the primary role of FMR1 in fragile X syndrome. *Nat. Genet.* 1995;10:483-5.
- [11] Strom CM, Crossley B, Redman JB, Buller A, Quan F, Peng M, McGinnis M, Fenwick RG Jr, Sun W. Molecular testing for fragile X syndrome: lessons learned from 119,232 tests performed in a clinical laboratory. *Genet Med. Off J. Am. Coll. Med. Genet* 2007;9:46e51.
- [12] Handt M, Epplen A, Hoffjan S, Mese K, Epplen JT, Dekomien G. Point mutation frequency in the FMR1 gene as revealed by fragile X syndrome screening. *Mol. Cell Probes.* 2014 Aug 27. pii: S0890-8508(14)00042-5. doi: 10.1016/j.mcp.2014.08.003.
- [13] Terracciano A, Pomponi MG, Marino GM, Chiurazzi P, Rinaldi MM, Dobosz M, Neri G. Expansion to full mutation of a FMR1 intermediate allele over two generations. *Eur. J. Hum. Genet.* 2004;12:333-336.
- [14] Zuniga A, Juan J, Mila M, Guerrero A. Expansion of an intermediate allele of the *FMR1* gene in only two generations. *Clin. Genet* 2005;68:471-473.
- [15] Fernandez-Carvajal I, Lopez Posadas B, Pan R, Raske C, Hagerman PJ, Tassone F. Expansion of an FMR1 grey-zone allele to a full mutation in two generations. *J. Mol. Diagn* 2009;11:306-310.
- [16] Eichler EE, Holden JJ, Popovich BW, Reiss AL, Snow K, Thibodeau SN, Richards CS, Ward PA, Nelson DL. Length of uninterrupted CGG repeats determines instability in the FMR1 gene. *Nat. Genet.* 1994;8:88-94.
- [17] Eichler EE, Macpherson JN, Murray A, Jacobs PA, Chakravarti A, Nelson DL. Haplotype and interspersed analysis of the FMR1 CGG repeat identifies two different mutational pathways for the origin of the fragile X syndrome. *Hum. Mol. Genet.* 1996;5:319-30.
- [18] Sherman S, Pletcher BA, Driscoll DA. Fragile X syndrome: diagnostic and carrier testing. *Genet Med.* 2005;7:584-7.

- [19] Cronister A, Teicher J, Rohlfs EM, Donnenfeld A, Hallam S. Prevalence and instability of fragile X alleles: implications for offering fragile X prenatal diagnosis. *Obstet. Gynecol.* 2008;111:596-601.
- [20] Seltzer MM, Baker MW, Hong J, Maenner M, Greenberg J, Mandel D. Prevalence of CGG expansions of the FMR1 gene in a US population-based sample. *American Journal of Medical Genetics Part B* 2012;159B:589-597.
- [21] Loesch DZ, Khaniani MS, Slater HR, Rubio JP, Bui QM, Kotschet K, D'Souza W, Venn A, Kalitsis P, Choo AK, Burgess T, Johnson L, Evans A, Horne M. Small CGG repeat expansion alleles of FMR1 gene are associated with parkinsonism. *Clin. Genet.* 2009a;76:471-6.
- [22] Bodega B, Bione S, Dalprà L, Toniolo D, Ornaghi F, Vegetti W, Ginelli E, Marozzi A. Influence of intermediate and uninterrupted FMR1 CGG expansions in premature ovarian failure manifestation. *Hum. Reprod.* 2006;21:952-7.
- [23] Aziz M, Stathopulu E, Callias M, Taylor C, Turk J, Oostra B, Willemsen R, Patton M. Clinical features of boys with fragile X premutations and intermediate alleles. *Am. J. Med. Genet. B Neuropsychiatr. Genet.* 2003;121B:119-27.
- [24] Loesch DZ, Godler DE, Khaniani M, Gold E, Gehling F, Dissanayake C, Burgess T, Tassone F, Huggins R, Slater H, Choo KH. Linking the FMR1 alleles with small CGG expansions with neurodevelopmental disorders: preliminary data suggest an involvement of epigenetic mechanisms. *Am. J. Med. Genet. A.* 2009b;149A:2306-10.
- [25] Ennis S, Murray A, Youings S, Brightwell G, Herrick D, Ring S, Pembrey M, Morton NE, Jacobs PA. An investigation of FRAXA intermediate allele phenotype in a longitudinal sample. *Ann. Hum. Genet.* 2006;70:170-80.
- [26] Bennett CE, Conway GS, Macpherson JN, Jacobs PA, Murray A. Intermediate sized CGG repeats are not a common cause of idiopathic premature ovarian failure. *Hum. Reprod.* 2010;25:1335-8.
- [27] Madrigal I, Xunclà M, Tejada MI, Martínez F, Fernández-Carvajal I, Pérez-Jurado LA, Rodríguez-Revenga L, Milà M. Intermediate FMR1 alleles and cognitive and/or behavioural phenotypes. *Eur. J. Hum. Genet.* 2011;19:921-3.
- [28] Sherman, S. L. Premature ovarian failure in the fragile X syndrome. *Am. J. Med. Genet.* 2000;97:189-194.
- [29] Garcia-Arocena, D., Hagerman, P. J.. Advances in understanding the molecular basis of FXTAS. *Human Molecular Genetic* 2010;19:R83-89.
- [30] Bailey DB Jr, Raspa M, Olmsted M, Holiday DB. Co-occurring conditions associated with FMR1 gene variations: findings from a national parent survey. *Am. J. Med. Genet A.* 2008;146A:2060-9.
- [31] Rousseau, F., Rouillard, P., Khandjian, E. W., & Morgan, K. Prevalence of carriers of premutation-size alleles of the FMR1 gene and implications for the population genetics of the fragile X syndrome. *American Journal of Human Genetics* 1995;57:1006-1018.
- [32] Toledano-Alhadeif H, Basel-Vanagaite L, Magal N, Davidov B, Ehrlich S, Drasinover V, Taub E, Halpern GJ, Ginott N, Shohat M. Fragile-X carrier screening and the prevalence of premutation and full-mutation carriers in Israel. *American Journal of Human Genetics* 2001;69:352-360.
- [33] Tzeng, C. C., Tsai, L. P., Hwu, W. L., Lin, S. J., Chao, M. C., Jong, Y. J., Chu, S. Y., Chao, W. C., and Lu, C. L. Prevalence of the FMR1 mutation in Taiwan assessed by

- large-scale screening of newborn boys and analysis of DXS548-FRAXAC1 haplotype. *Am. J. Med. Genet.* 2005;133A:37–43.
- [34] Genereux DP, Laird CD. Why do fragile X carrier frequencies differ between Asian and non-Asian populations? *Genes Genet. Syst.* 2013;88:211–24.
- [35] Dombrowski C, Lévesque S, Morel ML, Rouillard P, Morgan K, Rousseau F. Premutation and intermediate-size FMR1 alleles in 10572 males from the general population: loss of an AGG interruption is a late event in the generation of fragile X syndrome alleles. *Human Molecular Genetics* 2002;11:371–378.
- [36] Rodriguez-Revenga L, Madrigal I, Pagonabarraga J, Xunclà M, Badenas C, Kulisevsky J, Gomez B, Milà M. Penetrance of FMR1 premutation associated pathologies in fragile X syndrome families. *Eur. J. Hum. Genet.* 2009;17:1359–62.
- [37] Leehey MA, Berry-Kravis E, Min SJ, Hall DA, Rice CD, Zhang L, Grigsby J, Greco CM, Reynolds A, Lara R, Cogswell J, Jacquemont S, Hessler DR, Tassone F, Hagerman R, Hagerman PJ. Progression of tremor and ataxia in male carriers of the FMR1 premutation. *Mov. Disord.* 2007;22:203–6.
- [38] Leehey MA, Berry-Kravis E, Goetz CG, Zhang L, Hall DA, Li L, Rice CD, Lara R, Cogswell J, Reynolds A, Gane L, Jacquemont S, Tassone F, Grigsby J, Hagerman RJ, Hagerman PJ. FMR1 CGG repeat length predicts motor dysfunction in premutation carriers. *Neurology.* 2008;70:1397–402.
- [39] Greco CM, Soontrapornchai K, Wirojanan J, Gould JE, Hagerman PJ, Hagerman RJ. Testicular and pituitary inclusion formation in fragile X associated tremor/ataxia syndrome. *J. Urol.* 2007;177:1434–7.
- [40] Hamlin AA, Sukharev D, Campos L, Mu Y, Tassone F, Hessler D, Nguyen DV, Loesch D, Hagerman RJ. Hypertension in FMR1 premutation males with and without fragile X-associated tremor/ataxia syndrome (FXTAS). *Am. J. Med. Genet. A.* 2012;158A:1304–9.
- [41] Hagerman PJ, Hagerman RJ. Fragile X-associated tremor/ataxia syndrome--an older face of the fragile X gene. *Nat. Clin. Pract. Neurol.* 2007;3:107–12.
- [42] Rodriguez-Revenga L, Madrigal I, Alegret M, Santos M, Milà M. Evidence of depressive symptoms in fragile-X syndrome premutated females. *Psychiatr Genet.* 2008;18:153–5.
- [43] Bourgeois JA, Seritan AL, Casillas EM, Hessler D, Schneider A, Yang Y, Kaur I, Cogswell JB, Nguyen DV, Hagerman RJ. Lifetime prevalence of mood and anxiety disorders in fragile X premutation carriers. *J. Clin. Psychiatry.* 2011;72:175–82.
- [44] Hunter JE, Sherman S, Grigsby J, Kogan C, Cornish K. Capturing the fragile X premutation phenotypes: a collaborative effort across multiple cohorts. *Neuropsychology.* 2012;26:156–64.
- [45] Au J, Akins RS, Berkowitz-Sutherland L, Tang HT, Chen Y, Boyd A, Tassone F, Nguyen DV, Hagerman R. Prevalence and risk of migraine headaches in adult fragile X premutation carriers. *Clin. Genet.* 2013;84:546–51.
- [46] Leehey MA, Legg W, Tassone F, Hagerman R. Fibromyalgia in fragile X mental retardation 1 gene premutation carriers. *Rheumatology (Oxford).* 2011;50:2233–6.
- [47] Rodriguez-Revenga L, Madrigal I, Blanch-Rubió J, Elurbe DM, Docampo E, Collado A, Vidal J, Carbonell J, Estivill X, Mila M. Screening for the presence of FMR1 premutation alleles in women with fibromyalgia. *Gene.* 2013;512:305–8.
- [48] Farzin F, Perry H, Hessler D, Loesch D, Cohen J, Bacalman S, Gane L, Tassone F, Hagerman P, Hagerman R. Autism spectrum disorders and attention-deficit/hyperactivity

- disorder in boys with the fragile X premutation. *J. Dev. Behav. Pediatr.* 2006;27:S137-44.
- [49] Boyle L, Kaufmann WE. The behavioral phenotype of FMR1 mutations. *Am. J. Med. Genet. C. Semin. Med. Genet.* 2010 Nov 15;154C:469-76.
- [50] Cronister A, Schreiner R, Wittenberger M, Amiri K, Harris K, Hagerman RJ. Heterozygous fragile X female: historical, physical, cognitive, and cytogenetic features. *Am. J. Med. Genet.* 1991;38:269-74.
- [51] de Vries BB, Jansen CC, Duits AA, Verheij C, Willemsen R, van Hemel JO, van den Ouweland AM, Niermeijer MF, Oostra BA, Halley DJ. Variable FMR1 gene methylation of large expansions leads to variable phenotype in three males from one fragile X family. *J. Med. Genet.* 1996;33:1007-10.
- [52] Keysor CS, Mazzocco MM. A developmental approach to understanding Fragile X syndrome in females. *Microsc. Res. Tech.* 2002;57:179-86.
- [53] Hornstra I.K., Nelson D.L., Warren S.T., Yang T.P. High resolution methylation analysis of the FMR-1 gene trinucleotide repeat region in Fragile X syndrome. *Hum. Mol. Genet.* 1993;2:1659-1665
- [54] Loesch DZ, Sherwell S, Kinsella G, Tassone F, Taylor A, Amor D, Sung S, Evans A. Fragile X-associated tremor/ataxia phenotype in a male carrier of unmethylated full mutation in the FMR1 gene. *Clin. Genet.* 2012;82:88-92
- [55] Santa María L, Pugin A, Allende MA, Aliaga S, Curotto B, Aravena T, Tang HT, Mendoza-Morales G, Hagerman R, Tassone F. FXTAS in an unmethylated mosaic male with fragile X syndrome from Chile. *Clin. Genet.* 2014;86:378-82
- [56] Grønskov K, Brøndum-Nielsen K, Dedic A, Hjalgrim H. A nonsense mutation in FMR1 causing fragile X syndrome. *Eur. J. Hum. Genet.* 2011;19:489e91.
- [57] Myrick LK, Nakamoto-Kinoshita M, Lindor NM, Kirmani S, Cheng X, Warren ST. Fragile X syndrome due to a missense mutation. *Eur. J. Hum. Genet.* 2014;22:1185-9.
- [58] Mor YS, Schenker JG. Ovarian Hyperstimulation Syndrome and Thrombotic Events. *Am. J. Reprod. Immunol.* 2014 Aug 22. doi: 10.1111/aji.12310.
- [59] Devys D, Biancalana V, Rousseau F, Boue J, Mandel JL, Oberle I. Analysis of full fragile X mutations in fetal tissues and monozygotic twins indicate that abnormal methylation and somatic heterogeneity are established early in development. *Am. J. Med. Genet* 1992;43:208–216.
- [60] Frumkin T, Malcov M, Yaron Y, Ben-Yosef D. Elucidating the origin of chromosomal aberrations in IVF embryos by preimplantation genetic analysis. *Mol. Cell Endocrinol.* 2008;282:112–9.
- [61] Capalbo A, Wright G, Themaat L, Elliott T, Rienzi L, Nagy ZP. FISH reanalysis of inner cell mass and trophectoderm samples of previously array-CGH screened blastocysts reveals high accuracy of diagnosis and no sign of mosaicism or preferential allocation. *Hum. Reprod.* 2013;28:2298-307.
- [62] Xunclà M, Badenas C, Domínguez M, Rodríguez-Revenga L, Madrigal I, Jiménez L, Soler A, Borrell A, Sánchez A, Milà M. Fragile X syndrome prenatal diagnosis: parental attitudes and reproductive responses. *Reprod. Biomed. Online.* 2010;21:560-5.
- [63] Kronquist KE, Sherman SL, Spector EB. Clinical significance of tri-nucleotide repeats in Fragile X testing: a clarification of American College of Medical Genetics guidelines. *Genet Med.* 2008;10:845-7.

- [64] American College of Obstetricians and Gynecologists Committee on Genetics. ACOG Committee Opinion No. 469: Carrier screening for fragile X syndrome. *Obstet. Gynecol.* 2010;116:1008-10.
- [65] Finucane B, Abrams L, Cronister A, Archibald AD, Bennett RL, McConkie-Rosell A. Genetic counseling and testing for FMR1 gene mutations: practice guidelines of the national society of genetic counselors. *J. Genet Couns.* 2012;21:752-60.
- [66] Biancalana V, Glaeser D, McQuaid S, Steinbach P. EMQN best practice guidelines for the molecular genetic testing and reporting of fragile X syndrome and other fragile X-associated disorders. *Eur. J. Hum. Genet.* 2014 Sep 17. doi: 10.1038/ejhg.2014.185
- [67] Musci TJ, Caughey AB. Cost-effectiveness analysis of prenatal population-based fragile X carrier screening. *Am. J. Obstet. Gynecol.* 2005;192:1905-12.
- [68] Nolin SL, Brown WT, Glicksman A, Houck GE Jr, Gargano AD, Sullivan A, Biancalana V, Brøndum-Nielsen K, Hjalgrim H, Holinski-Feder E, Kooy F, Longshore J, Macpherson J, Mandel JL, Matthijs G, Rousseau F, Steinbach P, Väisänen ML, von Koskull H, Sherman SL. Expansion of the fragile X CGG repeat in females with premutation or intermediate alleles. *Am. J. Hum. Genet.* 2003;72:454-64.
- [69] Rifé M, Badenas C, Quintó L, Puigoriol E, Tazón B, Rodríguez-Revenga L, Jiménez L, Sánchez A, Milà M. Analysis of CGG variation through 642 meioses in Fragile X families. *Mol. Hum. Reprod.* 2004;10:773-6.
- [70] Lévesque S, Dombrowski C, Morel ML, Rehel R, Côté JS, Bussi res J, Morgan K, Rousseau F. Screening and instability of FMR1 alleles in a prospective sample of 24,449 mother-newborn pairs from the general population. *Clin. Genet.* 2009;76:511-23.

Chapter 93

TREATMENT OF FRAGILE X SPECTRUM: FXS, FXTAS AND FXPOI

F. J. Ramos¹ and M. P. Ribate²

¹Unidad de Genética-Pediatría, GCV-CIBERER,
Hospital Clínico Universitario “Lozano Blesa”,

Facultad de Medicina, Universidad de Zaragoza, Spain

²Facultad de Ciencias de la Salud, Universidad San Jorge,
Zaragoza, Spain

ABSTRACT

Fragile X Spectrum includes three different clinical conditions Fragile X Syndrome (FXS), Fragile X-Associated Tremor and Ataxia (FXTAS), and Fragile X-Associated Premature Ovarian Insufficiency (FXPOI). Treatment of FXS is mainly symptomatic and it is addressed to improve or make disappear some of the more dyscapacitating symptoms like hyperactivity, deficit attention disorder and behavioral problems of language anomalies. Clinical trials are in course focusing on new discovered therapeutical targets (i.e., mGluR). Treatment of FXTAS and FXPOI are also mainly symptomatic and should be individually prescribed and modified depending on the clinical evolution.

INTRODUCTION

Fragile X spectrum includes Fragile X Syndrome (FXS), Fragile X-Associated Tremor Ataxia Syndrome (FXTAS) and Fragile X-Associated Premature ovarian Insufficiency (FXPOI). All three conditions are associated with the CGG trinucleotide expansion in the FMR1 gene on the X chromosome: FXS with more than 200 CGGs (full mutation) and FXTAS and FXPOI with 50 to 200 CGGs (premutation). Patients with FXS and FXTAS are mostly males, although females can also be affected. All three conditions are treated symptomatically, combining pharmacological and non-pharmacological therapies directed at alleviate the main disabling symptoms present in each condition.

FRAGILE X SYNDROME

Fragile X syndrome (FXS) is the most common inherited cause of intellectual disability that affects approximately 1 in 4,000 males and 1 in 8,000 females. Its genetic cause was identified in 1991 and consisted of the expansion of the number of CGG trinucleotide repeats at the FMR1 gene, located on the distal region of the long arm of chromosome X. The normal gene has less than 55 CGGs but when it has more than 55 CGGs it tends to expand when transmitted by the mother to the next generation. Individuals who have between 55- 200 CGGs (premutation) are considered carriers, and those with more than 200 CGGs (full mutation) and have a hypermethylated gene are affected. This methylation leads to the silencing of the FMR1 gene and the subsequent absence, partial or complete, of the FMRP (Fragile X Mental Retardation Protein), considered as the ultimate cause of FXS [1]. The vast majority of FXS males with the full mutation develop intellectual disabilities, which occurs in only 25% of females due to the presence of a normal X chromosome that produces FMRP. However, most females with FXS have learning disabilities and/or emotional and behavior problems [2].

The clinical presentation of FXS is highly variable. In addition to intellectual disability, affected patients have neurodevelopmental problems such as attention deficit hyperactivity disorder, disruptive behavior or autism spectrum disorders. The physical characteristics of FXS include dysmorphic facial features (long face, large and prominent ears and prominent chin), and macroorchidism after puberty [3].

The FMRP Protein and the Synapse

FMRP is expressed ubiquitously in many tissues, predominantly in brain neurons. There it binds to different nuclear mRNAs, including the FMR1 gene mRNA, repressing the translation of the target mRNA to postsynaptic dendritic spines, where its activity is important for synaptic plasticity during transport. Dendritic spines are specialized and rapidly changing protrusions in the neurons whose creation and elimination are essential for learning and memory processing [4]. Microscopic analysis of brain samples of patients with FXS and *Fmr1* knockout mice revealed no gross morphological abnormalities; however, in certain areas of the brain there are spindly dendritic spines, which are interpreted as immaturity and affect adversely the synaptic plasticity [5]. The discovery of a morphological spine phenotype indicates a possible defect in synaptic plasticity in FXS that could result in the intellectual disability phenotype. Whether the abnormal spine morphology is a cause or a consequence of altered signal transmission is currently unknown [6].

The Glutamate Receptor (mGluR) Hypothesis

In the brain, there are two main types of neurotransmitter receptors in the synaptic membrane: metabotropic receptors and ionotropic receptors. Metabotropic receptors (mGluRs) have eight different subtypes (mGluR1-8) which in turn are divided into three groups (I, II and III) according to sequence similarities and pharmacological properties. Group I includes mGluR1 and mGluR5 receptors which activates the Gq protein coupled and phospholipase C.

Group II receptors (mGluR2 and mGluR3) and Group III (mGluR4, mGluR6, and Glu7 Glu8) are coupled to the proteins Gi / Go by inhibiting adenylate cyclase. Ionotropic receptors are ligand-gated ion channels in which the incorporation of a specific ligand induces a conformational change leading to opening of the receiver [7]. The opened receiver allows ion flow across the cell membrane by changing the excitability of the neuron. The main dependent neuronal glutamate ionotropic receptors are: AMPA (isoxazolepropionic amino-3-hydroxy-5-methyl-4-acid), NMDA (N-methyl-D-aspartic acid), and kainate receptors, whose activation produces a fast excitatory neurotransmission [8].

The mechanisms of long-term potentiation (LTP) and long-term depression (LTD) are the most important regulatory mechanisms of neuronal synaptic plasticity, defined as long-lasting changes in synaptic strength accompanied by abnormalities in the size and morphology of dendritic spines. The mechanism of strengthening of LTP produces a connection between presynaptic and postsynaptic neurons. The LTD mechanism is opposite to that of LTP and results in the weakening of synapses, mainly due to a reduction of glutamate-gated ionotropic AMPA-alpha in the postsynaptic membrane [9].

The ‘theory of glutamate receptor (mGluR)’, first published in 2004, tried to explain many aspects of clinical symptoms present in patients with FXS and in the Fmr1-KO mouse, including a higher density and immaturity of dendritic spines compared to normal individuals/mice. This theory stated that internalization of the AMPA receptor is excessive in FXS and likely caused by stimulation of group I mGluRs (mGluR1 and mGluR5). Such stimulation would induce local mRNA translation, resulting in the synthesis of new protein that triggers the internalization of the AMPA receptor, essential for the long-term plasticity of dendritic spines. FMR1 mRNA is also present in the postsynaptic compartment and FMRP is synthesized locally after mGluR activation [10].

On the other hand, FMRP seems to negatively regulate the translation of proteins that are important for the internalization of the AMPA receptor. The exact mechanism by which the FMRP local translation represses mRNA remains unknown. Moreover, phosphorylation and dephosphorylation of FMRP seems play a crucial role in the signaling cascade induced by stimulation of mGluR5 receptors, which stimulate local protein synthesis at the synapses [11].

Since the formulation of the mGluR theory, clinical researchers in FXS have sought the support of the pharmaceutical industry to collaborate in identifying therapeutic strategies aimed at mGluR5 in order to at least partially reverse some of the symptoms of FXS.

The MPEP (2-methyl-6-(phenylethynyl)-pyridine), a negative modulator of mGluR5, can, *in vitro*, counter the excess of activity of the receptor and rescue the loss of AMPA receptors after the loss of FMRP. In FXS animal models, researches have got the pharmacological rescue of audiogenic seizures, behavioral phenotypes and alterations in dendritic spines by using several negative modulators of mGluR5. MGLuR theory has focused research into the basic mechanisms underlying FXS, prompting the search for new therapeutic strategies [12].

The GABA Hypothesis

Besides the mGluR’s hypothesis, researches have suggested that the signaling receptor gamma-aminobutyric acid (GABA) is altered in patients with FXS. GABA is the major inhibitory neurotransmitter in the central nervous system (CNS) and, as such, plays a key role in the modulation of neuronal activity in the brain. GABA mediates its action through two

distinct systems, the ionotropic GABA-A and the metabotropic GABA-B receptors [13]. Many patients with FXS have epilepsy and sleep disorders, conditions associated with the signaling pathway of the GABA receptor. Interestingly, the mRNA encoding GABA-A receptor subunits are targets of FMRP. The *Fmr1*-KO mice have decreased levels of mRNA and protein in various subunits of GABA-A receptor compared to normal animals [14]. These studies support a model in which the FMRP mRNA would regulate the stability of the GABA-A receptor subunits, preventing its degradation by joining these mRNAs *in vivo*. In addition to GABA-A receptors, GABA-B receptors could also be related to FXS, therefore turning into another potential therapeutic target. The GABA-B receptor agonists inhibit the presynaptic release of glutamate and the postsynaptic signaling cascade downstream of mGluR5 [15].

All those studies have demonstrated a dysfunction in the mRNA and protein expression of several subunits of GABA receptors, mechanisms that would be likely involved in the occurrence of the FXS phenotype.

THERAPEUTIC APPROACHES IN FXS

Currently, the treatment of patients with FXS is mainly symptomatic. The two most commonly used medications are stimulants used to improve attention and hyperactivity, and selective inhibitors of serotonin reuptake to reduce the risk of aggression associated with anxiety.

FXS patients are not only treated with pharmacological agents, but also benefit from behavioral therapies for improving emotional and language problems. As already demonstrated in the mouse FXS behavior improves with an enriched environment and therefore this therapy might also be beneficial to humans [16]. Current therapeutic strategies, both pharmacological and non-pharmacological, are aimed at improving the symptoms but do not improve cognitive function. In recent years, new strategies have been developed for therapeutic interventions in FXS based on the theories of mGluR and GABA receptors. Several clinical trials using new designed drugs were initiated to correct the abnormal activity of the mGluR and GABA pathways in FXS [6]. Unfortunately, some of them have been recently discontinued due to the lack of demonstrable improvement.

mGluR5 Inhibitors

The *fenobam* is a potent and selective antagonist of mGluR5 and was the first negative mGluR5 modulator tested in patients with FXS. To test whether it had significant effect on the FXS phenotype, *fenobam* was administered as a single oral dose to twelve affected patients (six males and six females) that the pre-pulse flicker noise (PPI -Pre-Pulse inhibition was measured inhibition of Acoustic Startle-) before and after drug administration. Generally, patients with FXS show decreased PPI compared to healthy control individuals. Six of the twelve patients with FXS showed improvement after treatment with ~~PPI~~ *fenobam*, with no significant adverse effects observed [17].

Although results were promising, it was difficult to draw definitive conclusions because of the lack of a controlled study with placebo, the fact that the patients received only a single dose

of *fenobam*, and the low number of participants enrolled. Other inhibitors are the AFQ056 mGluR5, from Novartis; the STX107, developed by Merck and whose license was issued to Seaside Therapeutics, and the RO4917523, from Hoffman-La Roche [6]. Despite of the results of a preliminary clinical trial with AFQ056 (*Mavoglurant*) in which certain cognitive aspects improved in FXS patients with complete methylation of the FMR1 gene [18], the trial was discontinued by Novartis last April after results of phase IIb/III failed to show demonstrable improvement. Last month, Roche discontinued its trial as well because of negative phase II clinical results.

GABA-A Agonists

GABA-A receptor agonists are currently used as anticonvulsants, antidepressants or anxiolytics. *Benzodiazepines*, which enhance the function of the GABA receptor, are the best-known drugs of this group. Although their anxiolytic effects are useful in patients with FXS have unwanted side effects such as sedation or ataxia, and treatment discontinuation may cause withdrawal symptoms [6]. Besides the selective agonists of GABA-A receptor, neuroactive steroids that allosterically modulate them may be also effective, for example *ganaxolone*, which has a favorable safety profile and may be useful in patients with FXS [19].

The use of inhibitors of GABA-A receptors in patients with FXS is likely to be effective in reducing symptoms such as seizures or sleep disorders.

GABA-B Agonists

Arbaclofen (STX209), a drug proven effective and safe for gastroesophageal reflux disease (GERD), has been used to treat autism spectrum disorders (ASD) in several studies. In an animal model of FXS, *arbaclofen* demonstrated a reversal of behavioral, neurological, and neuropathological features associated with the disease. Results from one double-blind, placebo-controlled study that treated children and adults with FXS have been promising, with signs of improvement in social function, especially in the most severely socially impaired individuals. Two studies, one open-label and one double-blind, placebo-controlled, were also conducted in children, adolescents, and young adults with ASD showing improvements in socialization [20, 21].

Approximately 13-18% of patients with FXS have seizures. The GABA-B and mGluR5 receptors are thought to be involved in the occurrence of audiogenic seizures in *Fmr1*-KO mice. Treatment of these mice with racemic baclofen reduced the incidence of seizures.

Another positive effect of GABA-B receptor agonists could be the reduction of anxiety in patients with FXS. Treatment of *Fmr1*-KO mice with a GABA-B receptors agonist can inhibit audiogenic seizures, supporting the idea that GABA-B receptors are involved in the etiology of FXS [6].

AMPA Agonists

The increased internalization of AMPA receptors in neurons of *Fmr1*-KO mice plays a major role in altering the transmission of signals. CX516 is an ampakine which acts as a positive allosteric AMPA receptor modulator. This compound binds to the AMPA receptor-channel complex and induces a slower deactivation of the receptor, which results in a longer opening time, in a slower disappearance of the excitatory postsynaptic potential, and ultimately to an increased hippocampal LTP, compared to baseline state before administering the drug. Thus, CX516 enhances AMPA receptors after glutamate-mediated synaptic. CX516 has been tested in phase II, randomized, double-blind, placebo-controlled safety for four weeks in adult patients with FXS. Unfortunately no significant or cognitive measures and improvement in behavior were observed. This could be due to the use of low doses, to the CX516 short half-life in humans or, finally, to insufficient time of treatment. In the study, there were only minimal side effects and no serious adverse effects were observed [22].

Theoretically, treatments targeting AMPA receptors could improve behavior in FXS, although the beneficial effects of ampakines have yet to be convincingly demonstrated.

NMDA Receptor Antagonists

Memantine is a non-competitive antagonist of NMDA receptors that can slow the progression of Alzheimer's disease and has been tested to treat Pervasive Developmental Disorders (PDD). In the presence of low levels of synaptic glutamate, binding of *memantine* blocks NMDA receptors, which are unblocked when glutamate levels increase. Besides excessive signaling through mGluR5 is linked to an increased AMPA receptor internalization and to the dysregulation of NMDA receptor activity [23]. Therefore, *memantine* may have positive effects on behavioral problems of patients with FXS.

A clinical trial of *memantine* in six patients with FXS who had a comorbid diagnosis of PDD has been reported. Therapeutic effects were determined by the clinical evaluation of the "Clinical Global Impressions-Improvement" (CGI-I) scale during the time of treatment. The results did not show any significant improvement, although in four of the six patients certain signs of improvement were observed. Furthermore, the study was not a randomized placebo-controlled trial and, therefore, it was difficult to draw valid conclusions [24].

The NMDA receptor may be a target for pharmacological treatment of SXF because some brain regions of the *Fmr1*-KO mice showed impaired NMDA-dependent LTP. However, there is no convincing evidence that NMDA receptor antagonists have demonstrable beneficial effects on behavior [6].

Additional Treatments

Lithium has been used for many years as a mood stabilizer. Most studies that have linked *lithium* with FXS have focused on the path of GSK3. *Lithium* inhibits the activity of GSK-3 β , which in turn inhibits the phosphorylation of microtubule-associated protein 1B (MAP1B). The MAP1B is one of the major targets of the mRNA to which it binds and translationally regulates FMRP. In 2008, a clinical trial with open-label *lithium* was published in FXS patients; although

the test was not randomized and placebo controlled, *lithium* appears that produced beneficial effects as decreased responses of aggression, abnormal vocalization improvement of self-harm and anxiety [25]. In conclusion, *lithium* appears to have beneficial effects on behavior and some cognitive functions but further research, such as long-term placebo-controlled studies are needed to check the effects of *lithium* in patients with FXS.

Minocycline is a tetracycline analog which can inhibit matrix metalloproteinase-9 (MMP-9) and reduce inflammation in the CNS. MMP-9 is an extracellular endopeptidase which breaks down the extracellular matrix proteins acting on synaptogenesis and morphology of dendritic spines. *Minocycline* is thought to act on the mGluR pathway and it has been tested in clinical trials to treat various neurological disorders (stroke, multiple sclerosis and autism). It has also been shown to have beneficial effects on the maturation of dendritic spines in cultured hippocampal neurons of Fmr1-KO mice [26]. A clinical trial of open-label studied the effects of *minocycline* in 50 patients with FXS was recently completed. The results were an improvement in language and behavior [27].

Acamprosate (calcium acetyl-homotaurine) is a commercially available drug used for the maintenance of abstinence from alcohol. This compound appears to have several mechanism of action: antagonist of mGluR5, weak antagonist of NMDA receptors and agonist of GABA-A receptors. Although the mechanism of action of *acamprosate* is not fully understood, a small clinical trial was conducted in three patients with FXS to evaluate the response to treatment with *acamprosate* using the CGI-I scale. After a minimum of 16 weeks of treatment, all three patients improved. Surprisingly, they also showed improvement in language skills. Two subjects experienced nausea. Although these results appear promising, this trial was not placebo-controlled and the number of participants was too small to draw definitive conclusions [28].

Aripiprazole, an atypical antipsychotic that was approved in the USA for the treatment of children and adolescents with autism, has also been tested in patients with FXS. The effect of *aripiprazole* in patients with FXS was assessed by the CGI-I and ABC-I scales. All subjects who completed the study showed significant improvement in irritable behavior [29]. However, the study was not placebo-controlled and therefore the findings were difficult to interpret.

Oxidative stress has been implicated in some psychiatric disorders, including autism. Studies in animal models suggested that in FXS there is an increased sensitivity to oxidative stress, which may alter neuronal and glial function. Indirect evidence was obtained by preclinical treatments of FXS performed using antioxidants. Indeed, chronic alpha-tocopherol treatment normalized free radical production and oxidative stress in the brain of FXS mice, in which it improved many of the behavioral and learning deficits of these mice, such as exploratory behaviors, habituation abnormalities, anxiety responses and contextual fear conditioning [30]. Recently, a phase II randomized, placebo-controlled trial was protocol was designed for children and adolescents with FXS [31].

Chronic melatonin treatment of FXS mice normalized the glutathione levels and prevented lipid peroxidation in the brain and testes of the mice. Moreover, melatonin improved some abnormal behaviors observed in FXS mice such as context-dependent exploratory and anxiety behaviors and learning abnormalities [32]. In a 4-week double-blind, placebo-controlled study with melatonin in 12 children with FXS, ASD or both, results showed a favorable effect (longer) in night sleep duration [33].

Final Remarks

FXS patients are treated with medications to alleviate the symptoms of the disease and/or the behavior problems such as hyperactivity or anxiety. Attempts to discover new drugs for the treatment of patients with FXS are becoming more promising with time due to improved understanding of the molecular mechanisms involved in the pathogenesis of the syndrome. Preliminary trials of new drugs in KO-mice models will be essential before they can be used in clinical trials in humans, although this does not guarantee their success.

To date, several clinical trials have been carried out with promising preliminary results. Nevertheless, others had to be discontinued due to lack of positive results. Hopefully, more trials are currently being conducted or designed with improvements such as randomized double-blind methodology, sufficient number of patients and the use of outcome objective measurements to determine the therapeutic efficacy of the drug. For this, most clinical trials used one or more classic psychological questionnaires, however, the measurements of improvement were based on the likely subjective assessments made by relatives, caregivers and teachers. To ensure objectivity, it is important to develop reliable and objective outcome measurements (i.e.: questionnaires, scales or tests) to determine the true efficacy of the new drug [34]. Fortunately, and despite of some recent disappointing results, a number of clinical-therapeutic trials for FXS and/or ASD are currently in course or being designed.

It is remarkable that in less than 25 years since the discovery of the genetic defect that causes FXS, targeted therapeutic strategies have been already developed, with more or less success, to treat some of the most disabling symptoms of affected individuals. Although curative treatment seems to be far yet, it is likely that in the near future more effective symptomatic treatments will be available for FXS, especially for individuals with ASD. If expectations are met, they will significantly improve their quality of life and of their families.

FRAGILE X-ASSOCIATED TREMOR-ATAXIA SYNDROME (FXTAS)

Fragile X-Associated Tremor-Ataxia Syndrome (FXTAS) is a progressive neurological disease that affects mainly males over 50 who carry a premutation allele (55-200 CGG repeats) in the fragile X gene FMR1. Affected individuals usually have intention tremor, ataxia, signs of parkinsonism, cognitive progressive decline and peripheral neuropathy. Besides, ancillary findings include autonomic dysfunction and psychiatric symptoms such as anxiety, depression and disinhibition [35].

As in FXS, there is no curative treatment for FXTAS, although there are a number of symptomatic therapies that can improve the clinical manifestations in affected individuals. Moreover, there is enough evidence regarding the efficacy of various medications for treatment of other diseases (i.e., Alzheimer diseases) that have important clinical overlap with FXTAS.

The most common therapeutic interventions for FXTAS are listed in Table 1, which includes symptom-based treatments [36]. The drugs are used, basically, to alleviate symptoms related to tremor, equilibrium coordination, parkinsonism, sleep problems, anxiety, mood alterations, memory problems and pain.

FRAGILE X-ASSOCIATED PRIMARY OVARY INSUFFICIENCY (FXPOI)

Primary ovarian insufficiency or premature ovarian failure (POI/POF) is one of the causes of female infertility. POI consists in cessation of menstrual periods, increased levels of FSH and diminished levels of estrogens, all before the age of 40. POI occurs in about 1% of women between 30 to 40 years of age. Fertility of women with POI is severely diminished, but unlike menopause, POI may be accompanied with spontaneous ovarian activity and natural pregnancies [37].

The major causes of POI include autoimmunity, genetic and environmental factors. Among the most common genetic conditions that produce POI is Fragile X premutation, which is present in all female carriers.

Treatment of POI includes hormone replacement therapy to reduce complications due to impaired endocrine function of ovaries, and fertility preservation therapies that include ovarian cortex, oocyte and embryo cryopreservation, oocyte or embryo donation and adoption in women without any ovarian function [38].

Recent research findings in animals and humans showed that neonatal and adult ovaries have some oogonial stem cells (OSCs) that can stably proliferate for months and produce mature oocytes *in vitro*. Studies on the isolation of OSCs from ovaries of aged animals and production of mature normal oocytes in ovaries of young adult animals lead to the recognition of the importance of OSC niche and intraovarian environment on their differentiation to mature, normal oocytes. Therefore, cases of POI that result from defects in ovarian niche and its incapacity to support differentiation and growth of oocytes and also ovarian aging may be reversible in future [39].

These results have offered the opportunity for the application of OSCs as a target for POI therapy, restoration of ovarian function and, subsequently, restoration of normal fertility. However, clinical utility of these cells for treatment requires more evidence to confirm their safety, especially the effects from epigenetic changes during *in vitro* culture, and manipulation of produced oocytes and also resultant offspring.

Table1. Symptom-based treatments for patients with FXTAS

Symptoms	Treatment/Therapy
Tremor	Primidone, beta-blockers, benzodiazepines
Ataxia	Amantadine and physical therapy
Parkinsonism	Carbidopa/levodopa, pramipexole and eldepryl
Cognitive deficits and dementia	Donepezil, rivastigmine, galantamine, memantine
Psychiatric problems	Sertraline, citalopram, escitalopram, duloxetine, mirtazapine, venlafaxine and aripiprazole
Autonomic dysfunction	Bladder incontinency: Tricyclic antidepressants, muscarinic receptor antagonists, cystoscopy with injection of Botox Swallowing difficulties: pyridostigmine bromide
Pain	Antidepressants, antiepileptics, topical analgesics, gabapentin and/or pregabalin

* Modified from Capelli et al., (2010). [36]

CONCLUSION

There is no curative treatment for Fragile X Syndrome. Symptomatic paliative therapies include behavioral and cognitive interventions, speech therapy, occupational and physical therapy, as well as pharmacological treatments. All of them intended to improve intellectual abilities, familial interactions and social integration. Pharmacological treatments should be individualized depending on the age and the intellectual-cognitive level of the affected individual. Treatment for Fragile X-Associated Tremor and Ataxia is also symptomatic, and include physical therapy, psychological and/or psychiatric interventions, and pharmacological treatments. Fragile X Carrier females with Premature Ovarian Insufficiency should be treated by the appropriate specialists. Hormone replacement and fertility preservation therapies are the two main issues that should be addressed.

REFERENCES

- [1] Oostra, B. A. and Willemsen, R. FMR1: A gene with three faces. *Biochim Biophys Acta* 2009, 1790, 467-477.
- [2] Garber, K. B., Visootsak, J., Warren, S. T. Fragile X Syndrome. *Eur J Hum Genet* 2008, 16, 666-672.
- [3] Brown W. T. Clinical aspects of the Fragile X Syndrome. In: Denman R. B. Ed. *Modeling Fragile X Syndrome: Results and Problems in Cell Differentiation* n° 54. Berlin: Springer-Verlag; 2012; pp. 273-279.
- [4] Kim, M., Bellini, M., Ceman, S. Fragile X mental retardation protein FMRP binds mRNAs in the nucleus. *Mol Cell Biol* 2009, 29, 214-228.
- [5] Irwin, S. A., Patel, B., Idupulapati, M., Harris, J. B., Crisostomo, R. A., Larsen, B. P., Kooy, F., Willems, P. J., Cras, P., Kozlowski, P. B., Swain, R. A., Weiler, I. J., Greenough, W. T. Abnormal dendritic spine characteristics in the temporal and visual cortices of patients with fragile-X syndrome: A quantitative examination. *Am. J Med Genet* 2001, 98, 161-167.
- [6] Levena, J., de Vrij, F. M. S., Oostra, B. A., Willemsen, R. Potential therapeutic interventions for Fragile X síndrome. *Trends Mol Med* 2010, 16, 516-527.
- [7] Gross, C., Berry-Kravis, E. M., Bassell, G. J. Therapeutic Strategies in Fragile X Syndrome: Dysregulated mGluR Signaling and Beyond. *Neuropsychopharmacology* 2011, 37, 178-195.
- [8] Ferraguti, F., Crepaldi, L., Nicoletti, F. Metabotropic glutamate 1 receptor: Current concepts and perspectives. *Pharmacol Rev* 2008, 60, 536-581.
- [9] Collingridge, G. L., Peineau, S., Howland, J. G., Wang, Y. T. Long-term depression n the CNS. *Nat Rev Neurosci* 2010, 11, 459-473.
- [10] Bear, M. F., Huber, K. M., Warren S. T. The mGluR theory of fragile X mental retardation. *Trends Neurosci* 2004, 27, 370-377.
- [11] Ceman, S., O'Donnell, W. T., Reed, M., Patton, S., Pohl, J., Warren, S. T. Phosphorylation influences the translation state of FMRP-associated polyribosomes. *Hum Mol Genet* 2003, 12, 32295-32305.

- [12] Nakamoto, M., Nalavadi, V., Epstein, M. P., Narayanan, U., Bassell, G. J., Warren, S. T. Fragile X mental retardation protein deficiency leads to excessive mGluR5-dependent internalization of AMPA receptors. *Proc Natl Acad Sci USA* 2007, 104, 15537-15542.
- [13] D'Hulst, C. and Kooy, R. F. The GABA(A) receptor: A novel target for treatment of fragile X? *Trends Neurosci* 2007, 30, 425-431.
- [14] Selby, L., Zhank, C., Sun, Q. Q. Major defects in neocortical GABAergic inhibitory circuits in mice lacking the fragile X mental retardation protein. *Neurosci Lett* 2007, 412, 227-232.
- [15] Tu, H., Rondard, P., Xu, C., Bertaso, F., Cao, F., Zhang, X., Pin, J. P., Liu, J. Dominant role of GABAB2 and Gbetagamma for GABAB receptor-mediated-ERK1/2/CREB pathway in cerebellar neurons. *Cell Signal* 2007, 19, 1996-2002.
- [16] Meredith, R. M., Holmgren, C. D., Weidum, M., Burnashev, N., Mansvelder, H. D. Increased threshold for spike-timing-dependent plasticity is caused by unreliable calcium signaling in mice lacking fragile X gene *FMR1*. *Neuron* 2007, 54, 627-638.
- [17] Hessel, D., Berry-Kravis, E., Cordeiro, L., Yuhas, J., Ornitz, E. M., Campbell, A., Chruscinski, E., Hervey, C., Long, J. M., Hagerman, R. J. Prepulse inhibition in fragile X syndrome: feasibility, reliability, and implications for treatment. *Am J Med Genet B Neuropsychiatr Genet* 2009, 150B, 545-553.
- [18] Jacquemont, S., Curie, A., des Portes, V., Torrioli, M. G., Berry-Kravis, E., Hagerman, R. J., Ramos, F. J., Cornish, K., He, Y., Paulding, C., Neri, G., Chen, F., Hadjikhani, N., Martinet, D., Meyer, J., Beckmann, J. S., Delange, K., Brun, A., Bussy, G., Gasparini, F., Hilse, T., Floesser, A., Branson, J., Bilbe, G., Johns, D., Gomez-Mancilla, B. Epigenetic modification of the *FMR1* gene in Fragile X Syndrome is associated with differential response to the mGluR5 antagonist AFQ056. *Sci Transl Med* 2011; 3:64ra1.
- [19] Cornish, K., Turk, J., Hagerman, R. The fragile X continuum: New advances and perspectives. *J Intellect Disabil Res* 2008, 52, 469-482.
- [20] Cryan, J. F. and Kaupmann, K. Don't worry 'B' happy! A role for GABA(B) receptors in anxiety and depression. *Trends Pharmacol Sci* 2005, 26, 36-43.
- [21] Erickson, C. A., Veenstra-Vanderweele, J. M., Melmed, R. D., McCracken, J. T., Ginsberg, L. D., Sikich, L., Scahill, L., Cherubini, M., Zarevics, P., Walton-Bowen, K., Carpenter, R. L., Bear, M. F., Wang, P. P., King, B. H. STX209 (*arbaclofen*) for autism spectrum disorders: An 8-week open-label study. *J Autism Dev Disord* 2014, 44, 958-964.
- [22] Berry-Kravis, E., Krause, S. E., Block, S. S., Guter, S., Wu, J., Leurgans, S., Decle, P., Potanos, K., Cook, E., Salt, J., Maino, D., Weinberg, D., Lara, R., Jardini, T., Cogswell, J., Johnson, S. A.; Hagerman, R. Effect of CX516, an AMPA-modulating compound, on cognition and behavior in fragile X syndrome: a controlled trial. *J Child Adolesc Psychopharmacol* 2006, 16, 525-540.
- [23] Shang, Y., Wang, H., Mercaldo, V., Li, X., Chen, T., Zhuo, M. Fragile X mental retardation protein is required for chemically-induced long-term potentiation of the hippocampus in adult mice. *J Neurochem* 2009, 111, 635-646.
- [24] Erickson, C. A., Mullett, J. E., McDougale, C. J. Open-label *memantine* in fragile X syndrome. *J Autism Dev Disord* 2009, 39, 1629-1635.
- [25] Berry-Kravis, E., Sumis, A., Hervey, C., Nelson, M., Porges, S. W., Weng, N., Weiler, I. J., Greenough, W. T. Open-label treatment trial of *lithium* to target the underlying defect in fragile X syndrome. *J Dev Behav Pediatr* 2008, 29, 293-302.

- [26] Bilousova, T. V., Dansie, L., Ngo, M., Aye, J., Charles, J. R., Ethell, D. W., Ethell, I. M. *Minocycline* promotes dendritic spine maturation and improves behavioural performance in the fragile X mouse model. *J Med Genet* 2009, 46, 94-102.
- [27] Utari, A., Chonchaiya, W., Rivera, S. M., Schneider, A., Hagerman, R. J., Faradz, S. M., Ethell, I. M., Nguyen, D. V. Side effects of *minocycline* treatment in patients with fragile X syndrome and exploration of outcome measures. *Am. J Intellect Dev Disabil* 2010, 115, 433-443.
- [28] Erickson, C. A., Mullett, J. E., McDougle, D. J. Brief report: *Acamprosate* in fragile X syndrome. *J Autism Dev Disord* 2010, 40, 1412-1416.
- [29] Erickson, C. A., Stigler, K. A., Wink, L. K., Mullett, J. E., Kohn, A., Posey, D. J., McDougle, C. J. A prospective open-label study of *aripiprazole* in fragile X syndrome. *Psychopharmacology* (Berl). 2011, 216, 85-90.
- [30] De Diego-Otero, Y., Romero-Zerbo, Y., El Bekay, R., Decara, J., Sánchez, L., Rodríguez-de-Fonseca, F., del Arco-Herrera, I. Alpha-tocopherol protect against oxidative stress in the fragile X knockout mouse: An experimental therapeutic approach for the *Fmr1* deficiency. *Neuropsychopharmacology* 2009, 34, 1011-1026.
- [31] De Diego-Otero, Y., Calvo-Medina, R., Quintero-Navarro, C., Sánchez-Salido, L., García-Guirado, F., Del Arco-Herrera, I., Fernández-Carvajal, I., Ferrando-Lucas, T., Caballero-Andaluz, R., Pérez-Costillas, L. A combination of ascorbic acid and α -tocopherol to test the effectiveness and safety in the fragile X syndrome: study protocol for a phase II, randomized, placebo-controlled trial. *Trials* 2014, 15, 345-360.
- [32] Romero-Zerbo, Y., Decara, J., el Bekay, R., Sanchez-Salido, L., Del Arco-Herrera, I., de Fonseca, F. R., de Diego-Otero, Y. Protective effects of melatonin against oxidative stress in *Fmr1* knockout mice: a therapeutic research model for the fragile X syndrome. *J Pineal Res* 2009, 46, 224-234.
- [33] Wirojanan, J., Jacquemont, S., Diaz, R., Bacalman, S., Anders, T. F., Hagerman, R. J., Goodlin-Jones, B. L. The efficacy of melatonin for sleep problems in children with autism, fragile X syndrome, or autism and fragile X syndrome. *J Clin Sleep Med* 2009, 5, 145-150.
- [34] Berry-Kravis, E., Hessel, D., Abbeduto, L., Reiss, A. L., Beckel-Mitchener, A., Urv, T. K. and Outcome Measurements Working Groups. Outcome measures for clinical trials in Fragile X Syndrome. *J Dev Behav Pediatr* 2013, 34, 508-522.
- [35] Hagerman, R. J., Hall, D. A., Coffey, S., Leehey, M., Burgeois, J., Gould, J., Zhank, L., Seritan, A., Berry-Kravis, E., Olichney, J., Miller, J. W., Fong, A. L., Carpenter, R., Bodine, C., Gane, L. W., Rainin, E., Hagerman, H., Hagerman, P. J. Treatment of fragile X-associated tremor ataxia syndrome (FXTAS) and related neurological problems. *Clin Intervention Aging* 2008, 3, 251-262.
- [36] Capelli, L. P., Rodrigues-Gonçalves, M. R., Leite, C. C., Barbosa, E. R., Nitirini, R., Vianna-Morgante, A. M. The fragile X-associated tremor and ataxia syndrome (FXTAS). *Arq Neuropsiquiatr*, 2010, 68, 791-798.
- [37] Sadeghi, M. R. New hopes for the treatment of Primary Ovarian Insufficiency/ Premature Ovarian Failure. *J Reprod Infertil* 2013, 14, 1-2.
- [38] Blumenfeld, Z. Fertility treatment in women with premature ovarian failure. *Expert Rev Obstet Gynecol* 2011, 6, 321-330.

- [39] White, Y. A., Woods, D. C., Takai, Y., Ishihara, O., Seki, H., Tilly, J. L. Oocyte formation by mitotically active germ cells purified from ovaries of reproductive-age women. *Nat Med* 2010, 18, 413-421.

Chapter 94

PRIVACY AND PROGRESS IN WHOLE GENOME SEQUENCING*

Presidential Commission for the Study of Bioethical Issues

EXECUTIVE SUMMARY

Over the course of less than a decade, whole genome sequencing has progressed from being one of our nation's boldest scientific aspirations to becoming a readily available technique for determining the complete sequence of an individual's deoxyribonucleic acid (DNA)—that person's unique genetic blueprint. With this tremendous advance comes the accumulation of vast quantities of whole genome sequence data and complex questions of how—across a multitude of clinical, research, and social environments—to protect the privacy of those whose genomes have been sequenced. Collections of whole genome sequence data have already been key to important medical breakthroughs, and they hold enormous promise to advance clinical care and general health moving forward. To realize this promise of great public good ethically, individual interests in privacy must be respected and secured.

Large-scale collections of genomic data raise serious concerns for the individuals participating. One of the greatest of these concerns centers around privacy: whether and how personal, sensitive, or intimate knowledge and use of that knowledge about an individual can be limited or restricted (by means that include guarantees of confidentiality, anonymity, or secure data protection). Because whole genome sequence data provide important insights into the medical and related life prospects of individuals as well as their relatives—who most likely did not consent to the sequencing procedure—these privacy concerns extend beyond those of the individual participating in whole genome sequencing. These concerns are compounded by the fact that whole genome sequence data gathered now may well reveal important information, entirely unanticipated and unplanned for, only after years of scientific progress.

Another privacy concern associated with whole genome sequencing is the potential for unauthorized access to and misuse of information. For example, in many states someone could

* This is an edited, reformatted and augmented version of the Presidential Commission for the Study of Bioethical Issues, dated October 2012.

legally pick up a discarded coffee cup and send a saliva sample to a commercial sequencing entity in an attempt to discover an individual's predisposition to neurodegenerative disease. The information might then be misused, for example, by a contentious spouse as evidence of unfitness to parent in a custody case. Or, the information might be publicized by a malicious stranger or acquaintance without the individual's knowledge or consent in a social networking space, which could adversely affect that individual's chance of finding a spouse, achieving standing in a community, or pursuing a desired career path.

Realizing the promise of whole genome sequencing requires widespread public participation and individual willingness to share genomic data and relevant medical information. This, in turn, requires public trust that any whole genome sequence data shared by individuals with clinicians and researchers will be adequately protected. Current U.S. governance and oversight of genetic and genomic data, however, do not fully protect individuals from the risks associated with sharing their whole genome sequence data and information. In particular, a great degree of variation exists in what protections states afford to their citizens regarding the collection and use of genetic data. Only about half of the states, for example, offer protections against surreptitious commercial genetic testing.

Currently, the majority of the benefits anticipated from whole genome sequencing research will accrue to society, while associated risks fall to the individuals sharing their data. This report focuses on reconciling the enormous public benefits anticipated from whole genome sequencing research with the potential risks to privacy of individuals, and the protections that must be foremost in our minds as we focus our policies to facilitate such privacy and progress.

Basic Ethical Principles for Assessing Whole Genome Sequencing

Laws and regulations cannot do all of the work necessary to provide sufficient privacy protections for whole genome sequence data. The Commission has been mindful of how the five ethical principles set out in its first report, *New Directions: The Ethics of Synthetic Biology and Emerging Technologies*, apply to the ethics of whole genome sequencing. These principles—which flow from the ideal of respect for persons—are public beneficence, responsible stewardship, intellectual freedom and responsibility, democratic deliberation, and justice and fairness. This report, *Privacy and Progress in Whole Genome Sequencing*, enlists these principles along with those set forth in the *Belmont Report* (a landmark statement of ethics for research involving human participants). *Privacy and Progress* focuses on recommendations aimed at pursuing and securing the public benefits anticipated from whole genome sequencing while minimizing the potential privacy risks to individuals.

These principles suggest ethically important and practically useful guidelines for whole genome sequencing. Chief among these is the principle of respect for persons, which requires strong baseline protections for privacy and security of data, while public beneficence requires facilitating ample opportunities for data sharing and access to data by clinicians, researchers, and other authorized users. Respect for persons further requires that any collection and sharing of individual data be based on a robust process of informed consent. Responsible stewardship calls for oversight and management of whole genome sequence information by funders, managers, professional organizations, and others. The principle of intellectual freedom and responsibility provides further support for pursuing whole genome sequencing and seeking models for broad data sharing by promoting regulatory parsimony. Democratic deliberation

urges all parties to consider changes to policies and practices in light of the evolving science and its implications for enduring ethical values. Finally, justice and fairness requires that we seek to channel the benefits of whole genome sequencing to all who can potentially benefit, and to ensure that the risks are not disproportionately borne by any subset of the population, including vulnerable or marginalized groups.

Recommendations

Currently we are in a period of intense transition with respect to integrating whole genome sequencing into clinical care, as well as facilitating access to and use of whole genome sequence data for research purposes. Moreover, the challenges we face today are not precisely the same challenges we will face in one, five, or ten years, as genomic technologies continue to develop and mature. Due to the rapid development of technology, we need to craft policies that are flexible and agile enough to ensure that we do not constrain our ability to adapt to evolving technology and social norms related to privacy and access.

Recognizing that ethical obligations reach beyond what is legally enforceable, the Commission examines both the relevant ethical principles and the relevant legal requirements to offer guidance as to what (ethically) *ought* to be done and what (legally) *must* be done. This is the foundation on which the Commission builds its *Privacy and Progress* recommendations.

Strong Baseline Protections while Promoting Data Access and Sharing

Presently, many national and state policies are in place to guard personally identifiable health information and records of participation in research. These policies should apply to all handlers of the data, from those who collect the data, to researchers who use them, to third-party storage and analysis providers (e.g., hosts of cloud computing services). Privacy protections should guard against unauthorized access to, and illegitimate uses of, whole genome sequence data and information while allowing for authorized users of these data to advance individual and public health.

Recommendation 1.1.

Funders of whole genome sequencing research; managers of research, clinical, and commercial databases; and policy makers should maintain or establish clear policies defining acceptable access to and permissible uses of whole genome sequence data. These policies should promote opportunities for models of data sharing by individuals who want to share their whole genome sequence data with clinicians, researchers, or others.

Strong baseline privacy protections require a spectrum of policies starting with data handling through the protection of persons from future disadvantage and discrimination arising from misuse of their whole genome sequence data. It is critical, however, to ensure that privacy regulations allow individuals to share their own whole genome sequence data with clinicians, researchers, and others in ways that they choose.

Recommendation 1.2.

The Commission urges federal and state governments to ensure a consistent floor of privacy protections covering whole genome sequence data regardless of how they were obtained. These policies should protect individual privacy by prohibiting unauthorized whole genome sequencing without the consent of the individual from whom the sample came.

Treating like data alike is crucial to ensuring consistent protections for whole genome sequence information across the United States. Although states should enact genomic policies that are most relevant and important to their constituents, bringing such protections to a minimum standard that addresses privacy—while still allowing individuals to share their own data—would provide just and fair protections regardless of where one happens to reside.

Data Security and Access to Databases

Data privacy requires data security. Data security requires ethical responsibility and accountability from all those who handle whole genome sequence data. It must further be supported by policies and infrastructure to protect safe sharing of data.

Recommendation 2.1.

Funders of whole genome sequencing research; managers of research, clinical, and commercial databases; and policy makers should ensure the security of whole genome sequence data. All persons who work with whole genome sequence data, whether in clinical or research settings, public or private, must be: 1) guided by professional ethical standards related to the privacy and confidentiality of whole genome sequence data and not intentionally, recklessly, or negligently access or misuse these data; and 2) held accountable to state and federal laws and regulations that require specific remedial or penal measures in the case of lapses in whole genome sequence data security, such as breaches due to the loss of portable data storage devices or hacking.

Many observe that absolute privacy is not possible in this, or many other realms. The greater potential for harm is not by virtue of authorized others *knowing* about one's whole genome make-up, but rather through the *misuse* of data that have been legally accessed.

Recommendation 2.2.

Funders of whole genome sequencing research; managers of research, clinical, and commercial databases; and policy makers must outline to donors or suppliers of specimens acceptable access to and permissible use of identifiable whole genome sequence data. Accessible whole genome sequence data should be stripped of traditional identifiers whenever possible to inhibit recognition or re-identification. Only in exceptional circumstances should entities such as law enforcement or defense and security have access to biospecimens or whole genome sequence data for non health-related purposes without consent.

The consent process should communicate limits on access to and use of genomic data to those having their whole genome sequenced in clinical care, research, and consumer-initiated contexts. These policies should apply to the original recipient of the data as well as to all parties who work with the data, from those who collect the sample or data through third-party storage and analysis service providers. Those who work with whole genome sequence data should remain current on regulations regarding data privacy and security.

Recommendation 2.3.

Relevant federal agencies should continue to invest in initiatives to ensure that third-party entrustment of whole genome sequence data, particularly when these data are interpreted to generate health-related information, complies with relevant regulatory schemes such as the Health Insurance Portability and Accountability Act and other data privacy and security requirements. Best practices for keeping data secure should be shared across the industry to create a solid foundation of knowledge upon which to maximize public trust.

Whole genome sequence data not stripped of traditional identifiers are considered “protected health information” and are covered under the Health Insurance Portability and Accountability Act’s Privacy, Security, and Enforcement Rules and the federal Common Rule for protecting human research participants. The same regulations, policies, and ethical guidelines that protect such health information should also be in place to govern the sharing of whole genome sequence data with third-party storage and analysis service providers. Public and the private sector parties should share their lessons learned to promote efficiency and avoid duplicating efforts.

Consent

Not unique to whole genome sequencing, a well-developed, understandable, informed consent process is essential to ethical clinical care and research. To educate patients and participants thoroughly about the potential risks associated with whole genome sequencing, the consent process must include information about what whole genome sequencing is; how data will be analyzed, stored, and shared; the types of results the patient and participant can expect to receive, if relevant; and the likelihood that the implications of some of these results might currently be unknown, but could be discovered in the future. Respect for persons requires obtaining fully informed consent at the outset of diagnostic testing or research.

Recommendation 3.1.

Researchers and clinicians should evaluate and adopt robust and workable consent processes that allow research participants, patients, and others to understand who has access to their whole genome sequences and other data generated in the course of research, clinical, or commercial sequencing, and to know how these data might be used in the future. Consent processes should ascertain participant or patient preferences at the time the samples are obtained.

Recommendation 3.2.

The federal Office for Human Research Protections or a designated central organizing federal agency should establish clear and consistent guidelines for informed consent forms for research conducted by those under the purview of the Common Rule that involves whole genome sequencing. Informed consent forms should: 1) briefly describe whole genome sequencing and analysis; 2) state how the data will be used in the present study, and state, to the extent feasible, how the data might be used in the future; 3) explain the extent to which the individual will have control over future data use; 4) define benefits, potential risks, and state that there might be unknown future risks; and 5) state what data and information, if any, might be returned to the individual.

Each Common Rule agency has its own enforcement authorities to protect research participants. All agencies should work together as they develop clear and consistent guidelines for their informed consent forms. Clinical consent documents for whole genome sequencing will have to address a number of issues specific to whole genome sequencing: an explanation of the science, whether whole genome sequence data collected for clinical applications will be made available for research purposes, and what types of results will be produced through whole genome sequencing. For example, an important unsettled issue is the ethics of reporting incidental findings to individuals— that is, information gleaned from whole genome sequencing research or clinical practice that was not its intended or expected object.

Recommendation 3.3.

Researchers, clinicians, and commercial whole genome sequencing entities must make individuals aware that incidental findings are likely to be discovered in the course of whole genome sequencing. The consent process should convey whether these findings will be communicated, the scope of communicated findings, and to whom the findings will be communicated.

Recommendation 3.4.

Funders of whole genome sequencing research should support studies to evaluate proposed frameworks for offering return of incidental findings and other research results derived from whole genome sequencing. Funders should also investigate the related preferences and expectations of the individuals contributing samples and data to genomic research and undergoing whole genome sequencing in clinical care, research, or commercial contexts.

Individuals undergoing whole genome sequencing in research, clinical, and commercial contexts must be provided with sufficient information in informed consent documents to understand what incidental findings are, and to know if they will or will not be notified of incidental findings discovered as a result of whole genome sequencing.

Facilitating Progress in Whole Genome Sequencing

Currently, large amounts of patient data are being collected in the health care setting, stripped of traditional identifiers, analyzed, and fed into research that might one day improve clinical care. This “learning health system” model both translates advances in health services research into clinical applications and collects data during clinical care to facilitate further advances in research. Learning health system advocates and others support standardized electronic health record systems and infrastructure to facilitate health information exchange so that data can be easily aggregated and studied. Integrating whole genome sequence data into health records in the learning health system model can provide researchers with more data to perform genome-wide analyses, which in turn can advance clinical care.

Recommendation 4.1.

Funders of whole genome sequencing research, relevant clinical entities, and the commercial sector should facilitate explicit exchange of information between genomic researchers and clinicians, while maintaining robust data protection safeguards, so that whole genome sequence and health data can be shared to advance genomic medicine.

Current sequencing technologies and those in development are diverse and evolving, and standardization is a substantial challenge. Ongoing efforts are critical to achieving standards for ensuring the reliability of whole genome sequencing results, and facilitating the exchange and use of these data.

Recommendation 4.2.

Policy makers should promote opportunities for the public to benefit from whole genome sequencing research. Further, policy makers and the research community should promote opportunities for the exploration of alternative models of the relationship between researchers and research participants, including participatory models that promote collaborative relationships.

Respect for persons implies not only respecting individual privacy, but also respecting research participants as autonomous persons who might choose to share their own data. Public beneficence is advanced by giving researchers access to plentiful data from which they can work to advance health care. Regulatory parsimony recommends only as much oversight as is truly necessary and effective in ensuring an adequate degree of privacy, justice and fairness, and security and safety while pursuing the public benefits of whole genome sequencing. Therefore, existing privacy protections and those being contemplated should be parsimonious and not impose high barriers to data sharing. While the Commission supports the intellectual freedom this access will encourage, clinicians and researchers must also act responsibly to earn public trust for the research enterprise.

Public Benefit

Thousands of citizens have participated in whole genome sequencing research personally, and all citizens help to support government investment in whole genome sequencing through their general participation in and support of our political system. Therefore, all citizens should have the opportunity to benefit from medical advances that result from whole genome sequencing.

Special caution should be taken on the part of researchers to ensure that their participants accurately reflect as much as possible the rich diversity of our population. Different groups have genomic variants at different frequencies within their populations, and sufficiently diverse data must be collected so that advances arising from whole genome sequencing can be used for the benefit of all groups.

Recommendation 5.1.

The Commission encourages the federal government to facilitate access to the numerous scientific advances generated through its investments in whole genome sequencing to the broadest group of persons possible to ensure that all persons who could benefit from these developments have the opportunity to do so.

Government investment in genomic research has resulted in public benefit through improved health care and in economic return on investment. The principle of justice and fairness requires that the benefits and risks of whole genome sequencing be distributed equitably across society. Research funded with taxpayer contributions should benefit all members of society. To these ends, researchers should be vigilant about including individuals from all sectors of society in their studies, so that research findings can be translated widely into improved clinical care. The federal government should follow through on its investment

in research and assure that the discoveries of whole genome sequencing are integrated with clinical care to benefit the health of all.

INTRODUCTION

The Potential of Whole Genome Sequencing

In 1996, Retta Beery gave birth to apparently healthy twins Alexis and Noah.¹ It soon became clear, however, that something was wrong; the twins cried nonstop and had developmental problems. Over the next two years, Retta and her husband Joe endured the physical, emotional, and financial costs of visiting numerous specialists, putting their young twins through countless tests, and having their children undergo surgery. None of these steps provided results or solutions.

In 1998, the twins were diagnosed with cerebral palsy and a related course of treatment was outlined. Although the treatment yielded some symptomatic improvement, Retta felt that the diagnosis was incorrect. In 2002, the Beerys were starting to look at wheelchairs and feeding tubes when Retta, after four years of research, stumbled upon an article on DOPA-responsive dystonia (also known as Segawa's dystonia) and suspected that this was the disease that the twins had. The Beerys contacted a specialist, and after a physiological test the twins were diagnosed with Segawa's dystonia. They began a new course of treatment to increase brain dopamine, which yielded a dramatic improvement in their health.

In 2009, Alexis developed breathing problems and was forced again to endure multiple emergency room visits and a battery of tests and visits to specialists. In August 2010, the Beerys went to Baylor College of Medicine for diagnostic whole genome sequencing. By November, Alexis's and Noah's whole genomes had been sequenced. Their data were compared to other whole genome sequences in databases, such as the Baylor Human Genome Sequencing Center's database, to reveal what was unique about the Beery twins' genomes. Clinicians now had answers for the family. The geneticists had uncovered an extremely rare and only recently recognized genetic cause of DOPA-responsive dystonia producing a deficiency of not only dopamine but also serotonin production in the brain. Armed with this new information, the Beerys returned to their neurologist, who amended the treatment regimen for Alexis and Noah with an over-the-counter supplement. Within a month, Alexis's breathing problems disappeared.

As a result of that final piece of the puzzle—the information provided by whole genome sequencing—Alexis is able to breathe normally and can now even compete in sports. Both children have a definitive diagnosis, and are expected to live long, healthy lives.

THE CHALLENGE OF PRIVACY

Victoria Grove's sisters struggled with a difficult genetic diagnosis: alpha-1 antitrypsin deficiency. The genetic illness meant that her sisters' bodies did not make enough of a protein that protected their lungs and liver from damage, which could lead to emphysema and liver disease.

Victoria wanted to help them, and in 2004 agreed to enroll in a research study of families with alpha-1 antitrypsin deficiency. *"I just knew I didn't have it, so I signed up for the study."* But the tests came back positive—Victoria had the same genetic mutation as her sisters. She did not yet have any symptoms, and wanted to keep her test results private, so she did not tell her doctor.

In 2005, Victoria got tested again to confirm the research study results. She used a private company and had the results sent directly to her. Victoria's second test came back positive, and she chose again not to send the results to her doctor, fearing that the information would be included in her medical record. Victoria worried that this information could lead her insurance company to drop her coverage or charge her higher rates. Victoria kept her genetic results private for nearly three years.

The pivotal moment came when Victoria felt she was coming down with a bout of pneumonia but could not convince the nurse practitioner who saw her to order the X-ray necessary to prescribe antibiotics. Victoria went home without antibiotics, her condition worsened, and she called back a few days later. The nurse asked Victoria to come in again, but Victoria told them she could not drive across town in the snowstorm that had immobilized the city. She could, however, get to a pharmacy near her house if the office called in the antibiotics. The nurse on the phone insisted this was not possible. *"My emotions just took hold and I cried 'I have alpha-1 and I need that antibiotic,'"* Victoria said. *"At that point the cat was out of the bag."*

Today Victoria gets regular treatments for her condition. She recognizes that fear kept her from providing her clinicians with crucial information. Still, she can't convince either her brother or her son to get tested for alpha-1. Victoria says both men are aware that there is federal protection from discrimination in employment and health insurance, but fear that these laws will not provide sufficient protection. Her son already has to buy his own health insurance; he does not want any information in his medical record that could jeopardize his job or his access to health insurance. *"I can imagine in a job situation, it's expensive to take on someone if they're ill. And you can always get rid of people for other reasons. I assume that's going on."*

Whole genome sequencing offers great promise of medical advances that could benefit all of society, but this promise is tempered by the concerns of individual privacy. This tension between medical progress and the risks to privacy from whole genome sequencing is the subject of this report. To use whole genome sequencing to discover the changes in deoxyribonucleic acid (DNA) that underlie disease, scientists and clinicians must have access to whole genome sequence data from many individuals (for the definitions of scientific terms used in this report, see Appendix I: Glossary of Key Terms). Continued advances therefore depend on large numbers of individuals who are willing to share their whole genome sequence data for research purposes. Further, scientists are better able to make connections between variations in whole genome sequence data and specific diseases when additional health and demographic information accompanies these data. But this additional information might make it easier to identify an individual and discover his or her private health information.² Thus, while society stands to benefit from advances in improved medical treatment and diagnosis from whole genome sequencing, the privacy risks associated with sharing whole genome sequence data fall predominantly on the individuals themselves.

Whole genome sequencing is a technique that determines the complete sequence of DNA in an individual's cells (See Figure 1. For more information regarding the science of whole

genome sequencing, see Appendix II: Genetic and Genomic Background Information). Whole genome sequencing reveals the genetic blueprint for a person, generating information on every gene in the nucleus of one’s cells. Each person’s DNA is unique, and changes in DNA can lead to disease. The ability to link variations in DNA with health and disease outcomes, a process still in its infancy, holds promise for substantial public benefit.³ These benefits have the potential to alter the way we treat cancer, heart disease, diabetes, Alzheimer’s disease, schizophrenia, and countless other illnesses.

The Commission believes that the ethical principles and recommendations in this report should not be limited to whole genome sequencing. Whole genome sequencing is the focus of this report because of its current promise to advance health. The ideas in this report, however, apply broadly to all studies using large-scale genetic and genomic data and information, including whole exome sequencing, genome-wide SNP analysis, and other large-scale genomic studies. The tools used to decipher and study whole genomes are evolving rapidly, and it is not clear what additional technologies will emerge in the near future. What is clear is that all current genetic and genomic research can be measured against the ethical principles and recommendations described in this report, and the Commission is optimistic that its recommendations will specifically accommodate future advances in large-scale sequencing and analysis.

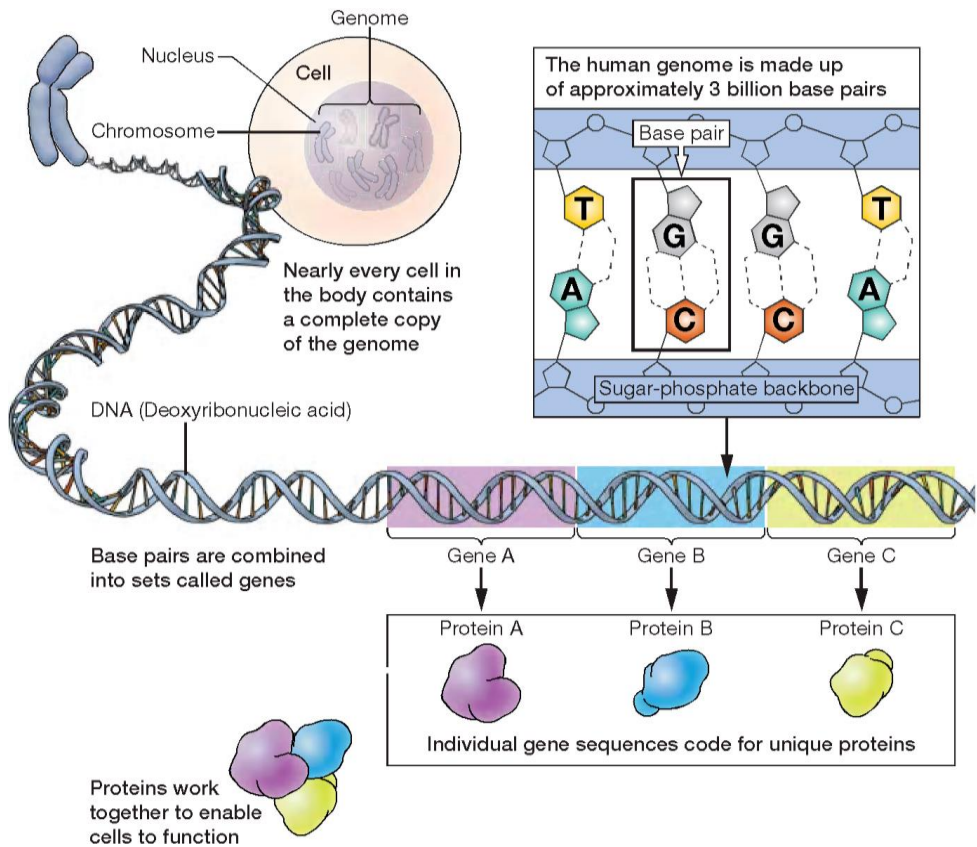


Figure 1. The Structure of DNA.

TERMINOLOGY

Whole genome sequencing: determining the order of nucleotide bases—As, Cs, Gs, and Ts—in an organism’s entire DNA sequence

Whole genome sequence data: the file of As, Cs, Gs, and Ts that results from whole genome sequencing

Whole genome sequence information: facts derived from whole genome sequence data, such as predisposition to disease

Genomics: the study of all the DNA (the genome) in an individual, and how parts of the genome interact with each other and the environment

Genetic test: a discrete test that examines a specific genetic location or a single gene, such as the test for Huntington’s disease

Genotyping: analyzing a handful to thousands of discrete variants across the genome (i.e., more than a discrete genetic test, but less than whole genome sequencing)

For additional terminology, please see Appendix I: Glossary of Key Terms and Appendix II: Genetic and Genomic Background Information.

Current clinical uses of DNA information are limited mostly to specific genetic tests. If clinicians suspect a particular disease with a known genetic cause, such as Huntington’s disease, they can order a genetic test looking at one specific gene among the more than 20,000 genes in the human genome. These tests examine only a few of the whole genome’s three billion pairs of building blocks. The price of sequencing a whole genome is dropping rapidly, however, and soon it will be less expensive to sequence an entire genome than to perform a few individual genetic tests. Once this happens, whole genomes might be sequenced in lieu of discrete genetic tests, and such information can be stored in a patient’s medical records. Then, if a clinician would like to find out something about a patient’s DNA in the future, he or she could examine the whole genome sequence data already stored in that patient’s record. For example, a patient’s response to a particular dose of warfarin, a drug that helps prevent blood clotting, is partly dependent on his or her genetic makeup. A clinician with access to a patient’s whole genome sequence can use it to identify drug sensitivity and reduce the time required to achieve the optimal dosage.⁴

The sheer amount of information contained in our genomes is what makes whole genome sequence data different from other medical information. Our whole genome sequence data can reveal predispositions to diabetes, cancer, or psychiatric conditions. The data can also reveal variations in DNA that are not yet understood. For example, an apparently healthy individual could be missing a small piece of DNA. The person seems healthy, but will that variant cause a problem in the future?

Over 20,000 individual human genes have been identified. A major recent advance by the National Institute of Health’s (NIH) Encyclopedia of DNA Elements (ENCODE) project greatly enhanced our knowledge of the function of the genome through a flurry of scientific publications, finding that 80 percent of the genome has a “biochemical function.”⁵ For years,

that number had stood at 10 percent.⁶ Eric Lander, president of the Broad Institute, compared the results of the Human Genome Project, which sequenced the first full human genome, to a picture of the Earth from space, and compared the ENCODE project to Google Maps. The ENCODE Project is a major step toward demonstrating the function of the whole genome sequence that was determined in the Human Genome Project, much like Google Maps can refine a snapshot of the Earth by showing traffic, alternate routes, and the location of landmarks.⁷ The function of 20 percent of the non-coding regions—regions of DNA that do not contain specific instructions for making proteins—is still unknown, but these regions might have functions that are yet to be determined.

Unlike genetic testing—which looks at the specific parts of the genome to reveal a variant at a specific location of a single gene indicating a particular disease—whole genome sequencing reveals an individual’s entire genome, including all variants within the genome. These variants are changes in the DNA sequence and range in size from small changes like a single base pair change, to larger changes such as a deletion of a portion of the DNA strand. As more information about our genomes becomes available, variants that might be revealed by whole genome sequencing include: specific known disease variants; variants of unknown significance (e.g., an unknown variant in the region that increases risk for heart disease); nonmedical genetic traits, including hair and eye color; carrier status variants, including variants that do not cause disease in the individual but could be passed on, such as mutations for hemophilia or cystic fibrosis; susceptibility genes, such as those that slightly increase susceptibility to diabetes, heart disease, or some cancers; and genes for conditions with late onset that will not affect an individual until much later in life, such as Alzheimer’s disease and Huntington’s disease.⁸ Only a small number of the genetic variants that whole genome sequencing might reveal have yet been studied enough to substantiate their connection to disease.⁹

“And so having the genome may be not incredibly powerful right now, but it opens the door to outrageous rates of discovery, which I’m pretty certain are going to happen over the next five to ten years.”

Leonard D’Avolio, Associate Center Director for Biomedical Informatics, Massachusetts Veterans Epidemiology Research and Information Center, Department of Veterans Affairs, Instructor, Harvard Medical School. (2012). Genomic Privacy, Data Access, and Health IT. Presentation to PCSBI, May 17, 2012. Retrieved from [http:// bioethics.gov/cms/node/713](http://bioethics.gov/cms/node/713).

Whole genome sequencing also raises many potential concerns for individuals. One might shoulder the burden of knowing medical information regarding future adverse health conditions for which there is currently no treatment. Whole genome sequencing raises concerns about our privacy as well. Just as patients would not want to give anyone access to their medical record, many people might not want others to have access to their whole genome sequence data and information. With unauthorized access comes concerns about misuse of information. For example, someone could pick up a discarded coffee cup and send a sample of saliva—which contains DNA—from the rim of the cup to a commercial sequencing entity in an attempt to discover an individual’s predisposition to neurodegenerative disease. The information might then be misused by publicizing it in a social networking space, which could derail that

individual's chance of finding a spouse, achieving standing in a community, or pursuing a certain career path.

To yield medically useful information, an individual's genomic sequence data needs to be coupled with clinical information about disease and compared to other genomic sequence data. Further, genomic research is complex because each person's DNA naturally has thousands of variants, and the vast majority of these variants do not cause disease. The research and clinical power of whole genome sequencing lies in being able to compare a large number of whole genome sequence data sets that are linked with relevant health and disease states. This type of study allows researchers to identify sequence variations and associations between whole genome sequence variations and disease. For this reason, scientists need whole genome sequence data to be linked to clinical, laboratory, and socio-demographic data. This linking can be done by entering only relevant information (e.g., disease state or symptoms) and excluding personally identifiable information, such as an individual's name or address; but without access to relevant medical data, links between whole genome sequence variations and disease could not be identified.

Recent technological advances have facilitated storing and sharing of whole genome sequence data. Whole genome sequence data and associated health information can now be stored in genomic databases and biorepositories that contain digital information and physical samples, respectively, from large numbers of persons. By using these resources, researchers will have the volume of data they need to advance medical understanding for the public good through genomics. However, this data storage and sharing raises its own questions: How does one securely store these huge data files? Who should have access to these data files? How can these data be used productively, and how might they be misused? What constitutes "misuse"? What should the penalties be for misusing these data? The summation of all these issues—the unknowns, privacy, consent, data security, and data storage involved in whole genome sequencing—will require careful and sustained ethical attention.

This report delves into two crucial questions: What information about an individual's whole genome should remain private, and when should it remain private? The Commission explores how, when, and why genomic information should remain subject to clear rules of confidentiality, secrecy, information security, decisional autonomy, and freedom from unwanted intrusion out of respect for individuals. Without trust in the confidentiality and security of the data, individuals could be less likely to participate in research. Conversely, with well-founded trust that their sense of privacy will be honored, individuals are treated with the respect to which they are entitled and might be more likely to contribute to the research enterprise that promises important public benefits.

This report therefore aims to pursue and secure the public benefit anticipated from whole genome sequencing while minimizing the potential privacy risks to individuals. The recommendations draw upon the principles that flow from the ideal of respect for persons, and are set forth in the *Belmont Report*, a landmark statement of ethics for research involving human participants, and those outlined in the Commission's first report *New Directions: The Ethics of Synthetic Biology and Emerging Technologies*.¹⁰

The Promise of Whole Genome Sequencing

Whole genome sequencing can help researchers and clinicians better understand the unique qualities of a disease, and, especially when combined with other information, might help select treatment methods.¹¹ Researchers already have been able to help clinicians aid some children born with rare birth defects by sequencing and analyzing their whole genomes to diagnose and treat their illnesses.¹² Researchers are also teaming up with clinicians in using whole genome sequence data to advance personalized medicine, including predicting an individual's risk for a heart attack or determining the best dosage of medication for an individual.¹³ Researchers recently determined a fetal whole genome sequence using a blood sample from the mother, an innovation that could soon reach the clinic.¹⁴ And this is only the beginning of the whole genome sequencing era, which has the potential to revolutionize medicine.

In 2000, the cost of sequencing a single human genome was estimated to be 2.5 billion dollars; it is anticipated that this cost will soon be \$1,000. As the cost falls, whole genome sequencing will be increasingly integrated into clinical care. Clinicians can—and many will—incorporate whole genome sequence information into the clinic to promote the practice of personalized medicine.¹⁵ Nevertheless, little has been written about the ethical concerns of integrating whole genome sequencing into the clinical context, which is particularly problematic given the speed with which this could occur.¹⁶ The Commission therefore presents its recommendations mindful of the changing uses and implications of whole genome sequencing. Although this report focuses on issues related to privacy and sharing of whole genome sequence data, the Commission recognizes that another important unsettled issue is the ethics of reporting incidental findings to individuals—that is, information gleaned from whole genome sequencing research or clinical practice that was not its intended or expected object. The Commission plans to take up the issue of incidental findings in the future.

Privacy Concerns

At age 13, Brian Hurley learned from an ophthalmologist that he had retinitis pigmentosa and that at some unknown point in his life he would go blind. During high school, Brian learned about careers in law and thought this was something he could do well, regardless of eyesight. When he started law school, however, he realized he did not like it. Brian needed to find something he could do and wanted to do—not just something a blind person was considered capable of doing.

Brian felt tremendous pressure to resolve his career path before he lost his vision: “In the beginning of a career, you try to figure out what you are good at and hopefully enjoy, but I was more concerned about could I do it well when blind.” Brian spent hours online searching for careers that might work, and successful blind professionals that he could use as role models. “It was like having a time bomb inside of me,” Brian said.

After college, Brian experienced a steady decline in his peripheral vision. At age 27, Brian stopped driving. During this time, Brian's actual symptoms did not match the decline of his emotional state. Brian said he was so panicked that it took the joy out of his last few years before becoming legally blind. “If you took my mental condition, I might as well have been blind already.”

Then, at 33, Brian lost the majority of his eyesight. “The irony is, anticipation was much worse than the actual loss. It was a relief to stop worrying when the loss would occur.”

Today, at 39, and relieved of the anticipation, he enjoys his current role as a Public Affairs Program Director. Brian refuses to let his vision loss be an obstacle to his professional and personal goals.

Brian recently learned of the eyeGENE® program at the National Institutes of Health. He wants to help with eye research—specifically research related to retinitis pigmentosa. He knows research is important and wants to contribute his data to help others. He does not, however, want his whole genome sequenced in the course of participating in research.

Before enrolling in the eyeGENE® program, Brian spent three days with a lighted magnifier and 20 pages of consent forms to ensure that researchers would not sequence his entire genome and that they will not divulge findings about other diseases to him. Having lived with one time bomb, Brian understands its collateral damage. He never wants to carry that burden again. In his situation, Brian feels that having less information is better.

With regard to whole genome sequence data, privacy concerns are more complex than a simple decision about whether to undergo whole genome sequencing and, if so, whether the data should be included with an individual’s medical record. Individuals might have good reason for wanting to share particular parts of their genomic data—such as for the purposes of research— but might also want to limit the extent to which others can access these data.

The prevention of unauthorized use or disclosure of medical information about specific individuals has long been a serious ethical concern. Whole genome sequencing dramatically raises the privacy stakes because it necessarily involves examining and sharing large amounts of biological and medical information that is not only inherently unique to a single person but also has implications for blood relatives. Genomic information is inherited and determines traits like hair and eye color. Unlike a decision to share our hair or eye color, which does not reveal anything about our relatives that is not observable, a decision to learn about our own genomic makeup might inadvertently tell us something about our relatives or tell them something about their own genomic makeup that they did not already know and perhaps do not want to know. More than other medical information, such as X-rays, our genomes reveal something both objectively more comprehensive and subjectively (to many minds) more fundamental about who we are, where we came from, and the health twists and turns that life might have in store for us.

The fact that whole genome sequence information is uniquely connected to our conceptions of self is what could cause the inappropriate disclosure or misuse of this information to be so harmful. In theory, whole genome sequence information could be used to deny financial backing or loan approval, educational opportunities, sports eligibility, military accession, or adoption eligibility.¹⁷ Disclosing genomic information could affect the opportunities available to individuals, subject them to social stigma, and cause psychological harm. The full extent of what whole genome sequencing can reveal is unknown, but we know that having one’s whole genome sequenced today could reveal genetic variants that increase the risk for certain conditions such as Alzheimer’s disease, which many people either do not want to know about themselves or others to know about them.

“[H]arm is not the act...of distributing data. Harm comes from actions that are taken once the data have been distributed.”

John Wilbanks, Founder, Consent to Research; Senior Fellow, Kauffman Foundation; Research Fellow, Lybba. (2012). Privacy II – Control, Access and Human Genome Sequence Data. Presentation to PCSBI, February 2, 2012. Retrieved from <http://bioethics.gov/cms/node/659>.

It is understandable, therefore, that whole genome sequencing heightens concerns about how unauthorized disclosure can threaten one’s individual privacy. But determining what privacy requires in the whole genome context is not straightforward. In the legal context, privacy is multidimensional and includes physical, informational, decisional, proprietary, associational, and intellectual aspects.¹⁸ While there is no consensus definition of privacy, in this report we consider privacy to be a general concept that includes confidentiality, secrecy, anonymity, data protection, data security, fair information practices, decisional autonomy, and freedom from unwanted intrusion.¹⁹ Whole genome sequencing calls for serious consideration of each of these components and their related ethical concerns. It also is important to recognize at the outset that, in some significant respects, parts of our genomic information are not and cannot be wholly private. When we routinely provide a blood sample in a clinical exam, decide to submit a DNA sample to be used in research, or unintentionally leave behind traces of DNA on a coffee cup that we discard in a public waste bin, we are providing some other individuals the opportunity to learn something about us.

While doing everything possible to prevent any use of whole genome sequence data certainly would provide strong privacy protection, it would fail to allow the anticipated public benefit that is to be achieved by sharing whole genome sequence data and advancing science. Because preventing all whole genome sequence data sharing would stifle potentially life-saving and life-enhancing medical progress, we must focus on how best to protect confidentiality of data, ensure security of information from unauthorized access and uses, preserve decisional autonomy as to possible uses, and guarantee the freedom of individuals from unwanted and unwarranted intrusion.

Policy and Governance

“If you sequence people’s exomes you’re going to find stuff,” said Gholson Lyon, a physician and researcher previously at the University of Utah, now at Cold Spring Harbor Laboratory.

As part of his research, Dr. Lyon worked with a family in Ogden, Utah. Over two generations, four boys had died from an unknown disease with a distinct combination of symptoms—an aged appearance, facial abnormalities, and developmental delay.

Dr. Lyon sought to identify the genetic cause of this disease, and collected blood samples from 12 family members who had signed consent forms. The family members understood these forms to mean that they would have access to their results.

Dr. Lyon conducted exon capture and sequencing of the X chromosome—a process that analyzes specific regions of the X chromosome and is a less expensive alternative to whole genome sequencing—to analyze the blood samples. Dr. Lyon and his colleagues identified a genetic mutation, and named the disease Ogden Syndrome after the family’s hometown.

After Dr. Lyon and his team identified the genetic basis of Ogden Syndrome, one of the family members contacted him. This young mother of one daughter had submitted a blood sample for Dr. Lyon's research. She had not been pregnant at the time, but was now four months pregnant with her second child. She knew that she was carrying a boy and wanted to know if she was a carrier of the mutation. She wanted to be able to mentally and emotionally prepare herself and her family.

By reexamining his research data, Dr. Lyon was able to see that the expectant mother was a carrier of Ogden Syndrome. This meant that her son had a 50 percent chance of being born with the disease. Dr. Lyon could not, however, legally share this important information with the family because he had conducted the original sequencing in a research laboratory that had not satisfied federally mandated standards designed to ensure the accuracy of clinical genetic results.

Instead, Dr. Lyon worked to have the mutation validated at a laboratory that satisfied those federal standards; this involved overcoming substantial bureaucratic hurdles and other obstacles that held up the process. During this time, the baby boy was born and died of Ogden Syndrome at four months of age. While knowing the results would not have changed the outcome, Dr. Lyon feels he should have been able to do more for the family.

Dr. Lyon has become an outspoken advocate for conducting whole genome sequencing in laboratories that satisfy the federal standards so that researchers can return results to participants, if appropriate. Dr. Lyon wants clear guidance for laboratories conducting genetic research and clear language in consent forms that clarifies the results that participants should expect to have returned from the researchers.

Realizing the promise of whole genome sequencing requires widespread public participation and individual willingness to share genomic data and relevant medical information. This requires public trust that any whole genome sequence data shared by individuals with researchers and clinicians will be adequately protected. Individuals must trust that their whole genome sequence data will not be either intentionally or inadvertently disclosed or misused. Current U.S. governance and oversight of genetic and genomic data, however, do not fully protect individuals from the risks associated with sharing their whole genome sequence data and information.

The Genetic Information Nondiscrimination Act of 2008 (GINA) is the leading federal protection of genetic information, but it offers only prohibition of genetic discrimination in health insurance and employment. GINA does not regulate access, security, and disclosure of genetic or whole genome sequence information across all potential users, nor does it protect against discrimination in other contexts. U.S. state laws on genetic information vary greatly in their protections of individuals, and they also fail to provide uniform privacy protections. In an era in which whole genome sequence data are increasingly stored and shared using biorepositories and databases, there is little to no systematic oversight of these systems.

Ethical Principles

Laws and regulations cannot do all of the work necessary to provide sufficient privacy protections for whole genome sequence data. Individuals who obtain their whole genome sequence data also have a responsibility to thoughtfully consider to what extent they ought to

act to protect their own privacy beyond current legal protection when considering whether to share their data and information publicly.

In its previous reports, the Commission established an ethical framework for considering the implications of scientific advances, including emerging technologies, that can be applied in similar situations. That framework outlines principles developed to apply particularly to emerging biotechnologies that do not directly involve human therapy or human experimentation. These guiding principles are 1) public beneficence, 2) responsible stewardship, 3) intellectual freedom and responsibility, 4) democratic deliberation, and 5) justice and fairness.

As biomedical science has evolved over time, the lines between clinical care, human research, and research not involving human participants have become blurred. The principles developed by this Commission, which flow from the concept of respect for persons, are described in detail in *New Directions: The Ethics of Synthetic Biology and Emerging Technologies*, and also apply when considering the ethics of whole genome sequencing.²⁰ As applied to the science of whole genome sequencing, these principles, along with the principle of respect for persons, guide us to focus on pursuing public benefit while minimizing both personal and public risk.

Respect for Persons

Respect for persons provides a strong, enduring, and widely accepted foundation for this report's recommendations for protecting individual privacy in the pursuit of public benefit. As set forth in the *Belmont Report*, respect for persons requires one to give great "weight to autonomous persons' considered opinions and choices while refraining from obstructing their actions unless they are clearly detrimental to others."²¹ The *Belmont Report* recognizes that not all persons can act as autonomous agents, and makes clear that there are special responsibilities to those who cannot.

Public Beneficence

Public beneficence asks us to pursue and secure public benefits and minimize personal and public harm. It encompasses society's duty to promote activities that have great potential to improve the public's well-being.²² Public beneficence also supports scientific enterprises that benefit society by increasing economic opportunities.

Responsible Stewardship

Responsible stewardship calls upon governments and societies to proceed prudently in promoting scientific advancement by taking into account the interests and needs of those who are not in a position to represent themselves such as children, the mentally ill, future generations, or individuals that may be unaware of risks. Responsible stewardship expresses a shared obligation to act in ways that demonstrate respect for such individuals. Emerging technologies present particularly profound challenges for responsible stewardship because our understanding of their potential benefits and risks is incomplete and uncertain.²³ This makes it all the more important that we take great care not to make choices that have a substantial chance of causing irreversible harm to current or future generations.

Intellectual Freedom and Responsibility

Intellectual freedom grants scientists, acting responsibly, the right to use their creative abilities to advance science and the public good. Sustained and dedicated creative intellectual exploration produces much of our scientific and technological progress. Intellectual responsibility, the complementary part of this principle, calls upon scientists to adhere to the ideals of research; to avoid harm to others; and to abide by all applicable policies, rules, and regulations. Institutions, policies, and practices of a free society—along with the many citizens who support them—collectively provide the means for scientists to do their work, and the culture that recognizes and upholds intellectual freedom. As a result, scientists bear profound collective responsibility to society.²⁴

The Commission endorses the principle of regulatory parsimony, which encourages fostering an achievable balance of intellectual freedom and responsibility. Regulatory parsimony calls for “only as much oversight as is truly necessary to ensure justice, fairness, security, and safety while pursuing the public good.”²⁵ In this spirit, policy makers are obligated to avoid restrictive rules that offer few benefits and hinder progress in science, medicine, and health care.²⁶

Democratic Deliberation

Democratic deliberation is an approach to collaborative decision making that embraces respectful debate of opposing views and active participation by citizens. Democratic deliberation warrants engaging the public and fostering dialogue among the scientific community, policy makers, and persons concerned with the issues raised by scientific progress.²⁷ The principle of democratic deliberation acknowledges that while decisions must eventually be reached, those decisions need not (and often should not) be unalterable, particularly when subsequent developments warrant additional examination. It is in the spirit of democratic deliberation that the Commission was created, has undertaken its work in publicly open meetings, and offered all of its reports to the President and members of the public.

Justice and Fairness

The principle of justice and fairness relates to the distribution of benefits and burdens across society. A commitment to justice and fairness is a commitment to ensuring that the unavoidable burdens of technological advances do not fall disproportionately on any particular individual or group, and that the benefits are widely and equitably distributed.²⁸ The principle of justice and fairness counsels that the numerous scientific advances stemming from investments in science and medicine should be made accessible to the broadest possible number of persons, consistent with the ability to advance science and medicine for the true benefit of the public.

The Commission’s Process

In concert with the principle of democratic deliberation, the Commission invited experts from the public and private sectors to inform their deliberations. Over the course of four public meetings, speakers addressed issues of privacy, consent, data security, access to whole genome sequence data, views of the patient advocacy community, and relevant philosophical topics (for

a complete list of Commission speakers, see Appendix III: Guest Presenters to the Commission Regarding Privacy and Whole Genome Sequencing). The Commission also posed a data call to the 18 Common Rule departments and agencies, asking them to identify relevant statutes, agency regulations, guidance documents, and policies that govern privacy and access to genetic information generally and whole genome sequence data specifically.²⁹ Finally, a Request for Information was published in the *Federal Register* that elicited many thoughtful comments from individuals and professional societies.³⁰

The Commission identified the field of whole genome sequencing as an important topic for consideration because this rapidly advancing technology raises many ethical issues that have not been fully addressed. After careful consideration of where it could make the greatest contribution at the present time, the Commission chose to focus on privacy rather than address ethical issues that are currently under consideration or have been addressed by other high-level groups or federal agencies, including commercial genetic testing and other important and controversial topics relevant to whole genome sequencing.³¹

In focusing on the potential risks to individuals' privacy, the Commission also recognizes the anticipated societal benefit of the scientific and medical applications of advances in whole genome sequencing. Reconciling these goals means addressing the competing concerns of ensuring confidentiality of whole genome sequence data, granting access to and use of these data, and empowering participants who want to share their data without weakening privacy protections for others. The Commission reviewed rules and regulations already in place that protect privacy and prevent discrimination based on genetic information (currently there are no state or federal laws explicitly addressing whole genome sequence data), and heard testimony about the technological security systems used to protect whole genome sequence data. The Commission heard from experts about the ways whole genome sequencing is being, and will continue to be, integrated into clinical care. In addition, the Commission heard from the patient advocacy communities who expressed their wishes for more participatory models of research.

About This Report

With its guiding principles in mind, the Commission sought to reconcile the anticipated societal benefit of the scientific and medical applications of advances in whole genome sequencing with the potential risks to individuals' privacy. Recognizing that our ethical obligations reach beyond what is legally enforceable, the Commission examined both the relevant ethical principles and the relevant legal requirements to offer guidance as to what (ethically) *ought* to be done and what (legally) *must* be done.³² This is the foundation upon which the Commission builds its recommendations, which apply to both the public and private sectors.

Accordingly, Section 1 deploys and applies the relevant ethical principles. Section 2 summarizes the legal framework governing whole genome sequencing and the legal protections provided for persons who decide to share their whole genome sequence data. Finally, Section 3 offers recommendations and guidelines that are aimed at reconciling the existing tension between minimizing risks to individuals and maximizing the anticipated future societal benefits of whole genome sequencing. The Commission intends that any changes resulting from these recommendations be prospective and not apply retrospectively to specimens already collected or stored in the research or clinical setting.

WORK OF PREVIOUS COMMISSIONS

Previous bioethics commissions have issued reports on topics related to genetics. In 1982, the President's Commission for the Study of Ethical Problems in Medicine and Biomedical and Behavioral Research published a report, *Splicing Life*, which addressed the ethical and social implications of genetic engineering (<http://bioethics.georgetown.edu/documents/pcemr/splicinglife.pdf>). In 1983, the same commission issued a report on the ethical, social, and legal implications of genetic screening, counseling, and education programs, titled *Screening and Counseling for Genetic Conditions* (http://bioethics.georgetown.edu/pcbe/reports/past_commissions/geneticscreening.pdf).

Genetic issues were not revisited until the National Bioethics Advisory Commission (NBAC) discussed the issue of human cloning in 1997, in its report *Cloning Human Beings* (<http://bioethics.georgetown.edu/nbac/pubs/cloning1/cloning.pdf>). In 1999, NBAC issued *Research Involving Human Biological Materials: Ethical Issues and Policy Guidance*, which focused on research involving human biological materials (<http://bioethics.georgetown.edu/nbac/hbm.pdf>).

In 2002, the President's Council on Bioethics took up the issue of human cloning in its report, *Human Cloning and Human Dignity: An Ethical Inquiry* (http://bioethics.georgetown.edu/pcbe/reports/cloningreport/pcbe_cloning_report.pdf). The Council also published *The Changing Moral Focus of Newborn Screening*, which sought to establish ethical principles to guide newborn genetic screening (http://bioethics.georgetown.edu/pcbe/reports/newborn_screening/Newborn_Screening_for_the_web.pdf).

SECTION 1. ETHICAL PRINCIPLES

Whole genome sequencing offers the promise of tremendous public benefit, and is expected to change substantially our ability to assess risk, diagnose, and treat disease. Achieving this public benefit requires that researchers have access to large amounts of whole genome sequence data and associated medical information to assess correlations between underlying genomic variants and expressed disease. While many of the potential benefits arising from whole genome sequencing will accrue to the broader public, the risks associated with collecting and sharing whole genome sequence data will be borne disproportionately by the individuals whose data are being shared.

Because whole genome sequencing begins with obtaining a sample from an individual, to reconcile anticipated public benefits with potential individual harms the Commission begins with the principle of respect for persons. Respect for persons is among the most enduring and widely accepted foundations for protecting individual privacy in the pursuit of public benefit, and it is well formulated in the *Belmont Report*, a declaration of ethical principles regarding research involving human participants.³³ Since biomedical science has evolved significantly since the *Belmont Report's* publication in 1979 from clinically focused research to research for public benefit, the Commission also applies five additional ethical principles which flow from the principle of respect for persons—as outlined in *New Directions: The Ethics of Synthetic Biology and Emerging Technologies*—to the field of whole genome sequencing.³⁴ These five principles—public beneficence, responsible stewardship, intellectual freedom and responsibility, democratic deliberation, and justice and fairness—apply well not only to

emerging biotechnologies, but also to scientific advancement and innovation generally. The Commission's five principles are thus a useful supplement to the Belmont principles for the purpose of assessing the ethics of whole genome sequencing.

"The state of technology is that data acquisition is now... relatively inexpensive, and while the free access to genetic data has many positive benefits, we need to represent, of course, the tension of that with all of the other personal privacy issues..."

Richard Gibbs, Wofford Cain Professor, Department of Molecular and Human Genetics; Director, Human Genome Sequencing Center, Baylor College of Medicine. (2012). Ethics and Practice of Whole Genome Sequencing in the Clinic. Presentation to PCSBI, February 2, 2012. Retrieved from <http://bioethics.gov/cms/node/658>.

In the case of whole genome sequencing, as is true for many emerging medical technologies, there are tensions between some of these principles. Two of the principles—public beneficence and intellectual freedom and responsibility—support the continued pursuit of whole genome sequencing research because of the promise of intellectual gains and substantial public benefit. Simultaneously, other principles—respect for persons, responsible stewardship, and justice and fairness—counsel the adoption of protections to minimize the privacy risks that could befall individuals. Drawing upon the process of democratic deliberation, the Commission sought to reconcile the potentially conflicting practical implications of these principles. It did so by taking into account various paths to the anticipated promise of this rapidly advancing technology, while respecting the ethical concerns of the increasing numbers of individuals facing the prospect of whole genome sequencing: concerns, for example, about confidentiality, information security, decisional autonomy, and freedom from unwanted intrusion into personal lives.

The Public Benefit of Whole Genome Sequencing

Scientists predict that whole genome sequencing research will foster better understanding of the genetic factors that contribute to human health and diseases including cancer, heart disease, diabetes, and neuropsychiatric conditions, as well as many rare diseases. Further, whole genome sequencing is expected to usher in an era of personalized medicine, providing information that might allow clinicians to tailor treatments or manage the health of individuals based on their genomic profile.

The Commission's recommendations regarding the continued pursuit of whole genome sequencing to advance medical science are based primarily on the principles of public beneficence and intellectual freedom and responsibility. Public beneficence gives rise to a societal and governmental duty to promote individual activities and institutional practices, such as scientific and biomedical research, that have great potential to improve the public's wellbeing.³⁵

Public beneficence also supports scientific enterprises that advance the common good by increasing economic opportunities, a criterion that whole genome sequencing satisfies.³⁶ The U.S. government invested billions of dollars in the Human Genome Project—a collaborative research project with the ambitious goal of sequencing the entire human genome. This

investment has since generated \$244 billion in personal income and \$796 billion in overall economic impact.³⁷ In 2010 alone, the human genome sequencing projects and associated research and industry activity directly and indirectly generated over 300,000 jobs and brought in tax revenue of \$3.7 billion. While not unique to whole genome sequencing, increased economic productivity is often a positive by-product, consistent with public beneficence, of scientific and medical progress.

Intellectual freedom grants scientists—acting responsibly—the right to use their creative abilities to advance science. Creative, sustained, and dedicated intellectual exploration is an essential aspect of scientific, technological, and clinical progress. At the same time, it serves to expand our general understanding of the world.

However, both public beneficence and intellectual responsibility, the complement to intellectual freedom, caution against pressing forward with whole genome sequencing without regard to negative consequences. The principle of public beneficence requires both that public benefits be secured *and* that public harms be minimized. Likewise, intellectual responsibility calls upon all researchers and clinicians—including their staff and the institutions that support them—to adhere to the ideals of research, one component of which is avoidance of harm to others.³⁸ Pursuing whole genome sequencing without considering potential harms would violate the clear and compelling mandates of public beneficence and intellectual responsibility.

Privacy Concerns Raised by Whole Genome Sequencing

Respect for persons includes respect for the dignity and privacy of individuals. As a result, respect for privacy assumes special salience in discussions about ethics and genetics. Because whole genome sequence data provide important insights into the medical and related life prospects of individuals as well as their relatives (who most often did not consent to the sequencing procedure), whole genome sequencing poses real privacy concerns. These concerns are compounded by the fact that whole genome sequence data gathered now might reveal important information, entirely unanticipated and unplanned for, as science progresses. The potential power of the information contained in whole genome sequencing substantially raises the privacy stakes of medical information.

“Public trust is fundamental to the ongoing support of these activities and to participant willingness to actually contribute to the research. And without the participant willingness to contribute to the research, we will not move forward at all.”

Laura Lyman Rodriguez, Director of Office of Policy, Communications and Education, National Human Genome Research Institute. (2012). Presentation to PCSBI, August 1. Retrieved from <http://bioethics.gov/cms/node/749>.

Privacy and the Law

Concerns about privacy are not new; worries about the proper boundaries between self, others, and government extend as far back in recorded human history as ancient Greece and Rome.³⁹ The central role of privacy in U.S. culture and ethics is reflected in the tone of its laws. The word “privacy” does not appear in the U.S. Constitution. However, as American courts and scholars have observed, the Bill of Rights implicitly recognizes the value of privacy and

rights of privacy through provisions guaranteeing: 1) freedom of speech, freedom of religious, political and personal association, and related forms of anonymity (First Amendment); 2) freedom from government appropriation of one's home (Third Amendment); 3) freedom from unreasonable search and seizure of one's body and property (Fourth Amendment); 4) freedom from compulsory self-incrimination (Fifth Amendment); 5) freedom from cruel and unusual punishment, including unnecessarily extreme deprivations of privacy (Eight Amendment); and 6) other personal freedoms (Ninth Amendment). In addition to the Bill of Rights, the Supreme Court and state courts have marshaled the due process clause and language of "liberty" of the Fourteenth Amendment to strike down laws interfering with autonomous medical, marital, sexual, and family decision making.

"With advancing technologies it's increasingly hard to keep secret our genetic information. There's more data-sharing... but that doesn't mean that we don't have privacy interests here, it just means that we may need more explicit protections of those interests."

Sonia Suter, Law Professor at George Washington University. (2012). Presentation to PCSBI, August 1. Retrieved from <http://bioethics.gov/cms/node/748>.

A number of U.S. states have explicit privacy protection provisions in their constitutions that apply to privacy violations by state and, in some cases, private entities. The common law of some states includes a breach of confidentiality tort. Most states recognize one or more right to privacy torts, first proposed in the 1890 article "The Right to Privacy" by Samuel Warren and Louis Brandeis. This seminal article persuasively argued that courts should recognize a "right to be let alone" against unwanted intrusion and publicity.⁴⁰ Today personal injury suits can be brought alleging intrusion upon seclusion; publication of private facts; publication placing one in a false light; and appropriation of name, likeness, or identity.

The United States takes a sectoral approach to regulating privacy, which means that the United States specifically regulates privacy concerns in particular settings as they arise. In the past four decades, in response to pervasive new technologies and related business practices, state and federal authorities have enacted many statutes and agency rules protecting the privacy of data related to health, education, finances, taxes, the federal census, video rentals, lie-detection, motor vehicle records, library records, and electronic and telephonic communications. This sectoral approach means that a number of areas that have no specific laws currently do not receive even baseline privacy protections. By contrast, Europe regulates privacy comprehensively, providing privacy protections that are consistent across different types of data or information.⁴¹

The Meanings of Privacy

Privacy and associated terms, including confidentiality, anonymity, choice, and data protection, refer to related concepts. Discussions about ethics and whole genome sequencing sometimes inappropriately use these terms interchangeably. To enable clear ethical analysis in this report, we provide basic definitions of the family of privacy terms applicable to our work and map their relationships. Scholars differ in their precise definitions of the terms we use, but the language we present in this report is consistent with a general consensus view. The following definitions are meant to show how the Commission uses these terms and to help

guide future discussions regarding ethics and genetics. They should not be taken as formal arguments for precise definitions.

Restricted Access

The term *privacy* is used here (and in many ethical and legal contexts) broadly to mean states of affairs by virtue of which the accessibility of persons, personal information, or personal property is limited or restricted. What is valued as “personal,” “sensitive,” or “intimate” may be restricted by virtue of, for example, spatial distances, physical barriers, electronic passwords, social norms, or customs. In the United States and other developed societies, health information is widely considered personal, sensitive, or intimate, and genetic information especially so.

The term *informational privacy* refers generically to restricted access to information or data. “Confidentiality,” “anonymity,” and “data protection” are specific ways to protect informational privacy in the broad sense, with special relevance in clinical and health research settings.

Confidentiality is used to denote restricting access to information or data to groups of specifically authorized recipients. In the medical context, health information is often limited by custom to close family and friends and by law to health practitioners, insurers, and professional researchers. Patients and research participants may even choose to keep health conditions secret from intimate kin by deliberately concealing the information. Confidentiality is closely connected with trusting relationships. One can share private information with another person on the understanding that he or she can be trusted to keep that information secret (i.e., will not divulge it to others). Patients entrust clinicians with medical information provided that they have a “need to know,” and understanding that the clinician will keep the information confidential. In the context of whole genome sequencing, data must be kept confidential; databases must be secure and information must not be divulged to unauthorized users.

Anonymity is used to denote restrictions on access to personally identifiable information pertaining to individuals or groups, achieved through intentionally disguising or removing identifiers. A health record can be made more anonymous, for example, by removing a patient’s name, address, or social security number.

Data protection refers to measures designed to thwart deliberate or accidental disclosures of confidential or anonymous information. Health data that are electronically stored or transmitted can be protected with computer passwords and encryption. Health care providers employ technology to protect data, but ethical norms and business practices can also protect data from unauthorized access, use, and disclosure.

Autonomy

The term “privacy” has a second distinct use in ethics and law. Privacy is a rough synonym of autonomy with respect to self-regarding conduct and intimate relationships. Here, privacy denotes the absence of substantial government or other outside interference with individuals’ decisions and choices. In traditional bioethics, the “privacy” at issue in euthanasia, birth control, and consent to research is this second understanding of privacy, which involves the ability to make autonomous decisions.

We note that there are other uses of “privacy,” some health-related, that do not play a major role in this report. Seeking greater precision and focus, privacy scholars and the courts commonly qualify the term “privacy” using descriptive adjectives. Indeed, they commonly

speak of *informational privacy* in relationship to the collection, use, and sharing of information or data. They speak of *physical privacy* in relation to observing, concealing, and touching the human body, such as entering hospital rooms or respecting patient modesty. They refer to *spatial, geographical, and locational privacy* in relation to GPS and beeper technologies. They speak of *associational privacy* in relation to affiliation with like-minded people. They recognize *decisional privacy* in relation to independent decision-making. Less commonly, privacy scholars and the courts distinguish *proprietary privacy* in relation to repositories of personal identity and genetic ownership claims. And finally, they identify *intellectual privacy* in relation to interests in freedom of thought, conscience, and the right to read and access knowledge.

There is ample debate and disagreement about the value of particular privacies and the basis for laws and policies promoting or regulating each type of privacy. In this report, the Commission focuses on *informational* and *decisional* privacy as they pertain to whole genome sequencing. We use the term “privacy” in reference to both limited access to genetic information and data, and to the absence of interference with decisions about the collection, use, and sharing of genetic information. A person whose whole genome is sequenced might have both *decisional privacy* concerns (about who is permitted to decide whether whole genome sequencing data are shared) and *informational privacy* concerns about whether such data will be shared in confidence, securely, or in de-identified form.

Although the precise contours and content of privacy have changed substantially over time, with shifts in culture as well as technology, intense and widespread human interest in the protection of privacy is abiding, not only in the United States, but also around the world. Privacy protections promote a set of highly prized values. Although modern technology can facilitate unobserved and uninvited intrusions into homes, for example, what individuals choose to do in such a domain is generally valued as a matter of “privacy” and deemed legitimately “private,” unless that behavior violates particularly weighty ethical or legal limits. That is, there are constraints on what behavior can be considered legitimately private. In the inclusive understanding of what falls under the privacy umbrella we adopt here, what individuals choose to do at home is presumptively confidential, anonymous, intimate, secure, free from unwanted intrusion, and/or subject to decisional autonomy.⁴² Concern for privacy values (while additionally a means of enabling privacy at home and other vital privacies) also incorporates the increasingly elusive ideal of control over the flow of information regarding oneself, again subject to broad ethical and legal limits.⁴³

The Value of Medical Privacy

The Commission agrees that respect for patient and participant privacy can greatly benefit individuals and the general public. Under the principle of respect for persons, and for the sake of public beneficence and justice and fairness, those who collect, use, or share health data should employ practices that include confidentiality, anonymity, and informed consent to shelter clinical patients and research participants from the unwanted glare and control of others. It is important to ensure that respect for patient and participant privacy not be compromised, not only in clinical care and research, but also in the publication or archiving of medical lectures, scholarly articles, and personal papers. Medical privacy remains an important ethical principle, despite the recognition that many people voluntarily share their health information or data, including genetic information and data, and despite the practical reality that modern institutional practices presuppose that a great deal of sensitive health information can and will be lawfully shared among providers, insurers, researchers, and the government.

Medical privacy has many varieties of recognized public value. First, medical privacy encourages individuals to seek medical care. Individuals will be more inclined to pursue medical attention if they believe they can do so on a confidential basis. Practicing confidentiality assures that, in most cases, a patient can choose when to disclose an illness, condition, or genetic status. Confidentiality and anonymity enable individuals to exercise constitutionally protected liberties of autonomous medical decision-making by safeguarding information they do not choose to share because it is embarrassing or would expose them to discrimination or disapprobation.

“It just seemed safer to keep it to myself...I didn’t know what somebody would do with that information in the future...and I was very concerned about it.”

Victoria Grove, introductory vignette, referring to her decision to keep secret her positive genetic test for alpha-1 antitrypsin deficiency.

Second, medical privacy encourages frank disclosures in clinical and research settings. Individuals seeking care can be open and honest if they can trust that facts reported to or uncovered by clinicians or researchers will not be broadcast to the world at large. People are often embarrassed by symptoms, histories, and prospects of illness. Individuals concerned about discrimination, shame, or stigma have an interest in controlling the flow of information about their health. Some patients and participants believe they own personal information about themselves, especially genetic information, and should be able to control its release.

Third, if individuals believe they can decide whether to share data, information, and biospecimens under conditions of confidentiality, anonymity, and informed consent, they might be more likely to participate in research. In the context of health research, ethics committees and institutional review boards properly require researchers to protect the privacy of research participants and their medical records. Obligations of privacy may require the use of coded information rather than names or “de-identification” procedures such as data aggregation. Some have argued that researchers must publish genomic data in ways that obscure the identities of whole families. Even statistical use of individuals’ health data has raised privacy concerns, as some have argued that for cultural or social reasons individuals might have an ethical interest in the uses of data sets without personal identifiers that include data about them.⁴⁴

Fourth, alleviating the concerns about exposure and discrimination that keep patients away from clinicians enhances confidentiality, which can further the goals of health care cost savings by ensuring that patients seek early medical care rather than waiting until their conditions worsen and require more dramatic medical intervention.⁴⁵

The Commission recognizes that privacy, like most values, has ethical as well as practical limits. It is not an absolute public good. Certain diseases, conditions, and prescriptions must be reported to government to protect public health and safety. Health care providers and responsible adults are ethically obligated to report evidence of child neglect and abuse uncovered in treatment. Mental health providers have an ethical duty to warn police or potential victims of the credibly violent intentions of patients with mental illness. Situations arise in which medical confidentiality cannot be preserved because the media has a right to publish information or legal authorities have the authority to subpoena information for use in legal proceedings and investigations. Members of the military and civil servants serving in war zones may be also required to undergo mandatory genetic biobanking or testing for varied purposes.

Privacy in Whole Genome Sequencing

Currently, whole genome sequencing involves generating, storing, sharing, and analyzing large amounts of data. Although members of the public express general comfort with the idea of sharing genomic data in biorepositories, privacy ranks among participants' highest concerns.⁴⁶ Data also show that for many, privacy concerns are an important obstacle to participation in large cohort studies.⁴⁷ Although 60 percent of people surveyed said they would participate in a study that involved storing data in biorepositories, 91 percent of those potential research participants would be concerned about privacy.⁴⁸ Additional data indicate that although a large majority of survey participants trust clinicians and researchers, they are concerned that results of genetic tests could end up in the wrong hands and be used against them.⁴⁹ Most of the people interviewed following enrollment in one sequencing study indicated that their primary concern was that they be informed if there was a possibility that their data would be shared with other researchers and that it was important they maintain some control over who could have access to their genomic data. The participants wanted insurance companies and employers to be excluded from access to these data, but were comfortable with data sharing within the research community.⁵⁰

Informational and decisional privacy concerns about the unauthorized disclosure or misuse of whole genome sequence data are not only common and intensely important in the minds of potential research participants, they are also objectively linked to the potential for serious harms from such disclosure and misuse. Potential harms include the risk of lost opportunities in employment, long-term health care, disability and life insurance, loan approvals, education, sports eligibility, military accession, and adoption eligibility.⁵¹ In areas that are far less amenable to any legal protection or recourse, individuals could find themselves facing social stigma from disclosure of sensitive genomic information, and subsequent disruption of their home, family, and community life.⁵² Risks that are more internal to, and variable among, individuals include being subject to psychological harms upon learning information that can be difficult to bear, including that one has a predisposition to a disease such as cancer or Alzheimer's disease. Because whole genome sequence information directly implicates relatives, psychological harms often are not limited to the person whose genome is voluntarily being sequenced and publicly disclosed. Even individuals who learn that they do not carry a harmful variant may experience "survivor's guilt" if another family member is affected.⁵³

To date, the number of documented cases of discrimination on the basis of genetic test results is small.⁵⁴ This might be due to the relatively few conditions for which there are currently definitive genetic tests, coupled with the expense and difficulty of conducting these tests. As a result, genetic information is rarely available to third parties. Another reason for the small number of reported cases, now and potentially in the future, might be the difficulty of uncovering and documenting discriminatory use of data.⁵⁵ It is also possible that such discrimination might not occur, either because there are other more definitive bases on which to make insurance or employment decisions, or because all individuals have some form of disease predispositions. Regardless, legitimate concerns remain about the potential for differential treatment of individuals based on their genomic information, even if legally prohibited discrimination rarely occurs. If individuals lack assurances against misuse of their genomic information, their privacy concerns might motivate them to not share their whole genome sequence data, which could harm the research enterprise that generates life-saving discoveries.

Privacy and the Ethical Principles

A robust set of ethical principles—respect for persons, responsible stewardship, and justice and fairness—supports the adoption of norms to minimize the privacy risks that could befall individuals while enabling research and clinical care for public benefit to continue. Respect for persons requires one to give great “weight to autonomous persons’ considered opinions and choices while refraining from obstructing their actions unless they are clearly detrimental to others.”⁵⁶ Exercising autonomy includes self-determination, which requires that persons be allowed to make “important decisions about one’s life for oneself and according to one’s own values or conception of a good life.”⁵⁷ Respect for persons highlights an individual’s autonomy and recognizes that we should respect individuals’ ability to decide for themselves what they value, and how and when to act on those values. For example, an autonomous person should be able to decide whether to undergo a medical procedure based on personal considerations of risks, benefits, costs, and cultural and religious views. Forcing an individual to undergo a procedure, even for their medical benefit, would violate that person’s autonomy and would fail to demonstrate respect for the individual as a person. Respect for persons also encompasses respect for the individual’s dignity and privacy. Therefore, violation of an individual’s privacy, such as the misuse or unauthorized disclosure of whole genome sequencing data, demonstrates a violation of the principle of respect for persons.

Governments and societies that exercise responsible stewardship accept a duty to proceed prudently in promoting scientific advancement and emerging technologies. They recognize a shared duty to act in ways that demonstrate concern for all those who might be affected, and especially for those who are not in a position to represent themselves (e.g., children, the disenfranchised, vulnerable populations, and future generations). Rapidly advancing technologies such as whole genome sequencing present profound challenges for responsible stewardship because our understanding of the potential benefits and risks is largely incomplete and uncertain.⁵⁸ This makes it important that governments and societies take great care not to make decisions that have a substantial chance of causing irreversible harm to current or future generations, and especially those who have little or no say over such decisions. Responsible stewardship advises against decisions that are entirely precautionary (no action without complete certainty of security) or entirely proactionary (no limitations on science). Heeding the principle of responsible stewardship therefore neither thwarts the development of new scientific enterprises nor lets science advance unchecked on the fallible assumption that it is safe.

The principle of justice and fairness is, in important part, a commitment to ensuring that the unavoidable burdens of technological advances do not fall disproportionately on any individual or group, and that the benefits are widely distributed.⁵⁹ The principle of justice and fairness encompasses the idea of fair distribution in that it demands society ensure that risks not be disproportionately borne by any particular group and strive for “the broadest distribution of beneficial technologies.”⁶⁰ As such, the principle of justice and fairness entails protection for those who decide to share their whole genome sequence data to reduce the chances that they will be harmed by unauthorized disclosure or misuse.

These three principles, taken together, suggest that individuals are entitled to privacy protections that prevent undue and disproportionate burden. But these protections are not absolute. Prohibiting all gathering and sharing of whole genome sequence data would protect privacy absolutely, but still would fail to adequately respect persons. A total prohibition prevents individuals from choosing to participate in whole genome sequence research, even if they consider themselves adequately protected; it also fails to take into account individuals’

other interests, such as an interest in excellent medical care. Respect for persons demands respect both for individuals' privacy *and* for their interest in benefitting themselves and others from medical advances.

The Commission emphasizes that there is extremely good reason for individuals to choose to share information in a context where there is adequate protection for individual privacy: whole genome sequencing has the potential to be of substantial public benefit. The ability to share information is the sort of important decision that is central to autonomous action, which respect for persons commits us to recognize.

Respect for persons supports giving persons the opportunity to share their whole genome sequence information for scientific advancement, subject to strong baseline privacy protections. At the same time, individuals have a responsibility to safeguard their privacy as well as that of others, by giving thoughtful consideration to how sharing their whole genome sequencing data in a public forum might expose them to unwanted incursions upon their privacy and that of their immediate relatives. To be indifferent to the implications of disclosure of sensitive data and information about one's self is to act irresponsibly. That being said, it can be good and virtuous to share sensitive data about oneself in appropriate circumstances, for example, for the good of public health research or public education.

To determine what baseline privacy protections should be, we need to distinguish between access to, use of, and possession of whole genome sequence data. To *possess* whole genome sequence data is to have a copy of the data file and, therefore, to have access to it at any time. Having *access* to data implies the ability to manipulate and work with the data files. It is possible to access data that one does not possess; a researcher might be allowed to access data files in a secure database to address research questions without keeping a copy of the data. One can have access to data even if one does not (and either ethically or legally cannot) use it, as when whole genome sequence data are stored on a server available to download, but one does not download them. The *use* of data refers to seeking answers to questions by analyzing the data. A researcher could use data in a protected database without having either access to or possession of the data by submitting a query to the database manager and then receiving the results of the query from the database manager. In these ways, it is possible to allow researchers to work with whole genome sequencing data through access to or use of the data while maintaining the security of the data themselves and protecting the privacy of the individuals who contributed to the database. The confidentiality of information or data about persons can be maintained through a number of means designed to prevent unauthorized access to the data: these means are collectively called informational security or data security. Examples of data security mechanisms include legal limitations, locked drawers, and computer firewalls.

Presentations to this Commission indicated that whole genome sequence data could be used without actually possessing it: that is, technologies already are being developed to allow researchers to have limited computational access to select whole genome sequence data sets without physically transferring possession of all data files in the set.⁶¹ The researcher would be able to use the data for analysis, but would not maintain possession of the data. This means that possession of genomic information is neither necessary nor sufficient for its use. As with control of information, the use of information (including misuse and unauthorized use) in some cases will be of greater ethical salience than either access or possession.

Reconciling Competing Ethical Claims

The principles of public beneficence and intellectual freedom and responsibility support continued pursuit of whole genome sequencing to advance scientific understanding and medical progress. But these principles have components that suggest such pursuits should not be unrestrained. The positive argument for restraint is founded upon the principles of respect for persons, responsible stewardship, and justice and fairness, which together require implementing privacy protections and minimizing the chance of harm to individuals. But these principles do not suggest that privacy protections should erect absolute barriers to voluntary data sharing.

In moving forward with whole genome sequencing, respect for persons requires informing individuals about the foreseeable consequences of their decision to share their genomic data, including who has access to their whole genome sequence data and how these data might be used in the future. Respect for persons also counsels individuals who collect samples to determine patient and research participant preferences at the time samples are obtained so that they can choose whether to participate, or whether feasible limits on the use of their whole genome sequence data can be agreed upon. Providing individuals who are choosing whether to share whole genome sequence data with the information necessary to make a fully informed decision about the potential consequences—including who can access the data and how the data will be used—allows individuals to make an autonomous decision. The principle of respect for persons applies to all whole genome sequence data regardless of whether they were obtained in a research or a clinical context.

The Commission's principle of regulatory parsimony calls for "only as much oversight as is truly necessary to ensure justice, fairness, security, and safety while pursuing the public good."⁶² Regulatory oversight is appropriate in certain contexts—for example, disallowing certain types of research or permitting other types of research only when certain conditions are met. But some aspects of research—including data security protections for whole genome sequence data—remain outside most regulatory frameworks. For otherwise unregulated aspects of research, informed consent is one mechanism by which individuals can protect their own privacy. By informing individuals about the potential risks and benefits of participation in whole genome sequencing, along with information about the security protections in place, individuals can autonomously choose whether to provide a biological sample for use in whole genome sequencing research. In this way, informed consent is one means of reconciling the public good that can come from whole genome sequencing with the potential harms to individual privacy.

NEWBORN SCREENING

In the case of *Beleno v. Texas Department of State Health Services* parents sued, claiming that the Texas Department of State Health Services collected and stored newborn blood samples, subsequently making them available for research purposes, without seeking parental consent. The parents argued that the lack of proper consent was a violation of privacy. The out-of-court settlement that was reached resulted in the destruction of 4 million similar specimens that had been collected without parental consent.

Sources: *Beleno v. Lakey*, No. SA-09CA-188-FB (W.D. Tex. Sept. 17, 2009).; and Aaronson, B. (2010, December 8). Lawsuit alleges DSHS sold baby DNA samples. *The Texas Tribune*, TribBlog. Available at: <http://www.texastribune.org/texas-state-agencies/departamentof-state-health-services/lawsuitalleges-dshs-sold-baby-dna-samples/>.

The Commission is also mindful of democratic deliberation, an approach to collaborative decision-making that embraces respectful debate of opposing views and active participation by citizens. Democratic deliberation warrants engaging the public and fostering dialogue among the scientific community, policy makers, and those concerned with the issues raised by whole genome sequencing.⁶³ In this spirit, the Commission sought input from a broad range of voices, including members of the patient advocacy community calling for more participatory models of research and from researchers who feared further administrative burden.

The principle of democratic deliberation acknowledges that while decisions (e.g., recommendations, policies, and guidance documents) must be reached in a timely manner, those decisions need not—and generally should not—be unalterable, particularly when relevant new information emerges. Modern societies change rapidly, especially in the domain of science and technology, and decisions in changing realms are best considered provisional rather than permanent. Researchers and clinicians must be particularly mindful of the deliberative value of provisionality, of being tentative or temporary, as whole genome sequencing moves from the realm of research and enters the broader clinical context.⁶⁴ The transition is already raising new challenges, and the policies that were once created with the assumption that the research realm is clearly and cleanly separated from clinical contexts may no longer be either sustainable or desirable due to the reciprocal relationship that has developed between them. Clinical samples, stripped of identifiers and transferred to genomic databanks and biorepositories for broader use by researchers, contribute to the common good by making possible research that could not be done without large numbers of samples from which to generate data. Subsequently, medical benefits developed as a result of such research will be available to the broader population including the persons from whom the deidentified clinical samples were taken.

Conclusion

The Belmont principles and the principles articulated by this Commission suggest ethically important and practically useful guidelines for whole genome sequencing. Chief among these is that the principle of respect for persons requires strong baseline protections for privacy and security of data, while public beneficence requires facilitating ample opportunities for data sharing and access to data by clinicians, researchers, and other authorized users. Respect for persons further requires that any collection and sharing of an individual's data be based on a robust process of informed consent. The principle of responsible stewardship calls for oversight and management of whole genome sequence information by funders, managers, professional organizations, and others. The principle of intellectual freedom and responsibility provides further support for pursuing whole genome sequencing and seeking models for broad data sharing by promoting regulatory parsimony. Democratic deliberation is the foundation of the process that gave rise to this document, and others like it, and will continue to be the foundation moving forward. Democratic deliberation urges all parties to consider changes to policies and practices in light of the evolving science and its implications for enduring ethical values.

Finally, the principle of justice and fairness requires that we seek to channel the benefits of whole genome sequencing to all who may potentially benefit, and ensure that the risks are not disproportionately borne by any particularly vulnerable or marginalized group.

SECTION 2. POLICY AND GOVERNANCE

This section describes current policy and legal protections of genetic information and the ways in which genome sequence data are shared in the United States. There is no comprehensive federal law that protects genetic privacy. The Genetic Information Nondiscrimination Act (GINA) prohibits discrimination by employers and health insurers based on the results of genetic tests, but does not provide privacy protections. In addition, GINA does not address the complexity of large-scale genomic data. Many states have laws governing genetic information and some of these laws provide privacy protections, but the laws vary greatly from state to state. As a result, our laws lack the specificity required to encourage participation and secure public benefits from this emerging science, while still ensuring the protection of privacy.

To gain the most benefit from recent innovations in whole genome sequencing, researchers need as much data as possible, derived from broad public participation in whole genome sequencing research. Widespread participation will be achieved only if participants trust the research enterprise and are comfortable that their privacy interests are protected. Currently, the patchwork of state and federal laws does not provide uniform protection of genomic data privacy. Protecting privacy interests of individuals requires a spectrum of conditions to be in place, including ethical and trustworthy behavior by researchers and clinicians, sufficient security of information technology, and policies and laws that hold violators accountable.

Privacy Concerns about Genetic and Whole Genome Sequence Data

For as long as the nature of genetics and heritability has been understood, there have been concerns about misuse. During most of the 20th century, erroneous notions about genetics led to eugenic policies based on the idea that genetic “inferiority” should be eliminated. Since the launch of the Human Genome Project in 1990, scientific knowledge about genetic information has grown exponentially, especially in identifying genetic variations that cause disease. This new information has resulted in a heightened concern about privacy, and the implications of others knowing an individual’s genetic information.

To draw meaningful conclusions and answer broad research questions, researchers aggregate and share whole genome sequence data from large numbers of individuals. To garner widespread participation in research and maintain trust in the enterprise, users and holders of whole genome sequence data must guide themselves according to at least three facets of privacy and confidentiality. The first facet, the individual, requires fostering ethical behavioral norms for researchers and clinicians. Participants, patients, and consumers must be assured that those who have contact with identifiable data intend to use them in an ethical manner—namely, only for those uses for which the participant, patient, or consumer has given consent. Many individuals trust researchers and medical professionals to consider their needs along with the

greater good, despite substantial privacy concerns. A 2010 study of research participants' views on genomic research indicated that, while individuals expressed concerns about privacy and data security, they also understood the value of sharing whole genome sequence information. Overall, concerns about privacy did not outweigh their sense of the importance of sharing genomic data in the interest of a larger social good.⁶⁵

The second facet of privacy protection is information technology. Participants and patients must be assured that their data are secure. A 2006 survey queried the public's wariness about health information technology systems and found that 80 percent of survey participants were concerned about identity theft and fraud, 77 percent about health information being used for marketing purposes, and 55 percent about health information being misused by insurers or employers.⁶⁶ These concerns highlight the need for secure information technology systems tailored to sensitive biomedical information, including whole genome sequence data and information. These concerns build upon the need for fundamental trust in the ethical behavior of data users and in the security of the systems that store these data—participants and patients should be assured that they can rely on their consent to allow identified data to be used for certain purposes and not for others.

The third facet of privacy protection is policy. Policy-level protection requires that systems be in place to provide clear institution-level expectations of training and preparation to handle whole genome sequencing data and information, to ensure an atmosphere of trust and an expectation of security, and to provide recourse should individual and information technology privacy protections fail.

While rapid advancement of genomic science in the past decade has led to vast potential for valuable research and societal benefits through medical advances, privacy and confidentiality concerns persist. Without reliable protection from potential harms, perceived and real fears of privacy violation and discrimination could cause individuals to balk at sharing their whole genome sequence data, thus stifling scientific progress.

Current Sharing of Specimens and Whole Genome Sequence Data

The past few years have seen the rise of sharing whole genome sequence data through biorepositories (facilities that store large numbers of physical biospecimens containing genetic material and associated data and information that researchers can access) and databases. Biorepositories are categorized generally into four groups: disease-specific (e.g., cancer databases); longitudinal population studies (e.g., the United Kingdom biorepository); isolated populations (e.g., the Faroe Islands); or twin registries, used to distinguish between genetic and non-genetic bases for disease.⁶⁷ Biorepositories often have different missions and different governance structures and must reconcile the rights of individuals with potential societal benefit accordingly. Other organizations, such as academic institutions, government agencies, and private not-for-profit entities, store data in databases—repositories that do not contain physical biospecimens, but rather electronic versions of genome sequence files. For many purposes, it is no longer necessary to maintain actual stored DNA from an individual once the genome sequence data have been collected, because it is easier to share electronic data files than physical specimens.

Despite these differences, biorepositories and their associated databases share some commonalities. The collection of specimens and data and subsequent storage in biorepositories

and databases give rise to risks that might include minor harm to the donor in obtaining the biospecimen (such as bruising upon blood withdrawal); nonphysical harms such as discrimination, stigmatization, and untoward psychological impact upon discovering unwelcome information; group harms, like those incurred by the Havasupai; and ethical harms that arise when individuals are not treated with respect and dignity.⁶⁸ Various laws and regulations govern the ways that these data currently are collected, shared, and used in the United States and around the world.

U.S. Federal Agency Activity

In order to inform this report, the Commission sought information about human whole genome sequencing research sponsored by the 18 U.S. Common Rule agencies, and related privacy protections of the data generated in the research they sponsor (see Table 1). The Commission supplemented these responses with publicly available information.⁶⁹

Twelve of the responding agencies stated that they do not conduct research involving human genomics, have not advocated formally for policy changes, and do not anticipate policy changes related to genomics.⁷⁰

Table 1. Human genomics research in federal common rule agencies

Department/Agency	Conducts/ Sponsors Research Involving Human Genomics	Anticipates Proposing New Policies
Agency for International Development (USAID)	No	No
Central Intelligence Agency (CIA)	No	No
Consumer Product Safety Commission (CPSC)	No	No
Department of Agriculture (USDA)	No	Yes
Department of Commerce (DOC)	No	No
Department of Defense (DOD)	Yes	Yes
Department of Education (ED)	No	No
Department of Energy (DOE)	No	No
Department of Health and Human Services (HHS)	Yes	Yes
Department of Homeland Security (DHS)	Yes	No
Department of Housing and Urban Development (HUD)	No	No
Department of Justice (DOJ)	Yes	No
Department of Transportation (DOT)	Yes	No
Department of Veteran Affairs (VA)	Yes	As needed
Environmental Protection Agency (EPA)	No	No
National Aeronautics and Space Administration (NASA)	No	As needed
National Science Foundation (NSF)	No	No
Social Security Administration (SSA)	No	No

Six agencies—the Department of Homeland Security (DHS), the Department of Defense (DOD), the Department of Justice (DOJ), the Department of Health and Human Services (HHS), the Department of Veterans Affairs (VA), and the Department of Transportation (DOT)—currently sponsor genetic and/or genomic studies, and five maintain or support biorepositories and databases.⁷¹ The confidentiality, privacy, and security of samples and data stored by federal agencies are governed by a baseline of laws and regulations, including the

Health Information Technology for Economic and Clinical Health (HITECH) Act, the E-Government Act, the Federal Information Security Management Act, the Health Insurance Portability and Accountability Act (HIPAA), the Privacy Act, and the Policy for Privacy Act Implementation and Breach Notification.⁷² Several agencies have additional mission or function-specific policies that govern the entities they fund that perform whole genome sequencing studies.

DOD uses large-scale genomic data in the DNA Dog Tag program, a mandatory program that has collected and stored blood and tissue samples from every member of the Armed Forces since 1991. The program does not give service members the opportunity to opt out of this collection. DNA is extracted from the samples only if needed to assist in identifying human remains. Specimens stored in the repository are not used for any other purpose unless approved by the Assistant Secretary of Defense for Health Affairs. DOD has several policies for protecting and securing genetic information that address disclosure, medical records, and information systems.⁷³ DOD expects to increase the use of whole genome sequencing for forensic applications related to human remains identification.⁷⁴

Agencies within HHS routinely use or sponsor whole genome sequencing. The confidentiality and security of samples and data used by HHS are covered both by HHS-wide and agency-specific policies, laws, and regulations.⁷⁵ For example, one HHS agency, the Centers for Disease Control and Prevention coordinates efforts to conduct whole genome sequencing of residual dried blood spots archived by states after newborn screening with parental consent.⁷⁶ The Centers for Disease Control and Prevention also collects DNA specimens for its National Health and Nutrition Examination Survey, and the confidentiality of identifiable information collected is protected under the Public Health Service Act.⁷⁷ Another HHS agency, the National Institutes of Health (NIH), devotes resources to studying the influence of genetic factors on human health and illness. NIH has established a number of genetic data repositories, most notably the database of Genotypes and Phenotypes (dbGaP).⁷⁸ dbGaP stores various types of genetic information, including whole genome sequence data. Access to data stored in dbGaP is two-tiered: open access, which grants the public access to information about study design and aggregate phenotypic information; and controlled access, which grants researchers access to information including de-identified genotypes and phenotypes of individual study participants.⁷⁹ Researchers who seek controlled access must submit formal research requests that are reviewed and approved by NIH Data Access Committees.⁸⁰ NIH has implemented policies and procedures to which every researcher with access to dbGaP must adhere to protect the privacy and confidentiality of genetic and, specifically, whole genome sequence data.⁸¹

The Combined DNA Index System (CODIS) is a DNA database funded by the Federal Bureau of Investigation (FBI), a Department of Justice agency. CODIS consists of DNA profiles from the Convicted Offender Index, the Forensic Index, the Arrestee Index, the Missing or Unidentified Persons Index, and the Missing Persons Reference Index. The National DNA Index contains almost 11 million offender profiles.⁸² CODIS does not contain personally identifiable information, nor does it contain whole genome sequence data. To further protect the data in CODIS, access to computers containing CODIS software is limited to authorized users approved by the FBI. Unauthorized disclosure of DNA data in the National DNA database is subject to a criminal penalty.⁸³

DHS uses genetic data, but not whole genome sequence data, in several ways. DHS collects DNA from individuals who are arrested, facing charges, or convicted of federal or military

crimes. DHS also collects DNA from non-U.S. citizens who are detained under the authority of the United States. U.S. Citizen and Immigration Services can require genetic testing to establish familial relationships to determine immigration or refugee status. Finally, DHS is piloting a program for overseas refugees who request asylum for family members; refugees can be asked to voluntarily undergo familial relationship testing using a portable DNA testing device. DHS does not generally maintain or have access to the genetic information it collects from individuals; it sends DNA samples to the Department of Justice for processing and entry into CODIS.⁸⁴

VA has active research and clinical genomics programs. In 2012, VA launched the Million Veteran Program, which aims to collect one million biospecimens from veterans to explore the role of genes in health and disease.⁸⁵ VA treats genomic data as personally identifiable medical information protected under HIPA A, although it stores the biospecimens securely and without other traditional identifiers such as name. VA has applied for a Certificate of Confidentiality from NIH, and has several additional departmental policies to protect the privacy of identifiable medical information.⁸⁶ The Million Veteran Program database is accessible only to authorized researchers for projects that have been approved by appropriate VA oversight committees.⁸⁷

The Federal Aviation Administration, a Department of Transportation agency, is researching human factors related to aviation safety from a gene expression viewpoint (gene expression is the process by which genes are translated into proteins).⁸⁸ Specifically, the Federal Aviation Administration is researching how alcohol use, fatigue, and cosmic radiation change gene expression and is correlating changes in gene expression to human performance to improve aviation safety. The Federal Aviation Administration's intent is to have unique sets of molecular markers for these factors that are generally applicable across the broad human genetic spectrum with a high degree of specificity. Genetic data collected by the Department of Transportation are subject to a number of federal data security policies.

INTERNATIONAL BIOREPOSITORIES

While the United States has many publicly funded biorepositories of limited size, a number of countries have implemented or attempted to implement population-wide biorepositories.

In the United Kingdom, for example, a half million volunteers are being recruited to donate genetic material to be linked to medical records in a biobank. The biobank will obtain informed consent from its participants and will allow for withdrawal from the database. Participants can request: 1) complete withdrawal and destruction of existing samples, 2) discontinued participation but continued use of existing data, or 3) no further contact, but continued use of existing data.

Source: UK Biobank [website]. Retrieved from <http://www.ukbiobank.ac.uk/>.

Commercial Genetic Testing Companies

Over the past few years, accessibility and availability of commercial genetic testing and genotyping has greatly expanded. Companies like 23andMe, Navigenics, and Ancestry DNA provide an array of services including paternity testing, testing for predisposition to certain diseases and traits, genealogy and ancestry information, pharmacogenomics (the influence of genomic factors on drug response), and even private forensic tests to establish profiles of suspects not included in the federal CODIS database.⁸⁹ Most commercial genetic testing

companies currently do not conduct whole genome sequencing. Instead, they analyze hundreds of thousands of single-nucleotide polymorphisms (SNPs) or discrete variants throughout the genome, which they describe as “genotyping.”⁹⁰ Commercial genetic testing companies often conduct research that uses biospecimens submitted by their customers. Most recently, 23andMe patented one of their research discoveries, “polymorphisms associated with Parkinson’s disease.”⁹¹

Commercial entities face issues of data maintenance and storage similar to those of government-sponsored biorepositories and databases. They collect and analyze genetic and genotypic data and maintain electronic databases of consumer data. In addition, many commercial genetic testing companies have a research arm that conducts research on consumer data in biorepositories. Currently, there are no overarching federal or industry guidelines indicating how commercial genetic testing companies should operate, what privacy controls they should implement, or what limits they should put on the use of genetic data and information. Like government-sponsored biorepositories and databases, they can protect consumers by developing systems to promote ethical and trustworthy behavior of employees, strengthening the security of information technology systems, and developing company policies that hold violators accountable.

Privacy Regulations

Individuals who share their genomic information, like those who share any medical data, accept risks to their privacy and confidentiality should the data be improperly shared or used. Rather than a broad framework that provides general privacy protections, the United States has developed a patchwork of subject-specific regulations to protect the privacy of different types of information.⁹²

This system of subject-specific regulations includes, for example, regulations that protect census data, financial information, medical records, and video rental records, but does not include regulations that protect personally identifiable information that is not financial or medical, including name, address, occupation, affiliations, or internet activity.⁹³ As a matter of respect for persons as well as justice and fairness, a government can institute laws and regulations that help mitigate risks to individuals who share whole genome sequence data, and it can protect individuals from unwillingly or unwittingly sharing their whole genome sequence data. But it cannot eliminate all privacy risks while still effectively encouraging scientific, economic, and social progress. Just as the Commission strongly supports effective protections of privacy, it also emphasizes that sharing whole genome sequence data for the sake of medical research holds great potential for public benefit. The principle of public beneficence strongly encourages this sharing in a setting that provides adequate protections of privacy.

U.S. Privacy Regulations

The collection and protection of personally identifiable information is not new. The United States has collected personally identifiable information through the census and the tax systems since its early history. The government has recognized the importance of keeping this information secure and has implemented protections to ensure the privacy and security of these data.⁹⁴ Privacy laws and regulations permit but regulate cross-agency matching of collected data, and establish precedent that personal data shared by an individual for one specific purpose

should not be used to other ends, such as law enforcement or judicial proceedings, without their consent. In addition, traditional identifying information often is removed from the data files.⁹⁵

The United States has made several sectoral legislative attempts to regulate the privacy and security of personal data. These laws include the Fair Credit Reporting Act; the Privacy Act of 1974; the Confidentiality of Alcohol and Drug Abuse Act; the Family Educational Rights and Privacy Act of 1974; the Electronic Communications Privacy Act of 1986; the Video Privacy Act of 1988; the Children's Online Privacy Protection Act of 1998; and the Gramm-Leach-Bliley Act of 1999, also known as the Financial Services Modernization Act (requiring financial institutions to protect consumer privacy).⁹⁶

The laws cited above generally comport with "fair information practice" principles and practices first set forth in the Department of Health, Education, and Welfare (the precursor of HHS and the Department of Education) report, "Records, Computers and the Rights of Citizens."⁹⁷ The practices include the following principles: 1) there must be no personal data record-keeping systems whose very existence is secret; 2) individuals must be able to find out what personal information about them is in a record and how it is used; 3) individuals must be able to prevent information obtained for one purpose from being used for other purposes without consent; 4) individuals must be able to correct or amend a record of identifiable information; and 5) any organization creating, maintaining, using, or disseminating records of identifiable personal data must assure the reliability of the data for their intended use and must take reasonable precautions to prevent misuse of the data.⁹⁸

HIPAA, enacted in 1996, is the federal law most relevant to medical privacy.⁹⁹ Pursuant to the authority of Title II, HIPAA sets forth policies, procedures, and guidelines for maintaining the privacy and security of personally identifiable health information.

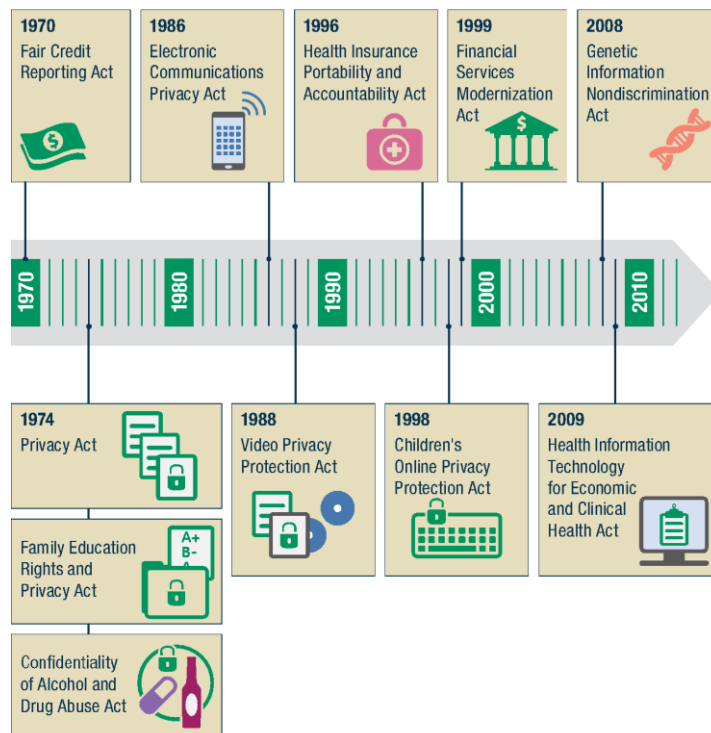


Figure 2. U.S. Federal Privacy Laws.

The HIPA A-mandated Privacy Rule was finalized in 2005. The Privacy Rule defines the circumstances in which an individual's protected health information—including any identifiable information—may be used or disclosed by a covered entity.¹⁰⁰ A covered entity is a health plan, a health care clearinghouse, or a health care provider that transmits any health information in electronic form.¹⁰¹ Under HIPA A, health information is not “identifiable” if there is “no reasonable basis to believe that the information can be used to identify an individual” or if it is stripped of the HIPAA identifiers.¹⁰² An individual's privacy rights under the Privacy Rule survive death.¹⁰³ While it is clear that genetic information is health information under HIPA A, HHS has stated that it is only covered by the Privacy Rule to the extent that it meets the definition of *protected* health information.¹⁰⁴ HHS has not clarified whether genetic or genomic information on its own is protected health information—that is, whether it falls under one of the HIPAA identifiers, such as “biometric identifier” or “any other unique identifying number, characteristic, or code.”¹⁰⁵

IDENTIFYING INFORMATION UNDER HIPAA

Names
 Address
 Dates
 Phone numbers
 Fax numbers
 Email addresses
 Social security numbers
 Medical record numbers
 Health plan beneficiary numbers
 Account numbers
 Certificate/license numbers
 Vehicle identifiers
 Device identifiers and serial numbers
 Web URLs
 Internet protocol (IP) addresses
 Biometric identifiers, including finger and voice prints
 Full face photographic images and any comparable images
 Any other unique identifying number, characteristic, or code (with certain exceptions)

A covered entity *must* disclose an individual's protected health information to him or her when specifically requested, and to HHS in the event of a compliance investigation or enforcement action.¹⁰⁶ A covered entity *may* disclose protected health information without consent in specifically enumerated circumstances, including for purposes related to treatment, payment, public health, and health care operations. A covered entity that discloses protected health information, however, must try to disclose only the minimum necessary to achieve its purpose.¹⁰⁷ There are no restrictions on the use or disclosure of de-identified health information, which is information that neither identifies nor provides a reasonable basis with which to identify an individual.¹⁰⁸

HITECH updated and revised HIPAA to extend slightly its privacy protections. HITECH adds business associates of covered entities to the list of those who can be subject to liability for disclosure of protected health information. It also strengthens the accounting requirements for the protection of health information, and imposes new notification requirements for covered entities to comply with when a breach has occurred.¹⁰⁹ The Office of the National Coordinator for Health Information Technology was created in 2004 through an Executive Order, and legislatively mandated in the HITECH Act. Its mission is to coordinate nationwide efforts to implement and use the most advanced health information technology and the electronic exchange of health information.¹¹⁰

While the requirements of HIPAA and HITECH apply only to “covered entities,” most academic institutions and federal agencies are required to follow the rules set forth for human research under the Common Rule. The Common Rule is a federal regulation governing human research in the United States that requires federally funded scientific research to be subjected to independent review by an institutional review board (IRB), have equitable subject selection, use procedures consistent with sound research design, minimize risks to participants, and obtain informed consent. Informed consent by participants must generally include, among other things, a description of the procedures in the research plan, an explanation of the risks and benefits to the participant, a description of the extent to which confidentiality of records will be maintained, and an explanation of the right to withdraw from the study.¹¹¹

Currently, whole genome sequence data obtained in the clinical context can be stripped of traditional identifiers and used for research purposes without IRB review or additional consent. This is because whole genome sequence data, when stripped of traditional identifiers (such as name or address), are not considered readily identifiable under the Common Rule.¹¹² The logic behind this is that while whole genome sequence data are unique to an individual, without a key that matches particular data to an identity, one could not readily ascertain *which* person the whole genome identifies. Similarly, while fingerprints are considered identifiable for law enforcement purposes, a fingerprint with no personal identifying information cannot point to whom that fingerprint belongs. In other words, a fingerprint does not have a name or address encoded directly in it. To discover the suspect’s identity, one must link the print to a database containing both traditional personal identifiers and fingerprints in order to know which person to arrest. Only research that uses data where the identity of the subject is, or may readily be, determined is considered human research under the Common Rule. Research using data stripped of traditional identifiers is not considered human research and therefore does not trigger Common Rule protections such as IRB review or consent.

Research using whole genome sequence data that have not been stripped of traditional identifiers (e.g., readily identifiable information) is considered human research. Accordingly, this research is governed by the Common Rule, meaning that IRB approval and informed consent must be obtained or waived by an IRB before the research can occur.

HHS recently published an Advanced Notice of Proposed Rulemaking (ANPRM), entitled *Human Subjects Research Protections: Enhancing Protections for Research Subjects and Reducing Burden, Delay, and Ambiguity for Investigators*, and collected comments on whether some types of genomic data should be considered identifiable. This ANPRM acknowledges that “there is an increasing belief that what constitutes ‘identifiable’ and ‘de-identified’ data is fluid” and that evolving technologies and the increasing accessibility of data could allow de-identified data to become re-identified.¹¹³ It also highlights the concern that “advances that have come in genetic and information technologies” might “make complete de-identification of

biospecimens impossible and re-identification of sensitive health data easier.”¹¹⁴ This is an ongoing discussion. A change to the Common Rule pertaining to identifiability could impact the collection and subsequent use of whole genome sequence data.

International Approaches to Regulating Genetic Information

The United States is not the only country deciding how best to prevent the misuse of genetic information. No international models yet exist regarding the misuse of and specific protections for whole genome sequence data. Some countries have enacted general privacy laws that encompass personal health information; patient rights’ acts that regulate, among other things, informed consent and confidentiality of medical information; and legislation that specifically regulates genetic information and genetic research. These laws differ from U.S. law, which is focused on prohibiting discrimination resulting from disclosure of genetic information rather than ensuring privacy of genetic information.

Many countries and foreign bodies have broad laws that regulate the use of personal information.¹¹⁵ Some of these, such as the European Union’s Data Protection Directive, offer special protection for more sensitive data, including personal health information.¹¹⁶ These privacy laws are far reaching—covering private and public institutions and many types of data—and are often overseen by data commissions or commissioners.¹¹⁷

In addition to these general data protection laws, many countries also have enacted patient rights’ laws that prohibit discrimination and require confidentiality of patients’ health information. These laws often require informed consent for disclosure of personal health information.¹¹⁸ Some of these laws, like those in the United States, also require that patients have access to their own medical records.¹¹⁹

In recent years, some countries have enacted laws specifically regulating genetic information and research. For example, Chile enacted a law in 2006 regulating genetic research that prohibits discrimination on the basis of genetic heritage and requires informed consent for research, confidentiality of genetic information, and anonymization of genetic data.¹²⁰ Some of these laws allow genetic testing only for individual health reasons or scientific research.¹²¹

Legal Protections of Genetic and Whole Genome Sequence Data

In light of mounting concerns about genetic privacy at the onset of the Human Genome Project, the U.S. Congress adopted legislation protecting against genetic discrimination. In 2008, Congress passed GINA, which aims to prevent genetic discrimination in the health insurance market (Title I) and in employment decisions such as hiring, firing, job assignments, and promotions (Title II).¹²² GINA does not protect against discrimination in the context of life insurance, disability insurance, or long-term care insurance. GINA’s protections apply to asymptomatic individuals, not those who have “manifested disease.”¹²³ Nor does it prescribe rules for genetic research.¹²⁴ GINA also expanded HIPA A privacy protections by applying prohibitions against genetic discrimination to all health insurers.¹²⁵

“There is certainly more room for legislation about privacy... the Genetic Information Nondiscrimination Act...is only a start. There are many more protections that the patient community would like that are not present in GINA.”

Greg Biggers, Council Member, Genetic Alliance; Chief Executive Officer, Genomera. (2012). Genomic Privacy, Data Access, and Health IT. Presentation to PCSBI, May 17, 2012. Retrieved from <http://bioethics.gov/cms/node/713>.

Under Title I of GINA, all health insurers are barred from: 1) using genetic information to determine coverage, eligibility, or premiums; 2) requesting or requiring genetic testing or genetic information for underwriting decisions; and 3) obtaining genetic information for underwriting purposes.¹²⁶ Additionally, insurers may not, on the basis of genetic information, impose a preexisting condition exclusion.¹²⁷ GINA extended HIPAA protections to cover persons purchasing individual, rather than group, health insurance policies.¹²⁸

GINA substantially expanded protections from genetic discrimination in employment. Under Title II of GINA, an employer with more than 15 employees cannot use an individual's genetic information when making employment decisions such as hiring, firing, job assignments, and promotions, nor can an employer request, require, or purchase genetic information about an individual employee or family member.

Although GINA prohibits specific types of misuse of genetic information by health insurers and employers, it does not address the use of or access to genetic data. In other words, GINA is an anti-discrimination law; it does not provide comprehensive privacy protections.

GINA provides a uniform federal law as a floor of protections against genetic discrimination, but also allows for state laws that provide additional safeguards.¹²⁹ Slightly fewer than half of all U.S. states have laws providing additional protection against discrimination in aspects of life, long-term care, or disability insurance not present in GINA.¹³⁰

About half of the U.S. states have policies governing genetic privacy. There is a great degree of variation, however, in what protections states afford their citizens regarding the collection and use of genetic data and, similar to the federal level, none have specific prohibitions for whole genome sequence data. Some states protect against the improper collection of genetic material without consent.¹³¹ Others protect against the improper disclosure of genetic information (and several of these states' laws do not specify to whom the disclosure is prohibited, whether to the donor or another party).¹³² Still others protect against improper retention of genetic information without consent.¹³³ The result of the variation in state laws is that there is no standard or comprehensive approach to the protection of genetic information in the United States, and the level of protection afforded to an individual's genetic information differs widely from state to state (for more information regarding the diversity of state law genetic protections, see Table 2 and Appendix IV: U.S. State Genetic Laws).

The U.S. Supreme Court has not established a constitutional right to informational privacy applicable to whole genome sequence data. Although the Supreme Court has addressed privacy rights of biomedical information in the context of the Fourth and Fourteenth Amendments, there is no case law addressing informational privacy in the context of whole genome sequencing.¹³⁴

Legal protections might be afforded if individuals have state property rights over their biospecimens, though courts have generally favored scientists over individuals from whom the specimen was taken.¹³⁵ The most famous case is *Moore v. Regents of the University of*

California in which the Supreme Court of California held that individuals are entitled to informed consent, but do not maintain property or privacy rights over cells after they have been removed from their body.¹³⁶

Table 2. Examples of State Genetic Privacy Laws

State	Protections
Arizona <i>AZ Rev. Stat. §12-2801-4, §20-448.02; 21</i>	Requires informed consent for genetic testing performed by health care providers, but does not address whether a non-health care provider may collect or analyze genetic material.
Oklahoma <i>Okl. St. § 1175</i>	Provides privacy protections for genetic information obtained from newborns, but does not provide similar protections for adult genetic data.
Hawaii <i>HRS §§ 431:10A-118</i>	Prohibits disclosure of genetic information by insurers, but does not specify the same for health care providers, nor does it protect against unwanted analysis of genetic material.
Missouri <i>§ 375.1309 R.S.Mo.</i>	Prohibits the disclosure of genetic information by persons who hold such information “in the course of business,” but does not address persons who have obtained it for any other reason.
Rhode Island <i>R.I. Gen. Laws §27-18-52, 52.3, §27-19-44, 44.1, §27-20-39. 39.1, §27-41-53, 53.1</i>	Prohibits the unauthorized disclosure of genetic information by insurance companies, but does not prohibit unauthorized disclosure by anyone else who may have access to genetic information.
Vermont <i>V.T. Stat. Ann. tit. 18, §9331 to 9335</i>	Prohibits people from performing genetic tests without consent and from disclosing the results of genetic tests without consent, but does not regulate the unauthorized obtaining or retention of genetic information.
Wyoming <i>Wyo. Stat. Ann. §14-2-701 to 710</i>	Prohibits the disclosure of genetic tests for paternity without consent, but does not address any other kinds of genetic tests.
Michigan <i>Mich. Comp. Laws §§ 333.17020, 333.17520</i>	Prohibits the performing of a genetic test by a health care provider without consent, but does not address performance of genetic tests by any other party, and does not prohibit unauthorized obtaining, retaining, or disclosure of genetic tests by any party.

State contract law also may provide legal protection if an individual has signed an informed consent document. In the context of genetic databases, researchers and participants can contractually determine who can access or use the data and on what terms, and the penalties for misusing protected information.

Conclusion

There is considerable concern in modern society about unauthorized or unintended disclosures of genetic information. While GINA prohibits genetic discrimination in the health insurance and employment contexts, it does not regulate use, access, security, or disclosure of genetic data, and does not specifically address whole genome sequence data or information. State-based privacy laws, consent forms, and IRBs collectively create a patchwork of privacy protections, but they neither comprehensively nor consistently protect the whole genome sequence data of individuals. In an era in which genomic data increasingly are stored and shared using biorepositories and genetic databases, there is little to no systematic oversight of these facilities.

To address the complex privacy and data security issues that arise in this arena, we need each of three robust facets of privacy and confidentiality protections of whole genome sequence data: individual researchers and clinicians; information technology systems; and laws, regulations, and institutional policies. Protection of personally identifiable data requires attention at all three broad facets of responsibilities. Individuals who collect, handle, store, and use data must recognize the ethical imperative of protecting the privacy of persons from whom they collect data. The information technology systems should be designed to protect persons by prohibiting the unauthorized access and release (intentional or unintentional) of identifiable data and protecting databases from intrusion. Laws and policies must protect persons from negative consequences of disclosure of information (e.g., discrimination) as well as enforce accountability and consequences for unauthorized access or disclosure.

It is clear that laws and regulations cannot do all of the work necessary to provide sufficient privacy protections for whole genome sequence data. Together with laws and regulations preventing misuse of data, individuals who receive such data have professional ethical obligations to protect the data that go beyond the limitations of the legal protections. Moreover, given how rapidly whole genome sequence technology is changing, it is in some ways preferable to adopt professional guidelines and policies rather than enact additional laws, since professional guidelines and policies are updated far more easily.¹³⁷ Guidelines and policies also might help those affected consider more deeply the privacy considerations at issue rather than focusing entirely on compliance with the letter of the law.¹³⁸

SECTION 3. ANALYSIS AND RECOMMENDATIONS

For this report on whole genome sequencing, the Commission has been mindful of the five ethical principles set out in its first report, as described in detail in Section 1. These principles for assessing emerging biotechnologies are public beneficence, responsible stewardship, intellectual freedom and responsibility, democratic deliberation, and justice and fairness. The Commission's principles complement and build upon the Belmont principles of respect for persons, beneficence, and justice. The Commission drew on both sets of principles to develop recommendations that facilitate responsible development of, access to, and use of whole genome sequencing.

The Commission focused on the principle of respect for persons by seeking to minimize risks to individuals willing to share their whole genome sequence data. Although individual benefits of whole genome sequencing are emerging, they are more elusive than predicted a decade ago. Many of the benefits anticipated from advances in whole genome sequence research will accrue to society generally through, for example, improved diagnosis and public health resulting from efficient medical treatment. Related privacy risks, however, primarily fall to individuals willing to share their genomic information. Risks might also fall to blood relatives of these individuals who carry similar genomic variants, thereby raising the stakes of privacy concerns in whole genome sequencing compared with most other types of research.

Strong privacy protections enable individuals to determine autonomously their preferred level of data and information sharing. When individuals have control and can govern sharing of their data at a level with which they are comfortable, they are more likely to have trust in the research or clinical enterprise, and are more likely to participate and share data, benefiting

society generally. These privacy interests are served by robust informed consent, data security provisions, and systematic oversight.

With the above in mind, the Commission identified the following areas for ethical analysis:

- Standards that allow individuals, if they wish, to access and share their whole genome sequence data and information;
- Security of whole genome sequence data and information and standards of access to and use of whole genome sequence databases;
- Informed consent to whole genome sequencing in the contexts of clinical care and research;
- Oversight of collection, storage, access, and use of whole genome sequence data and information; and
- Distribution of benefits from medical advances resulting from whole genome sequencing.

The recommendations presented in this section apply to individuals and entities that have an interest in, and work with, whole genome sequence data and information, both in the public and private sectors. Whole genome sequence data collected in the clinical setting are indistinguishable from whole genome sequence data collected in the course of research, and data increasingly move back and forth between the clinical and research settings. Ethical principles providing guidance in this area are based on a shared morality. While the implementation of recommendations that follow might be different depending on the entity involved in the collection, storage, access, and use of whole genome sequence data, the ethical issues at stake are the same.

The Commission's recommendations are also based on the fact that whole genome sequence data are inherently unique, meaning there is only one person in the world with that specific sequence. If the identity of the donor is not apparent to the user of the data, that individual is not readily identifiable. However, whole genome sequence data are often most useful when linked with information about physical characteristics, environmental factors, and medical records. These additional pieces of information, in turn, might make whole genome sequence data readily identifiable.

The Commission sees promise in the application of information technology to the field of whole genome sequencing. Information technology is able to tailor access to data with a degree of specificity not possible with traditional medical records, potentially making all types of whole genome sequence data more secure.

Uses of whole genome sequence data are rapidly evolving, and some of these uses do not fit easily into the current regulatory framework. The Commission, therefore, has crafted its recommendations to call attention to areas where it believes that current laws, regulations, and policies need to be reconsidered to honor applicable ethical principles and ensure that whole genome sequencing is most effectively used to the benefit of society and its individuals.

Strong Baseline Protections While Promoting Data Access and Sharing

Whole genome sequencing increasingly is being incorporated into clinical care and research. Presently, numerous national and state policies are in place to guard personally

identifiable health information and records of participation in research.¹³⁹ These policies should apply to all handlers of the data, from those who collect the data, to researchers, to third-party storage and analysis providers.¹⁴⁰ Privacy protection is an essential component of oversight of the use of whole genome sequencing in research and clinical care. Privacy protections should guard against unauthorized access to, and illegitimate uses of, data and information while allowing for authorized users of these data to advance public health.

For both ethical and practical purposes, it is important to carefully distinguish between access to, use of, and possession of whole genome sequence data. *Access* means being able to come in contact with the information, whether physically or electronically. It would be impossible to limit physical access to all sources of whole genome sequence data. We leave behind specimens containing our DNA in myriad public places—by discarding a coffee cup, for example—that could be used to perform whole genome sequencing. (It is more feasible of course to protect electronic access to whole genome sequence data in biorepositories and databases.) While individuals might have abandoned these genomic samples to public access, they nonetheless have a strong interest in whether the data they contain are collected and how they are used. On the other hand, sometimes persons might have authorized access to whole genome sequence data but misuse the information (e.g., by sharing information with a reporter). In certain cases, others simply have no right to know certain things about other people, no matter what they do with the information.¹⁴¹

Unauthorized access to data is not necessarily a problem in and of itself— despite having access to information, one can choose to not use it, and thereby not produce any harm. Misuse of information can therefore be more ethically significant than unauthorized access.

Laws and regulations can prohibit unauthorized parties from accessing or misusing whole genome sequence data, but it is impossible to guarantee that this will not occur. Laws and regulations can, however, provide deterrents to inappropriate access or misuse (such as fines), and compensation for the individuals whose data have been inappropriately accessed or misused.

Presentations to the Commission also indicated that authorized individuals can use data without having actual *possession* of those data.¹⁴² Technologies are being developed that allow “computational” access to data sets, which allow access and use without the user possessing the data set. In computational access, the data are possessed by a central party, but others can remotely perform analyses (i.e., use) of the data.

AN EXAMPLE OF COMPUTATIONAL ACCESS

Google, a major internet search engine, has collected data from its customers’ internet activity. Google views these data as a commercial asset and does not share possession of them. However, Google tools, such as Google Correlate and Google Trends, allow users to query Google’s collected data. A user can search for “stapler” to ascertain whether stapler and staple sales correlate, but users receive only the answer to their question (not access to the data mined by Google yielding the result). By using computational access, Google can give users access to answers, but not access to the data.

Source: Google. (n.d.). Google Trends. Retrieved from <http://www.google.com/trends/>.

Developments in the science of whole genome sequencing, which are progressing quickly, will require ongoing ethical consideration and democratic deliberation. Individuals and groups have differing sensibilities toward the privacy and publicity of whole genome sequence data, which might be relevant to distinguishing between acceptable and unacceptable uses of data. Perceived misuses of whole genome sequence data vary between cultures and individuals. For example, some individuals might be open to having a secondary researcher use his or her whole genome sequence data for an ancestry study. Members of the Havasupai tribe, on the other hand, strongly disapproved of their samples being used in ancestry studies, because these studies contradicted their traditional origin beliefs.¹⁴³ Some parents do not object to using Guthrie card newborn blood screening spots in future research without consent. Notable lawsuits in Minnesota and Texas, however, have indicated that some parents feel otherwise.¹⁴⁴ Requiring consent for future uses of readily identifiable whole genome sequence data, and encouraging consent for future uses of any data, are important to appropriate use. However, it is difficult to consent to all specific future uses in rapidly advancing scientific technology.

While privacy and confidentiality remain imperatives for protecting whole genome sequence data, it is also important to recognize that the American public, generally speaking, has become more open about communicating their health information. The development of online resources and communities reflects a shift in societal notions of what data should remain private (regardless of whether individuals would want to make it public). People now freely share information that was once considered inherently private or not suitable to be shared with a broad audience. The arrival of whole genome sequencing in health care has coincided with an era of greater openness about diseases that used to carry social stigma, such as HIV/AIDS, cancer, and mental health conditions. Many patients may choose to publicly share their stories, although others might not for privacy reasons.

Social attitudes about privacy are changing. There have been shifts not only in what information is considered private but also in how entities can realistically be expected to protect that privacy. Technological advances can trigger the creation of new privacy policies, such as the National Institutes of Health's (NIH) updated genome-wide association study policies.¹⁴⁵ While policy makers continue to focus on genetic non-discrimination policies that protect those whose privacy has been compromised, they have also begun to focus on data security policies that protect the data in the first place.¹⁴⁶ Finally, informed consent practices increasingly acknowledge that absolute privacy cannot be guaranteed.¹⁴⁷ Policies likely will evolve as notions of privacy continue to change. Future policies need to be flexible so that they can adapt to such advances in data security and information technology.

Recommendation 1.1.

Funders of whole genome sequencing research; managers of research, clinical, and commercial databases; and policy makers should maintain or establish clear policies defining acceptable access to and permissible uses of whole genome sequence data. These policies should promote opportunities for models of data sharing by individuals who want to share their whole genome sequence data with clinicians, researchers, or others.

Strong baseline privacy protections require a spectrum of policies starting with data handling through the protection of persons from future disadvantage and discrimination arising from misuse of their whole genome sequence data. It is critical, however, to ensure that privacy regulations allow individuals to share their own whole genome sequence data with clinicians, researchers, and others in ways that they choose.

Policy makers should also revisit efforts to strengthen protections against, and sanctions for, discrimination by treating the Genetic Information Nondiscrimination Act as a floor, not a ceiling, of protection. For example, because GINA does not cover symptomatic persons or address discrimination in life, disability, or long-term care insurance, persons with genetic diseases and predispositions are vulnerable to discrimination.¹⁴⁸

Advances in information technology should be pursued that promote appropriate use of whole genome sequence data while safe-guarding access to data files. For example, computational models that limit access to data files without preventing researchers' ability to analyze these data can be a valuable tool to protect privacy.

DATA PROTECTIONS THAT MOVE WITH THE DATA

The President's Council of Advisors on Science and Technology advocates a "tagged data element approach [that] allows for a sophisticated, fine-grained model of implementing strong privacy controls (including honoring patient-controlled privacy preferences where applicable) and strong security protection." This approach encourages privacy protections to move with the data across institutions, as opposed to changing protections based on the handler.

Source: President's Council of Advisors on Science and Technology. (2010, December). *Report to the President Realizing the Full Potential of Health Information Technology to Improve Health care for Americans: The Path Forward*, p. 52. Retrieved from <http://www.whitehouse.gov/sites/default/files/microsites/ostp/pcast-health-itreport.pdf>.

Last, policies regarding access to and use of data should take into consideration varying cultural, ethnic, and racial views about what might or might not constitute a misuse of data.¹⁴⁹

Currently, about half of the U.S. states have laws or regulations governing genetic privacy that outline illegitimate uses of these data. However, there is tremendous variation in these laws. In some instances, it is difficult to determine whether a state prohibits surreptitious testing of genetic material from an unwilling donor because of unclear language in the statutes. Some states prohibit unauthorized acquisition or analysis of genetic information, while others prohibit only unauthorized disclosure (and it is often unclear to whom disclosure is prohibited). Some laws at the state level encompass all genetic information, while others address only health-related information, or information obtained or used in particular settings (e.g., employment or insurance discrimination).¹⁵⁰ Therefore, whether genetic testing or whole genome sequencing without the consent of the donor is prohibited can depend on a combination of factors: who conducts the test, to whom the DNA belongs, what the test attempts to determine, how the results will be used, and in what state the testing takes place.¹⁵¹ Moreover, no states have laws or regulations specific to whole genome sequence data; some states have laws that include the words "DNA" and "genetic," although it is unclear whether these laws might be interpreted to cover whole genome sequence data and information.

Some of the topics specified in existing genetic laws could be used for whole genome sequencing laws as well. Types of regulations that would translate effectively into genomic protections include those regarding:

- Defining restrictions on what information can be stored in a biorepository, biobank, or genomic research database;

- Sharing of whole genome sequencing data, and if clinical data are shared with researchers, what type of information can be shared (e.g., stripped of traditional identifiers or not), and penalties for violations; and
- Using whole genome sequence information for life, disability, or long-term care insurance.

When individuals are asked about their concerns with respect to online health information, most focus on illegitimate uses of the data. They also cite discrimination, such as unauthorized use by insurers or employers, or use of their data for marketing purposes.¹⁵² However, the existing patchwork of state protections—with some states having no laws and the others having an inconsistent potpourri of legal prohibitions—does not protect all individuals from unauthorized uses. These uneven protections might also affect the development of trust in contexts where individuals are asked to share their whole genome sequence data for the public benefit in the course of research, clinical care, or commerce. Like all medical information, whole genome sequencing data should be ensured baseline privacy protections in all jurisdictions.

Recommendation 1.2.

The Commission urges federal and state governments to ensure a consistent floor of privacy protections covering whole genome sequence data regardless of how they were obtained. These policies should protect individual privacy by prohibiting unauthorized whole genome sequencing without the consent of the person from whom the sample came.

Currently federal and state laws protect data dependent on who collected them (i.e., a clinician, researcher, or consumer). Although a whole genome sequenced in the clinic is the same as a whole genome sequenced during research, data collected in the course of clinical care are governed by the Health Insurance Portability and Accountability Act, while data collected in the course of research are governed by the federal Common Rule for human research. The exact same data are treated differently depending on who collected the sample. Clinical data are collected to benefit the patient, while research data are collected to advance science and health care generally. However, the blurring of clinical and research lines, particularly in the field of whole genome sequencing, compels reconsideration of the differences between how clinical and research data are protected.

In addition, while the requirement for consent to whole genome sequencing is regulated in the clinical and research contexts (depending, to some extent, on whether or not traditional identifiers—such as name, address, or social security number—are attached to the sample), commercial genetic testing has opened a new loophole in privacy protections. One can now pick up a discarded coffee cup and send a saliva sample to a genetic testing company.¹⁵³ The potential consequences of unauthorized surreptitious testing could be profound (e.g., revealing disease risks to sway the disposition of a custody battle).¹⁵⁴ There are, of course, exceptions to this need for consent, such as use for legitimate law enforcement purposes. The Commission therefore recommends the prohibition of “unauthorized” whole genome sequencing—a term intended to carve out an exception for legitimate law enforcement.

Data protections should be tied to the *nature* of the data, not who collects them. Widely shared norms of justice and fairness dictate that similar kinds of data should be treated in similar ways, no matter in which state or health facility they are sequenced. If protections are inherent to the data, they should follow the data and dictate appropriate use. For example, meta-data

tags could be used to encode the level of security protections required for the data file and elements of consent (e.g., these data can/cannot be used in reproductive research). Using this approach, data will receive appropriately consistent use protections throughout their life span. More consistency in state protections of genetic and genomic data could also enhance privacy.

Treating like data alike is crucial to ensuring consistent protections for whole genome sequence information across the United States. Although states should enact genomic policies that are most relevant and important to their constituents, bringing such protections to a minimum standard that addresses privacy—while still allowing individuals to share their own data—would provide just and fair protections regardless of where one happens to reside.

Because the options for implementation of such protections are unclear, the Commission recommends that experts in federal law, state law, policy, and privacy be brought together to engage in further democratic deliberation regarding acceptable access to and permissible use of whole genome sequence data. The Commission will consider following up with stakeholders regarding: 1) suggested requirements to ensure a floor of protection of whole genome sequence data and data sharing in all states; and 2) the practical steps necessary to accomplish this goal, such as federal, state, or non-regulatory interventions.

Data Security and Access to Databases

Respect for persons requires honoring data privacy. Data privacy requires data security. Data security requires ethical responsibility and accountability from all those who handle whole genome sequence data and information. It must further be supported by policies and infrastructure to protect safe sharing of data.

Authorized users must have access to whole genome sequence databases to conduct research and make advances that will contribute to improved medical diagnostics and treatment for all. Security should allow only authorized individuals to access these data. However, breaches of unsecured protected health information have been publicized in the past, and can cause patients and research participants to doubt the security of their data. Unsecured health information can be accessed by unauthorized persons through means such as the loss or theft of unencrypted information on data storage devices, hacking of network servers, unauthorized disclosure, or improper disposal of paper records.¹⁵⁵ In a recent case, the unencrypted health information of over 800 patients was inadvertently embedded in PowerPoint presentations that were posted online.¹⁵⁶ In light of the possibility of data security breaches, it is important to address misuse of whole genome sequence data rather than wholly relying on preventing unauthorized access to these data.

“Technology can help save privacy, it can change your thinking, whether [it is] with respect to setting norms, [or] whether [it is] with respect to changing the way you set up the platform so that the platform can do more [computational] analysis... versus sharing the data around.”

Latanya Sweeney, Visiting Professor and Scholar, Computer Science Director, Data Privacy Lab, Harvard University. (2012). How Technology is Changing Views of Privacy. Presentation to PCSBI, August 1, 2012. Retrieved from <http://bioethics.gov/cms/node/748>.

When told of data hacking, some assume that the transition of private information to an electronic format makes it less safe. Quite the opposite might be true. In many respects, advances in information technology can be used to strengthen data security. For example, electronic files bear marks of who accessed them and when, allowing for more fine-tuned file tracking than is possible with paper records that may be surreptitiously accessed without a trace. In addition, current technology allows data files to be analyzed without the need to export the data files to other networks, that is, computational access can be allowed without data transfer. Even when individuals are willing to share their readily identifiable data and information for use in research, they might not want copies of their information saved on computers around the world. Access to and sharing of data files do not have to be one and the same. The Commission supports ongoing exploration and development of a set of best practice models that separate possession of, access to, and use of data.¹⁵⁷

Recommendation 2.1.

Funders of whole genome sequencing research; managers of research, clinical, and commercial databases; and policy makers should ensure the security of whole genome sequence data. All persons who work with whole genome sequence data, whether in clinical or research settings, public or private, must be: 1) guided by professional ethical standards related to the privacy and confidentiality of whole genome sequence data and not intentionally, recklessly, or negligently access or misuse these data; and 2) held accountable to state and federal laws and regulations that require specific remedial or penal measures in the case of lapses in whole genome sequence data security, such as breaches due to the loss of portable data storage devices or hacking.

Absolute privacy, many observe, is not possible in this as in many other realms. The greater potential for harm is not by virtue of authorized others *knowing* about one's whole genome make-up, but rather through the misuse of data that have been legally accessed.¹⁵⁸ For example, a clinician with a celebrity client would have legally authorized access to their client's whole genome sequence data for purposes of providing clinical care, but could not then sell that information to a tabloid. Researchers, clinicians, and others authorized to access whole genome sequence data should be guided by professional ethical standards so that they do not intentionally or inadvertently misuse these data.

In the event that data are mishandled or lost, those responsible should be aware of federal and state policies that require specific remedial actions, such as the requirement under the Health Information Technology for Economic and Clinical Health Act Breach Notification Rule to report breaches to the Department of Health and Human Services within the required number of days.¹⁵⁹ Those persons authorized to access whole genome sequence data should take part in regular training sessions to remain current on regulations governing whole genome sequence data privacy and security.

Public and private entities have different policies governing access to whole genome sequence databases by those seeking to use data for purposes other than that for which they were originally collected. Some policies create absolute prohibitions on releasing data to outside parties and associated penalties for violation, and some are more flexible, relying on the discretion of the person who holds the data.¹⁶⁰ Certificates of Confidentiality, for example, permit but do not require investigators to refuse access to research data by law enforcement officials and others.¹⁶¹ The use of Certificates of Confidentiality however, is limited; one study found that only 114 (0.04 percent) of 27,000 funded studies secured such a certificate.¹⁶²

Although empirical data on the use and effectiveness of these forms of privacy protection are not robust, scholars have questioned the strength of these protections, how well understood these protections are, and how they affect research participation.¹⁶³

Besides researchers, parties who might be interested in accessing information already compiled in whole genome sequence databases and biorepositories include law enforcement officials and marketing agencies. While commercial advertising can be a valuable tool in educating at-risk populations, this technique is often viewed as invasive when used as a way to sell products, for example, to selectively market a statin to someone with a genomic predisposition to high cholesterol.¹⁶⁴ In order to establish and maintain trust between members of the general public, clinicians, and the scientific research community, strong whole genome sequence data protections must be in place to secure data. Further, these limits on access must be communicated to those giving consent to have their whole genome sequenced in clinical, research, or consumer-initiated settings.

Obtaining a whole genome sequence data file by itself yields information about, but does not definitively identify, a specific individual. The individual still has “practical obscurity,” as his or her identity is not readily ascertainable from the data. Practical obscurity means that simply because information is accessible, does not mean it is easily available or interpretable, and that those who want to find specific information must expend a lot of effort to do so. While some experts might be able to determine an individual’s hair color or specific cancer risk from whole genome sequence data (a file of 6 billion As, Cs, Gs, and Ts), these data are not interpretable by the vast majority of individuals. In addition, even if we know that a whole genome sequence is from one individual, we cannot know which of the over 7 billion people on Earth that person is without a key linking the whole genome sequence information with a single person or their close relative. Therefore, while whole genome sequence data are uniquely identifiable, they are not currently readily identifiable.

Traditional identifiers have been stripped from samples or data in the clinical and research setting to mitigate the possibility of risks to the individual from whom the samples came. Removing traditional identifiers from samples and data can allow for research on samples previously collected for different purposes, deter users from illegitimately identifying individuals, and minimize the risk that users might recognize individuals and use this information subconsciously in their daily life.

Recommendation 2.2.

Funders of whole genome sequencing research; managers of research, clinical, and commercial databases; and policy makers must outline to donors or suppliers of specimens acceptable access to and permissible use of identifiable whole genome sequence data. Accessible whole genome sequence data should be stripped of traditional identifiers whenever possible to inhibit recognition or re-identification. Only in exceptional circumstances should entities such as law enforcement or defense and security have access to biospecimens or whole genome sequence data for non health-related purposes without consent.

The consent process should communicate limits on access and use to those having their whole genome sequenced in clinical care, research, and consumer-initiated contexts. These policies should apply to the original recipient of the data, as well as to all parties who work with the data, from those who collect the sample or data to third-party storage and analysis service providers.

An existing policy that could serve as a model is the Agency for Healthcare Research and Quality's confidentiality statute.¹⁶⁵ This statute was put in place to foster participation in research and provides a respected form of statutory protection for all identifiable data submitted to the Agency for Healthcare Research and Quality for research. The statute covers AHRQ, its grantees, and contractors. The statute also defines strict penalties for individuals who use these data for non-consented purposes.

Whole genome sequencing and related analyses generate enormous data sets. As of March 2012, the 1000 Genomes Project contained the sequence data of 1,700 people. The project database contained 200 terabytes of data, or the equivalent of 30,000 standard DVDs. This data set is a tremendous resource for biomedical researchers. At the same time, these data might not be useful to medical scientists and researchers without the computing power required to work with such a large data set. Exploring options for making these data available to qualified researchers is critical so that innovation and research are not slowed simply because researchers' computer networks cannot store these large data files.

"The explosion of biomedical data has already significantly advanced our understanding of health and disease. Now we want to find new and better ways to make the most of these data to speed discovery, innovation, and improvements in the nation's health and economy."

NIH Director Francis S. Collins, M.D., Ph.D., in a press release announcing the movement of the 1000 Genomes Project data set to the Amazon Web Services cloud. Retrieved from <http://www.nih.gov/news/health/mar2012/nhgri-29.htm>.

The question of how best to handle large data sets has gained attention throughout the government. The federal Office of Science and Technology Policy recently announced a "Big Data Research and Development Initiative," with the goal of "improving our ability to extract knowledge and insights from large and complex collections of digital data."¹⁶⁶ Six federal departments and agencies are part of the initiative. This initiative includes NIH, which recently made its 1000 Genomes Project public data set available on the Amazon Web Services cloud. NIH now expects that researchers can access and analyze the data at a fraction of the cost it would take to establish the computing capacity at their own institution.¹⁶⁷

Making whole genome sequence data accessible to researchers and clinicians is a promising step toward advancing medicine for the betterment of society. Moving data to third-party storage and analysis service providers, however, complicates the protection of individual data. When data are moved to third parties, an expanded range of data handlers and administrators have access to the data. Currently, a wide range of federal regulations govern the conduct of entities that handle protected health information.¹⁶⁸

Recommendation 2.3.

Relevant federal agencies should continue to invest in initiatives to ensure that third-party entrustment of whole genome sequence data, particularly when these data are interpreted to generate health-related information, complies with relevant regulatory schemes such as the Health Insurance Portability and Accountability Act and other data privacy and security requirements. Best practices for keeping data secure should be shared across the industry to create a solid foundation of knowledge upon which to maximize public trust.

Whole genome sequence data not stripped of traditional identifiers are considered “protected health information” and are covered under the HIPAA Privacy, Security, and Enforcement Rules and the Common Rule. The same regulations, policies, and ethical guidelines that protect such health information should also be in place to govern the sharing of whole genome sequence data with third-party storage and analysis service providers (those otherwise not considered covered health entities under HIPA A). Entities within the public and private sectors have developed a range of practices for protecting privacy. For example, the National Institute of Standards and Technology, the Office of the National Coordinator for Health Information Technology, and the Office for Human Research Protections are developing policies concerning access to and use of data by third parties. The National Institute of Standards and Technology recently released guidance on “Security and Privacy in Public Cloud Computing” and the Office of the National Coordinator for Health Information Technology worked to strengthen protections of identifiable health information handled by third parties.¹⁶⁹ Also, the Office for Human Research Protections issued guidance on research with coded private information or biological specimens.¹⁷⁰ Parties from the public and the private sectors should share their lessons learned to promote efficiency and avoid duplicating efforts. Because of the expansive potential of information technology, special attention should be paid to those practices that leverage information technology to protect privacy.

In order for the public to benefit as much as possible, best practices across the industry should be shared to ensure the privacy and security of whole genome sequence data and best gain the trust of those who have their whole genome sequenced in research, clinical, and consumer-initiated contexts. These best practices should include encrypting stored data and storing data without traditional identifiers when possible. Even when data are being accessed and used with informed consent, persons who access the data should be responsible and accountable for protecting the privacy of individuals and the confidentiality of the data. Respect for persons requires that these and other privacy protections do not become a competitive advantage for certain parties but rather serve, in both appearance and reality, as a reliable standard of individual protection.

Consent

Although not unique to whole genome sequencing, a well-developed, understandable, informed consent process is essential to ethical clinical care and research. Conveying the complexities of whole genome sequencing to an individual, however, is likely more difficult than for the average diagnostic test. To make the issue more complex still, informed consent documents are often overly legalistic and written at a reading level beyond the capacity of the average research participant.¹⁷¹ Studies have demonstrated varying levels of comprehension of consent documents, including reports of persons signing consent forms who are later either unable to recall whether they signed a consent form or describe to what they had consented.¹⁷²

To educate participants thoroughly about the potential risks associated with whole genome sequencing, the consent process must include in format ion about what whole genome sequencing is; how data will be analyzed, stored, and shared; the types of results the patient or participant can expect to receive, if relevant; and the likelihood that implications of some of these results might currently be unknown, but could be discovered in the future. As per usual consent protocol, permission to perform whole genome sequencing for a person who cannot

consent for him or herself should be obtained from an informed, legally authorized representative.

“How does consent change when a person lacks genetic health literacy, [or] when the health condition does not yet exist, but is a future probability, and some of those may be non-treatable conditions? When a health condition does not have implications for you, but it does for your offspring, what are the terms of consent there, especially if your offspring have different views about what they want to know about genetics, and then lastly, for these incidental findings versus disease specific testing..? I’ll just leave you with those questions, as the first of many that you will engage.”

Daniel Masys, Affiliate Professor, Biomedical and Health Informatics, University of Washington School of Medicine. (2012). Ethics and Practice of Whole Genome Sequencing in the Clinic. Presentation to PCSBI, February 2, 2012. Retrieved from <http://bioethics.gov/cms/node/658>.

Consent documents differ between research and clinical care. Research informed consent documents are often long and contain n elements such as a summary of the research, future uses of data, the option to opt out, potential risks of participation, conditions of compensation in case of injury, and potential benefits to the individual. Clinical consent documents contain some of the same elements but generally are shorter than, and not as detailed as, research consent forms. In fact, oral consent might be sufficient for low-risk clinical procedures. The reason clinical consent is less comprehensive is because clinical procedures are done for the direct benefit of the patient and thus pose less of a risk of conflicting interests. More substantive clinical written consent is required, however, for higher-risk procedures, such as those expected to produce pain, require anesthesia, or have a significant risk of complications.

In the research world, public opinion polls have found that individuals believe that being asked for consent throughout the course of research with their specimens or data would make them feel “respected and involved.”¹⁷³ Informed consent involves an autonomous decision to participate in research that results from a communication process between researchers and prospective research participants that describes the research and explains the risks and benefits associated with enrolling in the study. Respect for persons dictates that individual consent should be well-informed and honored, regardless of a person’s specific privacy preferences.

Clinical written consent documents for whole genome sequencing need not be as detailed as research consent documents, but these documents should still adequately explain whole genome sequencing and its potential impact upon privacy interests. A clinician should not frame whole genome sequencing as “just another type of blood test.” Consent procedures for clinical whole genome sequencing should build on those consent procedures already in place for discrete genetic tests. In the clinical context, as in research, individuals being asked to consent to whole genome sequencing should understand the volume of data and information to be generated, as well as the risks, benefits, and implications of the results of whole genome sequencing.

The Common Rule states that data and specimens collected in the clinic, when stripped of traditional identifiers, can be used in research without consent. Because consent requirements differ in clinical and research settings, researchers could theoretically seek out data and specimens collected in the clinic to bypass the more involved research consent requirements.

While it is acceptable to use clinical data and specimens in research, the Commission does not condone researchers circumventing Institutional Review Board approval by seeking out clinical data and specimens for use in research when they could not otherwise obtain IRB approval.

Whole genome sequencing involving minors raises additional ethical quandaries even when permission is properly obtained from an informed, legally authorized representative. First, federal privacy laws inconsistently define the age of consent—for the most part, the age of consent is 18 years old in the United States, yet in health care for certain contexts (e.g., mental health, contraception, or substance use), state laws allow consent by minors as young as age 14.¹⁷⁴ Second, the potential future risks raised by the current unknowns of whole genome sequencing are compounded in children who will see advancement in the science during their lifetime. While the function of all genes is not currently known, researchers will continue to determine the function of more genes, and could feel compelled to re-contact these children, as adults, with results that they are not prepared to receive or do not want. Third, whole genome sequence data obtained from a minor already could have been widely shared before the minor reached an age at which they could determine preferred data sharing limits themselves, thereby decreasing their autonomy. Whole genome sequencing in children, therefore, raises a number of unique issues with regard to fully informed decision making.¹⁷⁵

Some commentators are concerned that participants enrolled in research that requires especially large data sets, and who are given too much control over their data, will stifle the production of public benefits, such as improvements in clinical care, comparative effectiveness research, and epidemiological studies.¹⁷⁶ If individuals can choose not to participate in certain types of studies, the amount of data available to clinicians and researchers upon which to base their conclusions will be limited to some extent.

A range of consent frameworks are available that offer participants varying levels of control over their data. Most of these frameworks fall into four categories: 1) broad; 2) narrow; 3) tiered; and 4) participant-centric or dynamic approaches. Under broad consent, individuals are given the option to opt in or opt out of general, and often yet to be determined, future uses of their data. Narrow consent usually states that data will be used only by the research team carrying out a specific study or for a specific treatment in the clinic. Tiered consent processes allow individuals to specify acceptable and unacceptable uses of their sample and data at the outset of research.

“We also, as a research community, need to get used to the fact that there are patient-driven research objectives now and [patients] are coming together to do [research].”

Laura Lyman Rodriguez, Director, Office of Policy, Communications, and Education, National Human Genome Research Institute. (2012). Protection of Private and Public Genomic Databases. Presentation to PCSBI, August 1, 2012. Retrieved from <http://bioethics.gov/cms/node/749>.

Other consent models use computer-based participant-centered consent processes, which generally give participants freedom to determine their specific data sharing preferences up front, with some allowing participants to monitor and modify their preferences on an ongoing basis through a computer interface.¹⁷⁷ One prototype that has been implemented by a group called Consent to Research allows users to “attach” consent to the data they donate, and any researcher who can accommodate the provisions of that consent can use those data.¹⁷⁸ Alternatively, as databases become more technologically flexible, those donating biospecimens

can express preferences at the outset about permissible and impermissible uses that can be respected by future users of whole genome sequence data. Further, sample donors could electronically update their consent to encompass proposed new studies, with minimal hassle to the donor or the researcher. These models, however, can only be used by participants who have computer and internet access.

Some data have been collected on participant views of consent forms for biorepository research. Biorepository specimens and data files can be collected in clinical or research settings, and include (among other things) medical waste, newborn blood spot cards, and biopsy specimens. In one pair of studies, when asked about an opt out consent process, over 90 percent of participants agreed or strongly agreed that “DNA biobank research is fine as long as people can choose not to have their DNA included.”¹⁷⁹ Another study found that, despite privacy concerns, 60 percent of individuals surveyed would participate in a genetic biorepository, 48 percent of whom would prefer broad consent, while 42 percent would prefer project-specific consent with re-consent for each project.¹⁸⁰ These studies indicate that the majority of individuals enrolled in research are willing to share their data when asked, and the limited data available suggest that individuals vary widely across this spectrum of preferred form of consent.¹⁸¹ More research is needed, however, including on minority and marginalized populations where research participation is not as high.

The Common Rule, which governs most human research in the United States, requires that research consent be informed. Consent may be waived in some circumstances, and research with samples or data that are not readily identifiable is not considered human research (and thus does not fall under the Common Rule). Blanket authorization for all future uses of identifiable data, known and unknown, at the outset of a research study cannot legally satisfy the current requirements for informed research consent. However, the Common Rule Advanced Notice of Proposed Rulemaking (ANPRM) proposes a broad consent requirement that would give participants the opportunity to say “yes” or “no” to all future research uses of their data and specimens at the outset of research.¹⁸² The ANPRM also proposes that individuals could designate special categories of research in which they would not want their samples included, for example, reproductive research. By giving individuals the option to not participate in research to which they object, these individuals are respected as persons. Moreover, the option to not participate in a set of specific categories of research that one finds objectionable might actually encourage broader participation in research.

Broad consent at the outset of research might be a more practical solution than re-consent, or obtaining informed consent from every donor for a new use. Re-consent is difficult or, in some cases, impossible, as individuals frequently change residences, clinicians, phone numbers, and email addresses. Researchers also maintain that obtaining consent for each future study is burdensome and could hinder research.¹⁸³

Recommendation 3.1.

Researchers and clinicians should evaluate and adopt robust and workable consent processes that allow research participants, patients, and others to understand who has access to their whole genome sequences and other data generated in the course of research, clinical, or commercial sequencing, and to know how these data might be used in the future. Consent processes should ascertain participant or patient preferences at the time the samples are obtained.

Respect for persons requires obtaining fully informed consent at the outset of treatment or research. The informed consent process should cover the current proposed use of individuals' data, convey who might have access to their data, and explain potential future uses of these data, as well as what research results and incidental findings, if any, will be returned to the patients or participants.

Some patients might be surprised to discover that their whole genome sequence data obtained in the clinic could be used for research in the future without additional consent. With the blurring of the line between clinical care and research, data may be shared back and forth to improve clinical diagnosis and treatment.¹⁸⁴ Patients in the clinic should thus be explicitly informed that their whole genome sequence data could be used in research. When possible, individuals should be given the option to withhold their data from certain types of future research to avoid inadvertent complicity with research goals to which they are opposed. The Commission acknowledges the complexity of integrating individual options into the research enterprise, but if a framework is in place that accommodates identifying specific participant preferences at the time of enrolling in research, such as proposed in the ANPRM, these preferences should be honored.¹⁸⁵

As long as consent processes are equivalently effective in informing individuals about what they are consenting to, and as long as they do not unduly shape or undermine individuals' ability to make genuinely voluntary choices, there is no philosophical or ethical imperative to use one kind of consent process over another. In cases where the public stands to benefit from an activity and the research consent is fully informed and consistent with the ability to make autonomous choices, it might be advantageous to use consent processes that make it easier for individuals to participate—but most definitely not “trick” them into participating—at higher rates. In other words, the most important issue in consent is not the type, but rather that the consent is properly informed and consistent with voluntary choice.

Opt in consent policies assume that the default is not to go forward with some proposal, such as to consent to whole genome sequencing; the individual must actively consent to the proposal in order for anything to happen. Opt out consent means that, in the absence of a refusal, the default is participation, which tends to encourage higher rates of participation, a result particularly supportive of the public value of scientific and medical research that is otherwise ethically and legally sound.

BIOVU: AN OPT OUT DATABASE

Vanderbilt's BioVU database, which has collected DNA samples from almost 150,000 individuals, is an opt out database. Unless patients check a box indicating that they do not want their DNA in the BioVU database, their samples are included. In this way, BioVU is able economically to collect a large number of samples. To protect the data in its database, the samples are coded before being entered in the database. The computer system can match the DNA with information in medical records, but researchers working with the data do not know to whom the data belong. Data are stripped of identifiers before being shared with secondary researchers.

Vanderbilt University Medical Center. (2012). Vanderbilt BioVU. Retrieved from <http://www.vanderbilthealth.com/main/25443>.

Organ donation policies in Europe provide an example of opt out consent procedures. Austria, France, Hungary, Poland, and Portugal have opt out organ donation policies and all have organ donation consent rates above 99 percent.¹⁸⁶ The United States, on the other hand, uses an opt in system. Polls show that about 90 percent of Americans support organ donation, but only about 44 percent of people in the United States opt in to be organ donors.¹⁸⁷ This indicates that where Americans' values dispose them in favor of consent to organ donation, the often cumbersome and anxiety-inducing procedures of an opt in policy make them reconsider, passively resist, or fail to follow through with the extra steps (like filling out extra forms) required to opt in.¹⁸⁸

With some exceptions, federally funded research studies are required by law to obtain informed consent from all individuals enrolled in research or from their legally authorized representative.¹⁸⁹ The informed consent document is one component of the informed consent process. Current federal regulation requires that informed consent documents include, among other things, a description of the procedures in the research plan, an explanation of the risks and benefits to the participant, a description of the extent to which confidentiality of records will be maintained, and an explanation of the right to withdraw from the study.

By regulation, research participants can withdraw from research to which they consented at any time for any reason. However, complete destruction of whole genome sequence data is likely impossible. Although physical biospecimens and data files stored by the primary researchers can and will be destroyed at the time of withdrawal according to guidelines laid out in consent documents, the destruction of distributed copies of associated data files may not be feasible as distributed genome sequence data files can be stored on local computers or network servers. Therefore, those conducting whole genome sequencing research might not be able to promise complete withdrawal from a study.

Recommendation 3.2.

The federal Office for Human Research Protections or a designated central organizing federal agency should establish clear and consistent guidelines for informed consent forms for research conducted by those under the purview of the Common Rule that involves whole genome sequencing. Informed consent forms should: 1) briefly describe whole genome sequencing and analysis; 2) state how the data will be used in the present study, and state, to the extent feasible, how the data might be used in the future; 3) explain the extent to which the individual will have control over future data use; 4) define benefits, potential risks, and state that there might be unknown future risks; and 5) state what data and information, if any, might be returned to the individual.

Each government agency has its own enforcement authorities to protect research participants. For example, the Office for Human Research Protections has jurisdiction over human research conducted or supported by HHS, the Central Intelligence Agency has a Human Subject Research Panel, and the Department of Veterans Affairs uses a combination of Research Compliance Officers and its Office of Research Oversight. All these agencies should work together as each agency develops clear and consistent guidelines for their informed consent forms, enabling an individual to make a fully informed decision to participate in research.

Looking forward, clinical consent documents for whole genome sequencing will have to address a number of issues specific to whole genome sequencing: an explanation of the science, what types of results will be produced through whole genome sequencing, and whether whole

genome sequence data collected for clinical applications will be made available for research purposes.

Further, whole genome sequence data can provide information about many conditions, not just the condition under study. Acknowledging this, informed consent documents for studies involving whole genome sequencing should include which (if any) research results and incidental findings will be returned to individuals.¹⁹⁰

“Now, on the other side of the ledger...are the findings... which the patient is not expecting...which are going to have a dramatic impact of known consequence to them, and then the set of things for which there is much less certain impact.”

Richard Gibbs, Wofford Cain Professor, Department of Molecular and Human Genetics; Director, Human Genome Sequencing Center, Baylor College of Medicine. (2012). Ethics and Practice of Whole Genome Sequencing in the Clinic. Presentation to PCSBI, February 2, 2012. Retrieved from <http://bioethics.gov/cms/node/658>.

In whole genome sequencing, many individuals might want, and even expect, access to data or results.¹⁹¹ From the perspective of many individuals, the inability to receive or access their data denies them a fundamentally important sense of control over information about their own genomic makeup. While some individuals wish to share their data broadly for the advancement of science, others want control over their data to maintain their privacy, control information shared with intimate relations, or protect their right *not* to know results that might be discovered during whole genome sequencing. Individuals who seek return of data or results often feel that if someone else knows something unique about them, such as their risk for a particular disease, they ought to know it as well.¹⁹² On the other hand, some experts have said that although participant or patient preferences should be considered in the return of results, individual preferences are not a sufficient reason for agreeing to return results because of the importance of ensuring that the results are accurately communicated to individuals. These experts argue that the decision of whether to return incidental findings and other data should be in the hands of those who can more fully understand the broad implications of returning those findings, and what needs to accompany the return of raw results. They call for criteria to be developed, for example, by return of results committees.¹⁹³

There are, of course, reasons within our current research systems for not returning research results to individuals enrolled in research studies as well. First, by current law, only sequencing results from Clinical Laboratory Improvement Amendments (CLIA) compliant laboratories may be returned to individuals.¹⁹⁴ This requirement came about in the 1980s as a result of stories in the media that raised concerns about the quality of laboratory results, especially the return of false-negative Pap smear results.¹⁹⁵ This attention catalyzed the passage of CLIA in 1988, designed to improve quality and consistency in clinical laboratory testing. CLIA made it illegal to return to patients clinical results generated in a non-CLIA-certified laboratory.¹⁹⁶ Currently, most research is not conducted in CLIA-certified laboratories, including those laboratories performing whole genome sequencing.¹⁹⁷ In addition, researchers leading projects that are producing whole genome sequence data might not be qualified or trained to return sensitive, potentially devastating results directly to individuals, nor are grants usually structured to hire someone with the appropriate qualifications to do so.

Ethical analysis of whether and how individual research results and incidental findings should be returned is ongoing, and these questions are currently the subject of wide-ranging debate.¹⁹⁸ Many agree that participants should have the option to opt out of receiving research results and data from a study. There is less consensus on what should be done in cases where individuals want to receive incidental research results and data but, for example, researchers or clinicians did not themselves collect the information, are not trained in interpreting incidental results, did not perform the sequencing in a CLIA-approved lab, or have no prior knowledge of or relationship to the individual to appropriately convey the results. Alternatively, in some cases, investigators might feel personally obligated to provide research results that could be clinically meaningful.¹⁹⁹

One example that illustrates this dilemma is the Alzheimer's risk associated with certain variants of the ApoE gene. Individuals who carry the ApoE4 variant have a higher risk of developing Alzheimer's disease, but not everyone with this variant will develop Alzheimer's disease. Suppose that whole genome sequencing is being performed on a young adult for a breast cancer research study he or she is involved in, and the ApoE4 variant is discovered. Should this finding be returned? The finding is not clinically actionable—meaning that there is not an effective treatment or cure—and it is not certain that individuals with the ApoE4 variant will develop Alzheimer's disease. Some argue that the only acceptable reason to return an incidental finding is that the finding is clinically relevant and actionable, and the ApoE4 variant's association with Alzheimer's disease fails to cleanly meet these criteria.²⁰⁰

Others argue that it should be completely up to the individual whose whole genome is sequenced to make this decision.²⁰¹

A number of frameworks for return of research results and incidental findings have recently been proposed by broadly constituted groups. A recent consensus paper authored by academic researchers, legal scholars, and patient advocates determined that researchers should offer to return individual research results that 1) are analytically valid; 2) are in compliance with CLIA; 3) the patient has consented to receiving; 4) are clinically actionable; and 5) present an “established and substantial risk of a serious health condition.”²⁰² Another framework proposes grouping incidental findings into three “bins” including: “clinically actionable,” “clinically valid but not directly actionable,” (subdivided into low-, medium-, or high-risk incidental information groups), and “unknown or no clinical significance.”²⁰³ The bin into which the data fall in this model, in combination with other variants, determines if the result should be reported to the participant in a clinical context. Models also exist that are more finely tuned and consider multiple variables, such as participant preference (what results the participant does and does not want to know), significance of the result (analytic validity of the test and possibility for medical intervention), and communicability (literacy of the participant and clarity of the message).²⁰⁴

In contrast to these fine-tuned, multivariable return of results frameworks, many representatives of the patient advocacy community propose the wholesale return of whole genome sequence data to individuals. They argue that although universities or companies provide a service by performing whole genome sequencing, the individuals who supplied the samples should retain the right to control the use of the data, access to the data, and be able to share the data with whomever they choose (such as with researchers conducting other studies related to conditions affecting the individuals or the individuals' families).²⁰⁵

There is a difference however between the return of “data” and “information” in the context of whole genome sequencing. Some have suggested that regardless of whether meaningful

information (that is, analyzed data interpreted by experts) is made available, raw data might be valuable to individuals. Currently, the Food and Drug Administration is debating the classification of these data in the context of commercial genetic testing.

If companies are returning results with clinical or medical significance, commercial genetic services might be subject to regulatory requirements; but if they are simply returning unanalyzed whole genome sequence data files, regulatory requirements might not apply.²⁰⁶ For example, the commercial genetic test company Lumigenix does not interpret medically relevant genetic variants in-house. Rather, it provides customers with raw whole genome sequence data, inviting the consumer to use free genome analysis software to discover and interpret clinically relevant information on their own.²⁰⁷

This is a mere sampling of the many complex and detailed issues that need to be addressed before reaching a comprehensive set of actionable recommendations about whether and when incidental findings from whole genome sequencing can and should be returned to individuals with their fully informed consent.

Recommendation 3.3.

Researchers, clinicians, and commercial whole genome sequencing entities must make individuals aware that incidental findings are likely to be discovered in the course of whole genome sequencing. The consent process should convey whether these findings will be communicated, the scope of communicated findings, and to whom the findings will be communicated.

Recommendation 3.4.

Funders of whole genome sequencing research should support studies to evaluate proposed frameworks for offering return of incidental findings and other research results derived from whole genome sequencing. Funders should also support research to investigate the related preferences and expectations of the individuals contributing samples and data to genomic research and undergoing whole genome sequencing in clinical care, research, or commercial contexts.

Individuals undergoing whole genome sequencing in research, clinical, and commercial contexts must be provided with sufficient information in informed consent documents to understand what incidental findings are, and to know whether they will be notified of incidental findings discovered as a result of whole genome sequencing.²⁰⁸ Users of whole genome sequence data should continue supporting research into the management of incidental findings and individual research results obtained in both CLIA and nonCLIA-certified laboratories.

Previous research has generated many models and guidelines for returning incidental findings and other results obtained in clinical and basic research. In order to take the next step of translating these models into best practices for the return of results, additional data must be collected to inform the deliberations. In particular, research should be expanded to collect empirical data on participant, patient, researcher, and clinician opinions of each model, and the consequences and costs of implementing each model. These studies should examine the motivations of patients and participants enrolled in research, undergoing genome sequencing in the clinical context, or engaging in commercial whole genome sequencing to obtain their research results. Respect for patient and participant values is essential to guide the development of these tools ethically.

Facilitating Progress in Whole Genome Sequencing

Current protections for research participants emerged from a series of lapses in research ethics uncovered in the 1960s and 1970s in which clinicians and scientists conducted research without the fully informed consent or even knowledge of the research participants.²⁰⁹ One outcome of this history was the drawing of a bright line between clinical care and research. But this distinction is no longer so clear. Currently, large amounts of patient data are being collected in the health care setting, stripped of traditional identifiers, analyzed, and fed into research that might one day improve clinical care. This learning health system model both translates advances in health services research into clinical applications and collects data during clinical care to facilitate further advances in research.²¹⁰ With patient data increasingly being transitioned to electronic medical records, persons engaged in this type of research can also more easily access data to aggregate and analyze.²¹¹

Advocates of the learning health system model advocate encouraging intellectual freedom through clinical research and engaging in regulatory parsimony.²¹² Large amounts of data are essential for researchers to make correlations between genomic variants and disease states. Learning health system advocates and others call for standardized electronic health record systems and infrastructure to facilitate health information exchange so that data can be easily aggregated and studied.²¹³ Integrating whole genome sequence data into health records within the learning health system model can provide researchers with more data to perform genome-wide analyses, which in turn can advance clinical care. Several Institute of Medicine (IOM) working groups have supported these goals, outlining the desirability of establishing a universal health information technology system and learning environment that engages health care providers and patients. The IOM reports recommend that such a system include both genomic and clinical information, increased interoperability of medical records systems, and reduced barriers to data sharing.²¹⁴ The President's Council of Advisors on Science and Technology identified the lack of sharing electronic health records—with patients, with a patient's health care providers at other organizations, with public health agencies, and with researchers—as a barrier to improved health care.²¹⁵

Recommendation 4.1.

Funders of whole genome sequencing research, relevant clinical entities, and the commercial sector should facilitate explicit exchange of information between genomic researchers and clinicians, while maintaining robust data protection safeguards, so that whole genome sequence and health data can be shared to advance genomic medicine.

Performing all whole genome sequencing in CLIA-approved laboratories would remove one of the barriers to data sharing. It would help ensure that whole genome sequencing generates high-quality data that clinicians and researchers can use to draw clinically relevant conclusions. It would also ensure that individuals who obtain their whole genome sequence data could share them more confidently in patient-driven research initiatives, producing more meaningful data. That said, current sequencing technologies and those in development are diverse and evolving, and standardization is a substantial challenge. Ongoing efforts, such as those by the Standardization of Clinical Testing working group are critical to achieving standards for ensuring the reliability of whole genome sequencing results, and facilitating the exchange and use of these data.²¹⁶

In order for all persons to benefit from whole genome sequencing research, diverse populations must be involved in research. Consequently, it is incumbent upon the research community to earn and maintain the trust of individuals from a wide range of diverse populations across society. This trust is particularly important in minority and marginalized populations where levels of trust in the medical and research communities have been historically low.

To encourage such trust, some scholars and advocates have proposed alternative models for the interactions between researchers and individuals enrolled in research that attempt to increase transparency and shift the balance of control between these two parties.²¹⁷ As opposed to the traditional research model, in which there is usually little contact between the researcher and the individual enrolled in research beyond initial sample contribution, participant-centric initiatives put research participants at the center of the decision making, and are based on principles of respect and empowerment.²¹⁸ The federal government has shown an interest in giving patients a better understanding of disease, treatment, and care options through its establishment of the Patient-Centered Outcomes Research Institute.²¹⁹

The challenges we face today in whole genome sequencing are not (or only partially overlap with) the challenges we will face in the coming years as technologies continue to develop and mature. For example, one current concern is the integration of data into electronic medical records; in 20 years or less, society might have to decide if every newborn should have their whole genome sequenced and added to their electronic medical record. Due to rapid technological developments, today's policies must be crafted specifically enough to be actionable and targeted to address our current concerns, yet agile enough to ensure that we do not constrain our ability to adapt to evolving technology, research, and social norms related to privacy and sharing.²²⁰

Recommendation 4.2.

Policy makers should promote opportunities for the public to benefit from whole genome sequencing research. Further, policy makers and the research community should promote opportunities for the exploration of alternative models of the relationship between researchers and research participants, including participatory models that promote collaborative relationships.

Respect for persons implies not only respecting individual privacy, but also respecting research participants as autonomous persons who might choose to share their own data. Public beneficence is advanced by giving researchers access to plentiful data from which they can work to advance health care. Regulatory parsimony recommends only as much oversight as is truly necessary and effective in ensuring an adequate degree of privacy, justice and fairness, and security and safety while pursuing the public benefits of whole genome sequencing.

Therefore, existing privacy protections and those being contemplated should be parsimonious and not impose high barriers to data sharing.²²¹ While the Commission supports the intellectual freedom this access will encourage, clinicians and researchers must also act responsibly to earn public trust for the research enterprise.

Public Benefit

The federal government has made a substantial investment in genetics research, including whole genome sequencing, and the benefit of this investment has been realized in two major ways. First, disease diagnosis and treatment have been advanced, and the functions of many genes have been and will continue to be discovered, which will further improve clinical care in the coming years. Incorporating knowledge gained through advances in whole genome sequencing into the clinic could improve diagnosis and treatment of diseases that have brought turmoil and tragedy into the lives of individuals and their families. We have already begun to see some benefits resulting from these advances; for example, genetic variants that can lead to adverse drug reactions have been identified. In the future, as the genetic variations that underlie common diseases are discovered, clinicians will, in some instances, be able to detect predispositions to disease before those diseases occur, and begin treatment or recommend lifestyle changes before a patient exhibits symptoms. Second, an indirect economic benefit has been realized. The U.S. government invested \$3.8 billion in the Human Genome Project; it is estimated that this investment generated \$244 billion in personal income and \$796 billion in overall economic impact.²²² These health and economic gains not only benefit the public through improved health care but also through increased economic opportunities.

Thousands of citizens have participated in whole genome sequencing research personally, and all citizens help support government investment in whole genome sequencing through their participation in and support of our political system. Therefore, all citizens should have the opportunity to benefit from medical advances that result from whole genome sequencing.

Special caution should be taken on the part of researchers to ensure that their participants reflect as much as possible the rich diversity of our population. Different groups have genomic variants at different frequencies within their populations, and sufficiently diverse data must be collected so that advances arising from whole genome sequencing can be used for the benefit of all groups.²²³

Recommendation 5.1.

The Commission encourages the federal government to facilitate access to the numerous scientific advances generated through its investments in whole genome sequencing to the broadest group of persons possible to ensure that all persons who could benefit from these developments have the opportunity to do so.

Government investment in genomic research has resulted in public benefit through improved health care and in economic return on investment. The principle of justice and fairness requires that the benefits and risks of whole genome sequencing be distributed across society. Research funded with taxpayer contributions should benefit all members of society. To these ends, researchers should be vigilant about including individuals from all sectors of society in their studies, so that research findings can be translated widely into clinical care. The federal government should follow through on its investment in research and assure that the discoveries of whole genome sequencing are integrated with clinical care that can be accessed by all.

APPENDIX I: GLOSSARY OF KEY TERMS

Allele: a form of a gene at a particular location on a chromosome.

Biorepository: a stored collection of physical biological samples (e.g., blood or tissue) and associated data (e.g., medical information and policies). Sometimes called a biobank.

Carrier: an individual who has one normal and one mutated version of a gene.

Chromosome: X-shaped structure made of tightly wrapped DNA in the nucleus of the cell that carries genes from one generation to the next. Humans have 46 chromosomes (in 23 pairs).

Clinical utility: an assessment of the risks and benefits associated with a clinical test and the likelihood that the test will result in improved patient outcome.

Clinical validity: the degree to which a genetic test can predict clinical status, as measured by the strength of the association between the genotype and phenotype.

Copy number variations (CNVs): DNA mutations that occur when large sections of DNA are inserted or deleted during cell division.

Database: an organized collection of data or information (e.g., whole genome sequence data files and information).

Deoxyribonucleic acid (DNA): the molecule that contains the instructions to develop and direct the biological and chemical activities of a living organism.

DNA Sequencing: the process that identifies the order of the nucleotide bases in a strand of DNA.

Exome Sequencing: DNA sequencing of only the parts of the genome that make proteins (exons).

Exon: a stretch of DNA, part of a gene, that codes for a protein.

Gene: a piece of DNA that contains the information required for making a product that will have a biological function. A full set of genes is called a genome.

Gene-environment interaction: the environmental factors that can influence a gene's expression and the resulting phenotype.

Genetic test: a discrete test that examines a specific genetic location or a single gene, such as the test for Huntington's disease.

Genetic variation: differences in alleles of allele frequency between or among individuals or populations.

Genomics: the study of all the DNA (the genome) in an individual, and how parts of the genome interact with each other and the environment.

Genome: the full set of genes in an individual. Humans have about 20,000 to 25,000 genes in their genome.

Genome-wide association study (GWAS): compares large amounts of genetic data from individuals with and without a specific condition to identify DNA variants that correlate with diseases.

Genotype: the genetic make-up of an individual.

Genotype/phenotype correlation: the association between a certain mutation (genotype) and the resulting physical characteristic (phenotype).

Genotyping: analyzing discrete variants, from a handful to thousands, across the genome (i.e., more than a discrete genetic test, but less than whole genome sequencing).

Guthrie Card: piece of paper used to capture and store a few drops of blood collected from a newborn. DNA from the dried blood spot is then used to test for a range of genetic conditions and infections.

Heterozygous: when the genes or alleles on the two chromosomes are different.

Homozygous: when the genes or alleles on each of the two chromosomes are the same.

Incidental finding: a finding discovered in the course of clinical care or research concerning a participant that is beyond the aims of the clinical test or research but has potential health importance.

Individual research result: a finding discovered in the course of clinical care or research concerning a research participant that relates to the aims of the clinical test or research and has potential health importance.

Intron: part of a gene present between exons that does not directly code for a protein.

Locus: the location of a gene on a chromosome.

Mutation: a change in the DNA sequence. Mutations can arise from mistakes during cell division or from an outside source (e.g., radiation from the sun).

Nucleotide bases: the four chemical units that compose DNA. The bases are adenine (A), thymine (T), guanine (G), and cytosine (C). A always pairs with T on the opposite strand of DNA, and C always pairs with G. One A-T or G-C pair is called a base pair.

Phenotype: the expression of an individual's genotype. An individual's phenotype consists of their physical characteristics.

Public health utility: the likelihood that a clinical test will reduce disease burden and/or result in improved patient outcome in the population.

Single nucleotide polymorphisms (or SNPs): variations in the genome that involve single base pairs.

Structural variants: the insertion, deletion, duplication, translocation (the movement of DNA from one location to another on the same or another chromosome), or inversion (flipping over) of long DNA segments (greater than about 1,000 base pairs in length).

Whole genome sequence data: the file of As, Cs, Gs, and Ts produced as a result of whole genome sequencing.

Whole genome sequence information: facts derived from whole genome sequencing data, such as predisposition to disease.

Whole genome sequencing: determining the order of nucleotide bases— As, Cs, Gs, and Ts—in an organism's entire DNA sequence.

APPENDIX II: GENETIC AND GENOMIC BACKGROUND INFORMATION

Understanding Basic Genetic Architecture

Deoxyribonucleic acid (DNA) is the molecule that contains the instructions to develop and direct the biological and chemical activities of nearly all living organisms. DNA is a twisting pair of strands, called a double helix, made of four basic building blocks, or *nucleotide bases*. These bases are abbreviated A, T, C, and G. The As, Cs, Gs, and Ts are linked together in long strands. The A on one strand will link to a T on the other strand of the double helix, bringing

the two strands together at each point along the DNA strand, like rungs on a ladder. A always binds with T, and C always binds to G. One A-T or G-C pair is called a *nucleotide base pair*. If the DNA in a single human cell was stretched out, it would be about six feet long. If all the DNA in a human body was stretched out, it would reach almost 70 times from the earth to the sun and back.²²⁴ In order to fit this much DNA into cells, the long strands of DNA have to be stored compactly. In the cell, DNA is nearly always wrapped tightly into X-shaped structures called *chromosomes*, which prevent the long strands of DNA from tangling or being damaged. Chromosomes pass DNA from one generation to the next.

MENDELIAN GENETICS

Gregor Mendel, a 19th Century European monk, discovered the mechanism for trait inheritance in plants and animals. Mendel studied traits in peas, including flower color, stem length, seed shape, and seed color. Through selective pollination, he was able to observe how traits were expressed when two plants produced seed. He found that organisms have two copies of every inheritance “unit” (now called genes): one from each parent.

Chromosomes are located in the nucleus of a cell (a sub-compartment of the cell that stores DNA). Chromosomes are usually found in pairs, with one member of each pair coming from the individual’s genetic mother and the other from the genetic father (See Figure 3). Humans have 46 chromosomes, in 23 pairs. Of the 46 chromosomes, two are sex chromosomes (X and Y) that determine if an individual is male or female. In addition to the 22 pairs of chromosomes that all humans have, females inherit one X chromosome from each parent, making their 23rd pair XX, while a male inherits an X chromosome from his mother and a Y chromosome from his father, making his 23rd pair XY.

A complete set of DNA, or a full set of 46 chromosomes in a human, is called his or her *genome*. In humans, the genome is made up of approximately 3 billion nucleotide base pairs (A-T and G-C pairs). Nearly every cell in the human body contains a complete copy of the genome.

WHY IT IS GOOD TO HAVE TWO COPIES OF EACH CHROMOSOME

Having two copies of each gene ensures that if one gene on one chromosome of a pair is damaged, the gene on the other chromosome of the pair might not be damaged. In most cases, having only one functional copy of a gene is sufficient for normal function. This could be compared to having two kidneys: if one kidney is damaged, the healthy one can function well enough so that the individual can lead a relatively normal life.

Genes are specific regions of DNA on chromosomes. Genes are the basic physical unit of inheritance, and are passed from parents to their children. There are approximately 20,000-25,000 genes distributed over the 46 chromosomes. Together, these genes make up the blueprint for the body and how it functions. The location of a gene on a chromosome is called the *locus*, which is much like an address. For example, a gene can be found on chromosome 16 (e.g., the name of the street), on a particular end of the chromosome (e.g., the North or South end of a street), at a particular location (e.g., the house number).

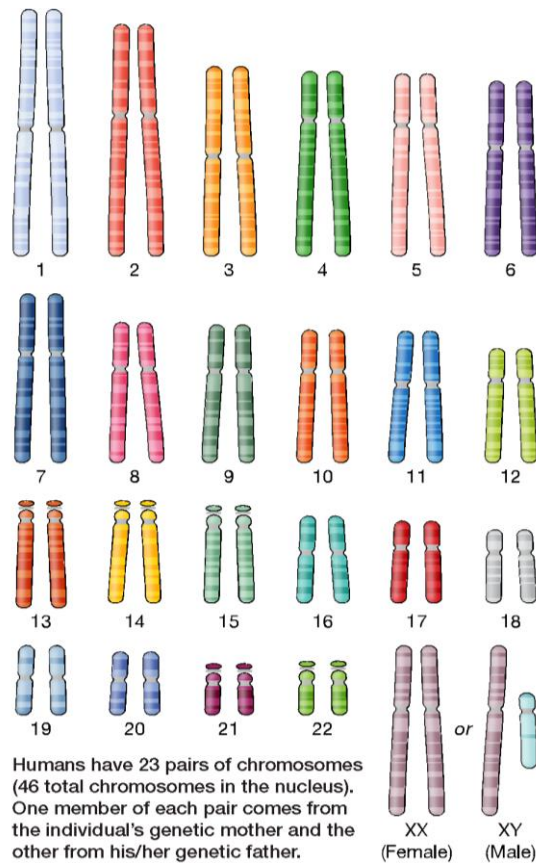


Figure 3. The 23 pairs of human chromosomes.

Almost all genes come in pairs with one copy (or *allele*) from each parent. While the alleles might or might not be identical, the genes are the same (just like we all have ears, but our ears do not all look exactly alike). Every person has the same number of genes, although they might have different alleles from one another; that is, every person has the genes for cystic fibrosis (*CFTR*) and breast cancer (*BRCA1/2*), but most of us do not have disease-causing mutations in these genes.

In order to go from the blueprint in the genes to a functioning human, information in DNA is turned into proteins. Genes contain the instructions for making proteins that make up the human body. Examples of proteins include collagen, which is a major component of our hair and skin; and enzymes, a special type of protein, some of which break down the food we eat. If the DNA coding for a protein is mutated, it could result in that protein not functioning. For example, if the enzyme that breaks down lactose (a protein found in milk) is not assembled properly, it cannot break down lactose effectively and an individual is said to be lactose intolerant.

Not every single part of our DNA contains the instructions for making a protein, only certain parts of genes make proteins. These regions are called *exons*. The function of the regions of DNA that do not code for proteins, called *introns*, is unknown. Introns were once called “junk” DNA, but scientists are learning that introns are likely essential for the rest of the gene to function properly.

GENE-ENVIRONMENT INTERACTIONS

Cystic Fibrosis is a recessive genetic disorder, which means that a child must inherit a mutated copy of the *CFTR* gene from each parent to have the disease. While the genetic cause is clear, the severity of disease is linked to environmental factors such as exposure to second hand smoke, stress, and poor nutrition. Smoke in particular has been shown to interact with the *CFTR* gene and a secondary gene as well, worsening lung function in the patient.

Source: Collaco, J.M., et al. (2008). Interactions between secondhand smoke and genes that affect cystic fibrosis lung disease. *Journal of the American Medical Association*, 299(4), 417-424.

The term *genotype* refers to an individual's collection of genes or to the two alleles inherited for a particular gene. The expression of the genotype, through making proteins, contributes to the individual's outward characteristics, called their *phenotype*. The association between a certain mutation or mutations (genotype) and the resulting physical characteristics (phenotype) is called the *genotype/phenotype correlation*. This association is at the core of genetic testing and research.

Gene-environment interaction refers to how environmental factors modify the expression of a gene and, therefore, the trait, or phenotype. Some phenotypic traits are strongly influenced by genes, while others are more strongly influenced by the environment. Most traits are influenced by one or more genes interacting in complex ways with the environment.

Genetic Variation

A *mutation* is a change in a DNA sequence (see Figure 4). Mutations can come from mistakes that happen when DNA is copied during cell division, exposure to chemicals or harmful radiation (like UV rays from the sun), or infection with certain viruses. Some mutations occur in the cells of an individual's body and are not passed on to offspring, such as DNA damage in the skin caused by sunburn. Other mutations occur in the eggs and sperm and can be passed on to offspring, such as a mutation in the gene for sickle cell anemia.

The term *genetic variation* refers to differences in alleles and other genetic changes between or among individuals. Genetic variation can also refer to how often those differences in alleles occur between or among populations.

Humans have about 99.9 percent of our genetic information in common, but there is considerable genetic variation. The differences in our genomes can explain why we are diverse as individuals or populations in appearance, predisposition to specific diseases, and adaptation to our environment. Understanding genetic variation is at the heart of understanding the role of genetics in disease.

Genetic variations involving only a single nucleotide base (an A, C, G, or T building block) are referred to as *single nucleotide polymorphisms*, or *SNPs* (pronounced "snips"). Most people have thousands of SNPs in their genomes, but they often occur in the parts of DNA that do not make proteins, so they do not cause disease. When SNPs occur within a gene, they might cause disease by affecting the gene's function. Researchers have found SNPs that might help predict how an individual responds to certain drugs, their susceptibility to environmental factors such as toxins, and their risk of developing particular diseases. SNPs have been used extensively to study diseases that are passed from one generation to the next in families.

Original sequence

THE SKY IS BLUE

SNP (single nucleotide polymorphism)

THE SKY IS BLUE → THE SEY IS BLUE

Deletion or insertion of stretches of DNA

THE SKY IS BLUE → THE SKY BLUE

THE SKY IS BLUE → THE SKY IS BLUE

VNTR (variable number of tandem repeats)

THE SKY SKY SKY SKY SKY SKY IS BLUE

CNV (copy number variant)

THE SKY YYY IS BLUE

Figure 4. Types of Genetic Variations.

Genetic variation can also involve much longer stretches of DNA. *Structural variants* involve the insertion, deletion, duplication, translocation (the movement of DNA from one location to a not her on the same or another chromosome), or inversion (flipping over) of long DNA segments (greater than about 1,000 base pairs in length). One type of structural variation is a copy number variant. *Copy number variations* (CNVs) can occur when large sections of DNA are inserted or deleted during cell division. Scientists are trying to understand how copy number variation contributes to health and disease. Each person carries roughly 100 copy number variants, but many do not appear to have a disease linkage.

SICKLE CELL ANEMIA

Sickle cell anemia is caused by a SNP in the gene for hemoglobin, a protein in red blood cells that is responsible for carrying oxygen. If the hemoglobin gene is mutated on both alleles, an individual will have sickle cell disease, which leads to a shortened life span. If an individual has one normal hemoglobin allele and one mutated allele, however, they will not have sickle cell disease (because they also have one functional allele) *and* they will have some protection against malaria. Sickle cell disease is most common in populations who live in malaria-prone regions of the world, because carrying this mutation is actually protective against malaria.

Source: CDC. (n.d.). Protective Effect of Sickle Cell Trait Against Malaria Associated Mortality and Morbidity. Retrieved from http://www.cdc.gov/malaria/about/biology/sickle_cell.html.

How Genetic Variants Translate into Disease

Today, clinical genetic testing is used in individuals with a family history of disease; in other words, the tests are limited to those who are considered at risk of carrying known genetic variants that are linked to a particular disease. However, some clinical studies are evaluating the use of whole genome sequencing in regular clinical practice.²²⁵ In addition, individuals can try to bypass the traditional health care system and use the services of companies that offer consumers SNP analysis, whole exome sequencing, and more.

Very few genetic variants are directly linked to a specific disease, however some examples include cystic fibrosis, sickle cell anemia, and Huntington's disease. Targeted genetic tests have been developed for many of these diseases. Many other diseases are suspected to have a genetic component, but scientists have not determined which genetic variants might cause them. For example, heart disease could be caused by genetic mutations, but it is certainly not a simple case of one mutation in one gene. The genetic component of heart disease could be many mutations throughout the genome that interact with the environment to cause heart disease. Whole genome sequencing could reveal complex interactions between genes and disease, where a particular mutation on a certain gene, in conjunction with another mutation on another gene, or several other mutations on other genes, come together to cause disease.

Some of the genetic variants discovered during whole genome sequencing will have clear links to disease, but the majority will be unknown. Based on how they translate to disease, genetic variants can fall into six categories:

- *Variants of unknown significance*: An example of this might be when a piece of DNA has been cut out of one location on a chromosome and inserted into another location on the chromosome. The fact that the DNA is different is clear, but what that difference means, or how it will relate to disease, is unclear.
- *Nonmedical genetic markers*: These are genes that code for things such as eye color. If there were a mutation in one of these genes, it would not be something that would require medical treatment.
- *Carrier status*: An individual is a *carrier* of a variant if they have one normal and one mutated version of a gene. Most often, the individual is not affected by the disease, but they can pass the gene on to their children. An example is sickle cell disease, where individuals with one mutated version of the gene and one normal version of the gene do not have the full-blown disease themselves.
- *Susceptibility genes*: These are genes that make it more likely, but not certain, that an individual will develop a particular disease, i.e., they are "susceptible" to it. An individual might carry genes that make them susceptible to diabetes, but with proper diet and exercise, they will not necessarily develop diabetes.
- *Late onset genetic conditions*: Late onset conditions present later in life. Examples are Alzheimer's disease, Huntington's disease, and some degenerative eye diseases.
- *Medical conditions found by current prenatal genetic tests*: These are conditions that, if an individual has one or two copies of the gene, they will have the disease, and the disease will affect their health and quality of life throughout their life span. An example is phenylketonuria. Individuals with phenylketonuria cannot break down a particular amino acid and must follow a diet that is low in that amino acid.

Sequencing Strategies

DNA sequencing is the process of determining the exact order of the bases (nucleotides) in a strand of DNA. Since base pairing is predictable (A always pairs with T; G always pairs with C), knowing the sequence on one strand automatically reveals the sequence on the other strand. Sequencing technology has rapidly advanced in recent years, allowing scientists to make discoveries about the regulation, variability, and evolution of the human genome.²²⁶

A consequence of decreasing cost and increasing accessibility of sequencing technologies is the increasing use of whole genome sequencing. *Whole genome sequencing* is the process of sequencing all the DNA in an organism, in contrast to testing for only a handful of known mutations or sequencing a particular gene. Whole genome sequencing reads more than 95 percent of the genome, compared to SNP genotyping, which typically covers less than 0.1 percent of the genome. That said, knowing one person's complete DNA sequence does not necessarily provide useful clinical information, because each person's DNA is different from the DNA of others at millions of places. One goal of whole genome sequencing research is to create a reference catalog of all common and rare genetic variants in human populations so that the relationship between variants and disease can be studied. By comparing one person's whole genome sequence with other whole genome sequences, reference sequences, and associated health information, one can find places in the genome where, for example, a group of people with the same DNA mutation at the same locus all have the same disease. Comparisons like this will hopefully lead to meaningful associations and ultimately guide clinical and personal health decisions.

Exome sequencing might be an efficient alternative to whole genome sequencing in some cases. Exome sequencing selectively sequences only the part of the genome that make proteins (exons). An estimated 85 percent of disease-causing mutations are found in the exome.²²⁷ Therefore, sequencing only the exons, which make up about 1 percent of the genome, should be faster and less expensive than sequencing the entire genome, and is likely to identify most disease-causing mutations. Increasingly, exome sequencing is being used in clinical diagnostic testing. However, now that 80 percent of the genome has been found to have "biochemical function," with non-coding regions of the genome influencing the activity of genes that are spatially distant, exome sequencing that does not find an answer could be complemented by targeted sequencing of non-coding regions.

A *genome-wide association study (GWAS)* is a method that has been used heavily in recent years to identify links between specific genetic variations and specific diseases. The method involves studying the genomes of many people with and without a disease of interest and searching for genetic markers (e.g., SNPs) that can be used to predict the presence of a disease. GWASs alone cannot specify which genes cause disease; however, by looking at hundreds of thousands of SNPs, researchers can identify mutations that are more frequent in people with the disease than without. These mutations are therefore considered "associated" with the disease. Disease-associated SNPs are used as markers or pointers to the region of the genome where a disease-causing mutation is likely to be found.

The Challenges of Analyzing Whole Genome Sequence Data and Identifying Disease Associations

The primary goal of whole genome sequencing research is to describe the relationship between genotype (genetic variants) and phenotype (physical characteristics, including

disease). Whole genome sequence data alone will not provide a complete understanding of disease. The data must be linked to phenotypic data, such as medical records. Environmental data will also be needed to fully understand gene-environment interactions.

A challenge of whole genome sequencing research is the hard-to-detect relationship between genetic variant and phenotypic trait, such as disease risk. To interpret an individual's disease risk, one must have reliable information about every validated genetic disease to use as a standard of comparison. Currently, there is no central, publicly available repository of all variants found to be associated with a clinically relevant trait or disease.

Refinements must also be made to take into account the genomic diversity of the human population. While no "private" variants have been found only in one population and not in others, many variants occur at different frequencies in different populations (for example, a particular SNP might be common in one population and rare in another). Studying genetic variation across populations can provide some, but not all, clues to the causes of health disparities.²²⁸

Finally, even if a specific mutation is linked to a disease, the expression of that gene and environmental interactions can result in different phenotypic effects in different people. In other words, one person carrying a particular mutation might develop the disease and another person with the same mutation might not, or that person might exhibit the disease in a more or less severe form. Further, a single mutation in one gene rarely leads to the particular phenotype of an individual.

The current clinical value of whole genome sequencing for linking genomic variants to disease remains challenging because of the many gene-gene and gene-environment interactions. Thus, the field continues to work toward establishing the *clinical validity* (future disease positive and negative predictive value stratified by exposure), *clinical utility* (targeted interventions to reduce disease risk among persons with the profile) and *public health utility* (comparing reduction of disease burden in the population based on genomic analysis) of whole genome sequence data.

APPENDIX III: GUEST PRESENTERS TO THE COMMISSION REGARDING PRIVACY AND WHOLE GENOME SEQUENCING

George Annas, J.D., M.P.H.

Chair, Health Law, Bioethics & Human Rights; William Fairfield Warren Distinguished Professor, Boston University School of Public Health

Richard Gibbs, Ph.D.

Wofford Cain Professor, Department of Molecular and Human Genetics; Director, Human Genome Sequencing Center, Baylor College of Medicine

Retta Beery

Mother of twins who benefitted from diagnosis made possible by whole genome sequencing

Jane Kaye, D.Phil., L.L.B.

Director, Centre for Law, Health and Emerging Technologies (HeLEX), Oxford University

Greg Biggers

Council Member, Genetic Alliance; Chief Executive Officer, Genomera

Ken Chahine, Ph.D., J.D.

Senior Vice President and General Manager, Ancestry DNA, LLC

Ellen Wright Clayton, J.D., M.D.

Craig-Weaver Professor of Pediatrics; Professor of Law and Director, Center for Biomedical Ethics and Society, Vanderbilt University

Leonard D'Avolio, Ph.D.

Associate Center Director for Biomedical Informatics, Massachusetts Veterans Epidemiology Research and Information Center (MAVERIC), Department of Veterans Affairs; Instructor, Harvard Medical School

James P. Evans, M.D., Ph.D.

Clinical Professor and Bryson Distinguished Professor of Genetics and Medicine, Department of Genetics, University of North Carolina School of Medicine

Madison Powers, J.D., D.Phil.

Professor, Department of Philosophy; Senior Research Scholar, Kennedy Institute of Ethics, Georgetown University

Laura Lyman Rodriguez, Ph.D.

Director, Office of Policy, Communications, and Education, National Human Genome Research Institute, National Institutes of Health

Mark A. Rothstein, J.D.

Herbert F. Boehl Chair of Law and Medicine, University of Louisville School of Medicine

Bartha Knoppers, Ph.D.

Director, Centre of Genomics and Policy; Canada Research Chair in Law and Medicine, McGill University

Daniel Masys, M.D.

Affiliate Professor, Biomedical and Health Informatics, University of Washington School of Medicine

Amy McGuire, J.D., Ph.D.

Associate Professor of Medicine and Medical Ethics; Associate Director of Research, Center for Medical Ethics and Health Policy, Baylor College of Medicine

Melissa Mourges, J.D.

Assistant District Attorney; Chief, Forensic Sciences/Cold Case Unit, New York County District Attorney's Office

Pilar Ossorio, Ph.D., J.D.

Associate Professor of Law and Bioethics, University of Wisconsin-Madison

Erik Parens, Ph.D.

Senior Research Scholar, The Hastings Center

Latanya Sweeney, Ph.D.

Visiting Professor and Scholar, Computer Science; Director, Data Privacy Lab, Harvard University

John Wilbanks

Founder, Consent to Research; Senior Fellow, Kauffman Foundation; Research Fellow, Lybba

Susan Wolf, J.D.

McKnight Presidential Professor of Law, Medicine & Public Policy; Faegre & Benson Professor of Law; Professor of Medicine; Faculty Member, Center for Bioethics, University of Minnesota

Professor of Law, George Washington
University

The map displays the following distribution of laws across the United States:

- Comprehensive Law (Dark Green):** CA, NV, OR, WA, NM, SD, IL, NY, VT, NH, ME, FL.
- Limited Law (Medium Green):** MT, ND, MN, WI, MI, IN, OH, PA, WV, VA, NC, SC, GA, AL, MS, AR, MO, IA, NE, WY, UT, AZ, CO, KS, OK, TX, LA, KY, TN, MS, AL, GA, FL.
- No Law on Point (Light Green):** ID, MT, ND, MN, WI, MI, IN, OH, PA, WV, VA, NC, SC, GA, AL, MS, AR, MO, IA, NE, WY, UT, AZ, CO, KS, OK, TX, LA, KY, TN, MS, AL, GA, FL.

An inset map shows Alaska (AK) and Hawaii (HI) in the dark green category.

* Map and table current as of March 2012.

State	Citation	Consent required to PERFORM genetic tests	Consent required to OBTAIN genetic information	Consent required to RETAIN genetic information	Consent required to DISCLOSE genetic information	Application/ context limited to
AL						no law on point
AK	Alaska Stat. §§ 18.13.010 through 18.13.100 (2011)	X	X	X	X	comprehensive law
AZ	Ariz. Rev. Stat. §§ 1-602, 12-2801 through 12-2804, & 20-448-02 (2011)	X			X	ordinary course of business
AR	Ark. Code. Ann. §§ 1643-1101, 20- 35-101 through 20-35-103 (2010)	X			X	ordinary course of business
CA	Cal. Civ. Code §§ 56.17, 56.265 (West 2010); Cal Ins Code 10123.35, 10148 through 10149.1 (2010)	X	X		X	ordinary course of business/ insurance
CO	Colo. Rev. Stat. §§ 10-3-1104.6 & 10- 3-1104.7 (2010)	X	X	X	X	healthcare provider or insurance company
CT						no law on point
DC						no law on point
DE	16 Del. Code Ann. §§ 1201 through 1208	X	X	X	X	comprehensive law
FL	Fla. Stat. § 760.40 (2010)	X			X	comprehensive law
GA	G.A. Code Ann. §§ 3354-1 through 33-54-8 (2011)	X	X		X	insurance
HI	Haw. Rev. Stat. §§ 431:10A-118 (2010)				X	insurance
ID						no law on point
IL	§ 225 Ill. Comp. Stat. 135/90, § 410 ILCS 513/1 through 513/91 (2011)				X	comprehensive law
IN	Ind. Code Ann. § 1639-5-2 (LexisNexis 2011)		X			insurance
IA						no law on point
KS						no law on point
KY						no law on point
LA	La. Rev. Stat. Ann. § 22:1023, 40:1299.6 (2011)		X	X	X	insurance
ME	22 Me. Rev. Stat. Ann. § 1711-C (2011)				X	healthcare providers
MD						no law on point
MA	Mass. Ann. Laws ch. 111, § 70G (2010)	X			X	healthcare providers
MI	Mich. Comp. Laws §§ 333.17020, 333.17520 (2011)	X				healthcare providers
MN	Minn. Stat. § 13.386 (2010)	X	X	X	X	government entity
MS						no law on point
MO	Mo. Rev. Stat. §375.1309 (2011)				X	ordinary course of business
MT						no law on point

State	Citation	Consent required to PERFORM genetic tests	Consent required to OBTAIN genetic information	Consent required to RETAIN genetic information	Consent required to DISCLOSE genetic information	Application/ context limited to
NE	Neb. Rev. Stat. § 71551 (2010)	X				healthcare providers
NV	Nev. Rev. Stat. Ann. §§ 629.141 through 629.201 (2010)	X	X	X	X	comprehensive law
NH	N.H. Rev. Stat. Ann. 141-H:1 through 141-H:6 (2010)	X	X		X	comprehensive law
NJ	N.J. Stat. Ann. §§ 10:5-44 through 10:5-49 (2011)	X	X	X	X	comprehensive law
NM	N.M. Stat. Ann. §§ 24-21-1 through 24-21-7 (2010)	X	X	X	X	comprehensive law
NY	N.Y. Civ. Rights Law § 79-1 (2011)	X	X	X	X	comprehensive law
NC						no law on point
ND						no law on point
OH						no law on point
OK	21 Okla. Stat. § 1175		X	X		newborns
OR	Or. Rev. Stat. §192.531 to 192.549 (SB 618)	X	X	X	X	comprehensive law
PA						no law on point
RI	R.I. Gen. Laws §27-1852, 52.3, §27-19-44, 44.1, §27-20-39. 39.1, §27-41-53, 53.1				X	insurance
SC	S.C. Code Ann. §3893-10 to §38-93-90; S.C. Code Ann. § 16-1-10	X			X	insurance
SD	S.D. Codified Laws §34-14-14 to -24	X				comprehensive law
TN						no law on point
TX	Tex. Ins. Code Ann. §546.001 et squ.	X			X	insurance
UT						no law on point
VT	V.T. Stat. Ann. tit. 18, §9331 to 9335	X			X	comprehensive law
VA	V.A. Code Ann. §38.2508.4				X	insurance
WA	Wash. Rev. Code §70.02.05 through 70.02.90; RCW 49.44.180				X	healthcare providers
WV						no law on point
WI	Wis. Stat. §942.07	X	X		X	employment
WY	Wyo. Stat. Ann. §14-2701 to 710	X			X	perform: comprehensive prohibition; disclose: paternity law

End Notes

- ¹ The individuals mentioned in these vignettes have given their permission to have their names included.
- ² For more detail, please see Section 2: Policy and Governance-Privacy Regulations.
- ³ The Genomics and Personalized Medicine Act of 2006 (S. 3822, 109th Cong. [2006]; subsequently reintroduced in the Senate in 2007 and in the House of Representatives in 2008 and 2010) encouraged accelerating genetics and genomics research and translating knowledge gained into clinical and public health applications. The proposed bill stated that “pharmaco-genetics has the potential to dramatically increase the efficacy and safety of drugs and reduce health care costs, and is fundamental to the practice of genome-based personalized medicine.” The Act identified several drugs that are more or less effective in people with particular genetic profiles, including the cancer drug Gleevac, the breast cancer drug Herceptin, and the acute lymphoblastic leukemia drug 6-mercaptopurine. All four bills were referred to committee and were not enacted.
- ⁴ American Medical Association (AMA). (2007). Personalized Health Care Report 2008: Warfarin and Genetic Testing. Retrieved from <http://www.ama-assn.org/ama1/pub/upload/mm/464/warfarin-brochure.pdf>.
- ⁵ Kolata, G. (2012, September 5). Bits of mystery DNA, far from ‘junk,’ play crucial role. *New York Times*, p. A1. Retrieved from http://www.nytimes.com/2012/09/06/science/far-from-junk-dna-dark-matter-proves-crucial-to-health.html?_r=2&hp; Young, E. (2012, September). ENCODE: The rough guide to the human genome. *Discover Magazine*. Retrieved from <http://blogs.discovermagazine.com/notrocketscience/2012/09/05/encode-the-rough-guide-to-the-human-genome/>.
- ⁶ McGuire, A.L., and J.R. Lupski. (2010). Personal genome research: What should the participant be told? *Trends in Genetics*, 26(5), 199-201.
- ⁷ Kolata, G., op cit.
- ⁸ Donley, G., Hull, S.C., and B.E. Berkman. (2012). Prenatal whole genome sequencing: Just because we can, should we? *Hastings Center Report*, 42(4), 28-40.
- ⁹ Cirulli, E.T., and D.B. Goldstein. (2010). Uncovering the roles of rare variants in common disease through whole-genome sequencing. *Nature Reviews Genetics*, 11, 415-425.
- ¹⁰ The National Commission for the Protection of Human Subjects of Biomedical and Behavioral Research. (1978). *The Belmont Report: Ethical Principles and Guidelines for the Protection of Human Subjects of Research*. Washington, DC: Department of Health, Education, and Welfare, DHEW Publication OS 78-0012. Retrieved from <http://www.hhs.gov/ohrp/human-subjects/guidance/belmont.html>; PCSBI. (2010, December). *New Directions: The Ethics of Synthetic Biology and Emerging Technologies*. Washington, DC: PCSBI.
- ¹¹ Pokorska-Bocci, A. (2010). Early examples of clinical use of whole-genome sequencing. Retrieved from <http://www.phgfoundation.org/news/5708/>.
- ¹² Beery, R., Mother of twins who benefitted from improved diagnosis gained by whole genome sequencing. (2012). The Beery Family Whole Genome Sequencing Success. Presentation to the Presidential Commission for the Study of Bioethical Issues (PCSBI), February 2, 2012. Retrieved from <http://bioethics.gov/cms/node/658>.
- ¹³ Topol, E. (2012). *The Creative Destruction of Medicine: How the Digital Revolution Will Create Better Health Care*. New York: Basic Books.
- ¹⁴ Fan, H.C., et al. (2012). Non-invasive prenatal measurement of the fetal genome. *Nature*, 487, 320-324; Motluk, A. (2012). Fetal genome deduced from parental DNA, *Nature News*, doi:10.1038.
- ¹⁵ Topol, E., op cit.
- ¹⁶ Donley, G., et al., op cit.
- ¹⁷ Geller, L.N., et al. (1996). Individual, family, and societal dimensions of genetic discrimination: A case study analysis. *Science and Engineering Ethics*, 2(1), 71-88; NOVA. (2012). *Cracking Your Genetic Code*, 46:30 et seq. Retrieved from <http://video.pbs.org/video/2215641935>; Suter, S., Professor of Law, George Washington University. (2012). How Technology is Changing Views of Privacy. Presentation to PCSBI, August 1, 2012. Retrieved from <http://bioethics.gov/cms/node/748>.
- ¹⁸ Allen, A.L. (2011). *Privacy Law and Society*, 2nd ed. St. Paul, MN: Thomson Reuters.
- ¹⁹ Ibid; DeCew, J. (2012). Privacy. In E.N. Zalta (ed.), *The Stanford Encyclopedia of Philosophy* (Fall 2012 Edition). Stanford, CA: The Metaphysics Research Lab, Stanford University. Retrieved from <http://plato.stanford.edu/archives/fall2012/entries/privacy/>.
- ²⁰ PCSBI. (2010, December). *New Directions: The Ethics of Synthetic Biology and Emerging Technologies*. Washington, DC: PCSBI.
- ²¹ National Commission for the Protection of Human Subjects of Biomedical and Behavioral Research. (1979). *The Belmont Report: Ethical Principles and Guidelines for the Protection of Human Subjects of Research*. Washington, DC: Department of Health, Education, and Welfare, DHEW Publication OS 78-0012. Retrieved from <http://www.hhs.gov/ohrp/humansubjects/guidance/belmont.html>.
- ²² PCSBI. (2010, December). *New Directions: The Ethics of Synthetic Biology and Emerging Technologies*. Washington, DC: PCSBI, pp. 24-25, 113.
- ²³ Ibid, pp. 25-27, 123-127.
- ²⁴ Ibid, pp. 27-28, 141-145.

²⁵ Ibid.

²⁶ Ibid.

²⁷ Ibid, pp. 28-30, 151.

²⁸ Ibid, pp. 30-31, 161-163.

²⁹ The Commission reached out to the heads of the 18 federal agencies and departments that have adopted the federal Common Rule for protecting human research participants. They are: Department of Agriculture; Department of Commerce; Department of Defense; Department of Education; Department of Energy; Department of Health and Human Services; Department of Homeland Security; Department of Housing and Urban Development; Department of Justice; Department of Transportation; Department of Veterans Affairs; Agency for International Development; Consumer Product Safety Commission; Environmental Protection Agency; National Aeronautics and Space Administration; National Science Foundation; Social Security Administration; and the Central Intelligence Agency.

³⁰ Request for Comments on Issues of Privacy and Access With Regard to Human Genome Sequence Data, 77 *Fed. Reg.* 18, 247 (March 27, 2012).

³¹ For example, the American College of Medical Genetics and Genomics is taking the lead in considering which results ought to be returned. A number of federal agencies have looked into direct-to-consumer testing, including the Secretary's Advisory Committee on Genetic Testing in its report "Enhancing the Oversight of Genetic Tests: Recommendations of the SACGT" (http://oba.od.nih.gov/oba/sacgt/reports/oversight_report.pdf); its successor, the Secretary's Advisory Committee on Genetics, Health, and Society in its report "U.S. System of Oversight of Genetic Testing: A Response to the Charge of the Secretary of Health and Human Services" (http://oba.od.nih.gov/oba/sacghs/reports/sacghs_oversight_report.pdf); and by the Government Accountability Office first in its 2006 report "Nutrigenetic Testing: Tests Purchased from Four Web Sites Mislead Consumers" (<http://www.gao.gov/new.items/d06977t.pdf>) and in its 2010 report "Misleading Test Results Are Further Complicated by Deceptive Marketing and Other Questionable Practices" (<http://www.gao.gov/products/GAO-10-847T>). The Food and Drug Administration has also considered the issue of direct-to-consumer testing, but has not yet published any formal recommendations. And the American Heart Association recently published a report entitled "Genetics and cardiovascular disease: a policy statement from the American Heart Association" that discussed gene patenting and insurance coverage of genetic tests (Ashley, E.A. et al. (2012). Genetics and cardiovascular disease: A policy statement from the American Heart Association. *Circulation*, 126).

³² Gutmann, A., Chair, PCSBI. (2012). How Technology is Changing Views of Privacy. Addressing PCSBI, August 1, 2012. Retrieved from <http://bioethics.gov/cms/node/748>.

³³ National Commission, op cit.

³⁴ PCSBI, (2010, December), op cit.

³⁵ Ibid, pp. 24-25, 113.

³⁶ The distribution of both scientific knowledge and subsequently of economic opportunity in the field of genome sequencing has been debated within the legal system. In a recent example, Myriad Genetics attempted to patent the *BRCA1* and *BRCA2* genes, which are associated with breast and ovarian cancer. Myriad Genetics' actions were contested by the American Civil Liberties Union, which argued that they were in violation of the First Amendment. In August 2012, a U.S. federal appeals court ruled that the company has the right to patent the two genes, stating that the patents would encourage innovation, but simultaneously denied the company's attempt to patent methods comparing or analyzing DNA sequences. Reuters. (2012, August 16). Court reaffirms right of myriad genetics to patent genes. *New York Times*. Retrieved from http://www.nytimes.com/2012/08/17/business/court-reaffirms-right-of-myriad-genetics-to-patent-genes.html?_r=1.

³⁷ Battelle. (n.d.). \$3.8 billion investment in Human Genome Project drove \$796 billion in economic impact creating 310,000 jobs and launching the genomic revolution. Retrieved from [http://www.battelle.org/media/news/2011/05/11/\\$3.8-billion-investment-in-human-genome-project-drove-\\$796-billion-in-economic-impactcreating-310-000-jobs-and-launching-the-genomic-revolution](http://www.battelle.org/media/news/2011/05/11/$3.8-billion-investment-in-human-genome-project-drove-$796-billion-in-economic-impactcreating-310-000-jobs-and-launching-the-genomic-revolution).

³⁸ PCSBI, (2010, December), op cit.

³⁹ Aristotle, *Politics*, (Benjamin Jowett trans., Dover first ed. 2000) (350 B.C.); DeCew, J.W. (1997). *In Pursuit of Privacy*. Ithaca, NY: Cornell Press, p.10.

⁴⁰ Warren, S.D., and L.D. Brandeis. (1890), *The Right to Privacy*, 4 Harv. L. Rev. 5.

⁴¹ Lowrance, W. (2012). *Privacy, Confidentiality, and Health Research*. New York: Cambridge University Press, p. 40-47.

⁴² Allen, A.L. (1997). Genetic privacy: Emerging concepts and values. In M. Rothstein (Ed.), *Genetic Secrets: Protecting Privacy and Confidentiality in the Genetic Era* (pp. 31-59). New Haven, CT: Yale University Press; Allen, A.L. (1999). *Coercing Privacy*, 40 Wm. & Mary L. Rev. 723 (1999); DeCew, J.W. (2004). Privacy and policy for genetic research. *Ethics and Information Technology*, 6, 5-14; Farahany, N.A. (2012). *Searching Secrets*, 160 U. Penn. L. Rev. 1239, 1255.

⁴³ Fried, C. (1970). *An Anatomy of Values*. Cambridge: Harvard University Press; *Moore v. Regents of the Univ. of Cal.*, 51 Cal. 3d 120 (Cal. 1990); Rachels, J. (1975). Why privacy is important. *Philosophy & Public Affairs*, 4(4), 323-333.

⁴⁴ Newcombe, H.B. (1994). Cohorts and privacy. *Cancer Causes and Control*, 5(3), 287-291.

- ⁴⁵ Allen, A.L. (2009). Confidentiality: An expectation in healthcare. In Ravitsky, V., Fiester, A., and A.L. Caplan (Eds.), *The Penn Center Guide to Bioethics*. (pp.127-135). New York: Springer Publishing Co.
- ⁴⁶ Goldstein, M., Associate Professor, Department of Health Policy, School of Public Health and Health Services, George Washington University, and McGraw, D., Director, Health Privacy Project, Center for Democracy and Technology. (2012). Comments submitted to PCSBI, May 25, 2012; Health Privacy Project. (2007). *Health Privacy Stories*. Retrieved from <https://www.cdt.org/healthprivacy/20080311stories.pdf>.
- ⁴⁷ Williams, S., et al. (2009). *The Genetic Town Hall: Public Opinion About Research on Genes, Environment, and Health*. Washington, DC: Genetics and Public Policy Center. Retrieved from http://www.dnapolicy.org/pub.reports.php?action=detail&report_id=27.
- ⁴⁸ Kaufman, D.J., et al. (2009). Public opinion about the importance of privacy in biobank research. *The American Journal of Human Genetics*, 85(5), 643-654.
- ⁴⁹ Genetics and Public Policy Center. (2007, April 24). U.S. Public Opinion on Uses of Genetic Information and Genetic Discrimination. Retrieved from http://www.dnapolicy.org/resources/GINAPublic_Opinion_Genetic_Information_Discrimination.pdf.
- ⁵⁰ McGuire, A.L., et al. (2008). DNA data sharing: Research participants' perspectives. *Genetics in Medicine*, 10(1), 46-53.
- ⁵¹ For a survey of self-reported genetic discrimination, see Geller, L.N., et al. (1996). Individual, family, and societal dimensions of genetic discrimination: A case study analysis. *Science and Engineering Ethics*, 2(1), 71-88; NOVA. (2012, March 28). *Cracking Your Genetic Code*, 46:30 et seq. Retrieved from <http://video.pbs.org/video/2215641935> (Discussing potential harms from whole genome sequencing at birth).
- ⁵² Powers, M., Professor, Department of Philosophy, Senior Research Scholar, Kennedy Institute of Ethics, Georgetown University. (2012). Theory and Practice of a Right to Privacy. Presentation to PCSBI, May 17, 2012. Retrieved from <http://bioethics.gov/cms/node/712>.
- ⁵³ Andrews, L.B. (2001). *Future Perfect*. New York: Columbia University Press.
- ⁵⁴ For example in *EEOC v. Burlington Northern and Santa Fe Railway*, the Federal District Court for the Northern District of Iowa addressed the scope of protection from genetic discrimination under the Americans with Disabilities Act. In this case, the railway required that employees who complained of carpal tunnel syndrome undergo genetic testing for a genetic predisposition. No. C01-4013 (N.D. Iowa Feb. 9, 2001).
- ⁵⁵ Sweeney, L., Visiting Professor and Scholar, Computer Science; Director, Data Privacy Lab, Harvard University. (2012). *How Technology is Changing Views of Privacy*. Presentation to PCSBI, August 1, 2012. Retrieved from <http://bioethics.gov/cms/node/748>.
- ⁵⁶ The *Belmont Report* recognizes that not all persons can act as autonomous agents, and makes clear that there are special responsibilities to those who cannot. National Commission for the Protection of Human Subjects of Biomedical and Behavioral Research. (1979). *The Belmont Report: Ethical Principles and Guidelines for the Protection of Human Subjects of Research*. Washington, DC: Department of Health, Education, and Welfare, DHEW Publication OS 78-0012. Retrieved from <http://www.hhs.gov/ohrp/humansubjects/guidance/belmont.html>.
- ⁵⁷ Brock, D. (1999). A critique of three objections to physician-assisted suicide. *Ethics*, 109(3), 523-524.
- ⁵⁸ PCSBI. (2010, December). *New Directions: The Ethics of Synthetic Biology and Emerging Technologies*. Washington, DC: PCSBI.
- ⁵⁹ *Ibid.*
- ⁶⁰ *Ibid.*, p. 31.
- ⁶¹ Sweeney, L., Visiting Professor and Scholar, Computer Science, Director, Data Privacy Lab, Harvard University. (2012). *How Technology is Changing Views of Privacy*. Presentation to PCSBI, August 1, 2012. Retrieved from <http://bioethics.gov/cms/node/748>.
- ⁶² PCSBI, (2010, December), *op cit*, p.5.
- ⁶³ *Ibid.*
- ⁶⁴ Gutmann, A., and D. Thompson. (1996). *Democracy and Disagreement*. Cambridge, MA: Harvard University Press.
- ⁶⁵ Trinidad, S.B., et al. (2010). Genomic research and wide data sharing: Views of prospective participants. *Genetics in Medicine*, 12(8), 486-495.
- ⁶⁶ Lake Research Partners and American Viewpoint. (2006). *National Survey on Electronic Personal Health Records*, conducted by the Markle Foundation. Retrieved from http://www.markle.org/sites/default/files/research_doc_120706.pdf.
- ⁶⁷ Gottweis, H., and K. Zatloukal. (2007). Biorepository governance: Trends and perspectives. *Pathobiology*, 74(4), 206-211.
- ⁶⁸ In the early 1990s, researchers obtained consent from the Havasupai Indian tribe to collect samples and conduct research on a genetic link to diabetes. The University of Arizona conducted the initial research, and later used the samples from the Havasupai to perform unrelated studies, including genetic and medical records analysis of inbreeding, schizophrenia, migration history, and genealogy. The Havasupai sued the University, alleging that it misused their genetic information, and that it conducted research for which it never obtained informed consent. The lawsuit resulted in a large settlement and destruction of the samples. Eriksson, S., and G. Helgesson. (2005).

Potential harms, anonymization, and the right to withdraw consent to biobank research. *European Journal of Human Genetics*, 13(9), 1071-1076; Harmon, A. (2010, April 21). Indian tribe wins fight to limit research of its DNA. *New York Times*. Retrieved from <http://www.nytimes.com/2010/04/22/us/22dna.html?pagewanted=all>.

- ⁶⁹ The Common Rule is a federal regulation regarding human subject research, adopted by 18 federal agencies; see footnote 29.
- ⁷⁰ Those agencies are: 1) USAID; 2) CIA; 3) CPSC; 4) USDA; 5) DOC; 6) DOE; 7) ED; 8) EPA; 9) HUD; 10) NASA; 11) NSF; and 12) SSA. SSA does receive genetic information that it considers personally identifiable information and deals with the information accordingly.
- ⁷¹ USDA indicated that its Agricultural Research Service expects to conduct human research programs in the future that may include more extensive use of whole genome analysis or sequencing.
- ⁷² Chief Information Officer, Office of Information Services, Centers for Medicare & Medicaid Services. (2007). Policy for Privacy Act Implementation & Breach Notification. Retrieved from <https://www.cms.gov/SystemLifecycleFramework/downloads/privacypolicy.pdf>; E-Government Act, 116 Stat. 2899 (2002); Federal Information Security Management Act (FISMA), 116 Stat. 2899 (2002); Health Information Technology for Economic and Clinical Health (HITECH), 123 Stat. 115 (2009); Health Insurance Portability and Accountability Act (HIPAA), 110 Stat. 1936 (1996); Policy for Privacy Act Implementation and Breach Notification, 5 U.S.C. § 552a.
- ⁷³ DOD. (2002). DOD Directive 6025.18, Privacy of Individually Identifiable Health Information in DOD Health Care Programs. December 19, 2002. Retrieved from <http://biotech.law.lsu.edu/blaw/dodd/corres/html/602518.htm>; DOD Regulation, DOD Directive 8580.02-R. (2007). DOD Health Information Security Regulation; DOD Regulation, DOD Directive 5400.11-R. (2007). Department of Defense Privacy Program.
- ⁷⁴ Letter from Earl C. Wyatt, Deputy Assistant Secretary of Defense, and Rapid Fielding, DOD to Amy Gutmann, Chair, PCSBI. (April 27, 2012).
- ⁷⁵ HITECH, Key provisions codified at 42 U.S.C. §§17931 et seq.; HIPAA, Key provisions codified at 42 U.S.C. §§ 1320d et seq.; FISMA, Key provisions codified at 44 U.S.C. §§ 3541 et seq. and 40 U.S.C. § 11331; Confidential Information Protection and Statistical Efficiency Act (CIPSEA), 44 U.S.C. § 3501 note, see also the OMB Implementation Guidance on CIPSEA, 72 *Fed. Reg.* § 33361 (2007); AHRQ general provisions, 42 U.S.C. § 299c-3(c); SAMHSA General Provisions, 42 U.S.C. § 290aa(n); Privacy Act of 1974, 5 U.S.C. § 552a; Confidentiality of information from health statistics and other activities, 42 U.S.C. § 242m(d).
- ⁷⁶ Letter from Kathleen Sebelius, Secretary, HHS to Amy Gutmann, Chair, PCSBI. (May 16, 2012).
- ⁷⁷ The Public Health Service Act, 42 U.S.C. §242m, part 301(d).
- ⁷⁸ Letter from Kathleen Sebelius, Secretary, HHS to Amy Gutmann, Chair, PCSBI. (May 16, 2012).
- ⁷⁹ *Ibid.*
- ⁸⁰ *Ibid.*
- ⁸¹ NIH. (2007). NIH Policy for Sharing of Data Obtained in NIH Supported or Conducted Genome-Wide Association Studies (GWAS). Retrieved from <http://grants.nih.gov/grants/guide/notice-files/NOT-OD-07088.html>; See updated data access policy at http://gwas.nih.gov/pdf/Data_Sharing_Policy_Modifications.pdf.
- ⁸² Federal Bureau of Investigation (FBI). (2012, June). CODIS-NDIS Statistics [web post]. Retrieved from <http://www.fbi.gov/about-us/lab/codis/ndis-statistics>.
- ⁸³ FBI. (n.d.). Frequently Asked Questions (FAQs) on the CODIS Program and the National DNA Index System [web post]. Retrieved from <http://www.fbi.gov/about-us/lab/codis/codis-and-ndis-fact-sheet/>.
- ⁸⁴ Collection and Use of DNA Identification Information from Certain Federal Offenders Act, 42 U.S.C. § 14135a; Collection of DNA Samples, 28 C.F.R. §28.12.
- ⁸⁵ VA. (2011). The Million Veteran Program: VA's Genomics Game-Changer Launches Nationwide [Press Release]. Retrieved from <http://www.va.gov/opa/pressrel/pressrelease.cfm?id=2090>.
- ⁸⁶ VA has departmental policies, including the Notice of Privacy Handbook Practice 1605.04 and Determining Service Connection for Congenital, Developmental, or Hereditary Disorders. See Letter from Robert A. Petzel, Under Secretary for Health, VA to Amy Gutmann, Chair, PCSBI. (May 1, 2012).
- ⁸⁷ VA. (2011). The Million Veteran Program: VA's Genomics Game-Changer Launches Nationwide [Press Release]. Retrieved from <http://www.va.gov/opa/pressrel/pressrelease.cfm?id=2090>.
- ⁸⁸ Letter from Judith S. Kaleta, Deputy General Counsel, DOT to Amy Gutmann, Chair, PCSBI. (August 1, 2012).
- ⁸⁹ Bregman-Eschet, Y. (2006). *Genetic Databases And Biobanks: Who Controls Our Genetic Privacy?* 23 Santa Clara Computer & High Tech. L.J. 1.
- ⁹⁰ 23andMe. (2012, April 10). What is the difference between genotyping and sequencing? [Frequently Asked Questions web post]. Retrieved from <https://customer.23andme.com/entries/21262606>.
- ⁹¹ US Patent Number 8,187,811 (filed Nov. 30, 2010).
- ⁹² Lowrance, W. (2012). *Privacy, Confidentiality, and Health Research*. New York: Cambridge University Press, p. 48.
- ⁹³ See e.g., Farahany, N.A. (2012). *Searching Secrets*, 160 U. Penn. L. Rev. 1277-83 (2012).
- ⁹⁴ Confidentiality and Disclosure of Returns and Return Information, 26 U.S.C. § 6103(d) et. seq.
- ⁹⁵ See e.g., Census Confidentiality Statute of 1954, 13 U.S.C. §9 (1954); HIPAA Privacy Rule, 45 C.F.R. § 164.514.

- ⁹⁶ Fair Credit Reporting Act, 15 U.S.C. § 1681 et seq.; Privacy Act of 1974, 5 U.S.C. § 552a; HHS. (2004). The Confidentiality of Alcohol And Drug Abuse Patient Records Regulation and the HIPAA Privacy Rule: Implications for Alcohol and Substance Abuse Programs. June. Retrieved from <http://www.nj.gov/humanservices/das/information/SAMHSA-Pt2-HIPAA.pdf>; Family Educational Rights and Privacy Act of 1974, 20 U.S.C. § 1232g; Electronic Communications Privacy Act of 1986, 18 U.S.C. § 2510-22; Video Privacy Act of 1988, 18 U.S.C. § 2710; Children's Online Privacy Protection Act of 1998, 15 U.S.C. §§ 6501–6506; and Gramm-Leach-Bliley Act, Pub. L. 106-102, 113 Stat. 1338.
- ⁹⁷ Department of Health, Education, and Welfare (DHEW). (1973). Records, Computers and the Rights of Citizens: Report of the Secretary's Advisory Committee on Automated Personal Data Systems. Retrieved from <http://aspe.hhs.gov/datacncl/1973privacy/tocprefacemembers.htm>. The Electronic Communications Privacy Act of 1986 is not a fair information statute, but rather a set of rules regulating government and other access to the many modes of communications used in daily life.
- ⁹⁸ DHEW. (1973). Records, Computers and the Rights of Citizens: Report of the Secretary's Advisory Committee on Automated Personal Data Systems. Retrieved from <http://aspe.hhs.gov/datacncl/1973privacy/tocprefacemembers.htm>.
- ⁹⁹ HIPAA, Pub. L. 104-191, 110 Stat. 1936, enacted August 21, 1996.
- ¹⁰⁰ HIPAA Privacy Rule, 45 C.F.R. § 164.501.
- ¹⁰¹ HIPAA Privacy Rule, 45 C.F.R. § 160, 164.
- ¹⁰² HIPAA Privacy Rule, 45 C.F.R. § 164.514.
- ¹⁰³ On the other hand, deceased individuals are not considered "human subjects" under, and are therefore not covered by, the Common Rule. 45 C.F.R. § 46.102(f).
- ¹⁰⁴ HIPAA Administrative Simplification: Standards for Privacy of Individually Identifiable Health Information, Proposed Rule, 74 *Fed. Reg.* 51698 (Oct. 7, 2009).
- ¹⁰⁵ Institute of Medicine (IOM). (2009). *Beyond the HIPAA Privacy Rule: Enhancing Privacy, Improving Health Through Research*. Washington, DC: National Academies Press.
- ¹⁰⁶ HIPAA Privacy Rule, 45 C.F.R. § 164.502.
- ¹⁰⁷ HIPAA Privacy Rule, 45 C.F.R. § 164.502(b).
- ¹⁰⁸ HIPAA Privacy Rule, 45 C.F.R. § 164.502.
- ¹⁰⁹ HITECH, 42 U.S.C. § 300jj.
- ¹¹⁰ The Office of the National Coordinator for Health Information Technology [web post]. (n.d.). Retrieved from http://healthit.hhs.gov/portal/server.pt/community/healthit_hhs_gov__onc/1200.
- ¹¹¹ Protection of Human Subjects, 45 C.F.R. § 46.
- ¹¹² Office for Human Research Protections (OHRP). (2008). OHRP Guidance on Research Involving Coded Private Information or Biological Specimens. October 16. Retrieved from <http://www.hhs.gov/ohrp/policy/cdebiol.html>.
- ¹¹³ Human Subject Research Protections: Enhancing Protections for Research Subjects and Reducing Burden, Delay, and Ambiguity for Investigators, 76 *Fed. Reg.* 44512, 44524.
- ¹¹⁴ *Ibid.*
- ¹¹⁵ See e.g., The Personal Information Protection and Electronic Documents Act, S.C. 2000, c. 5 (Canada, 2000); The Personal Data Act (523/1999) (Finland, 1999); The Federal Data Protection Act, BDSG 2003 (Germany, 2003); Act on the Protection and Processing of Personal Data, No. 77/2000 (Iceland, 2000); Personal Data Protection Code, Legisl. Ital. 196 (Italy, 2003); Data Protection Act, Cap. 440 (Malta, 2001); Personal Data Protection Act, 95/46/EC (Netherlands, 2001); Privacy Act, Public Act 1993 No 28 (New Zealand, 1993).
- ¹¹⁶ European Union, Directive 95/46/EC (1995).
- ¹¹⁷ See e.g., The Personal Information Protection and Electronic Documents Act, S.C. 2000, c. 5 (Canada, 2000); European Union, Directive 95/46/EC (1995).
- ¹¹⁸ See e.g., Patient Rights' Act, Law No. 482 (Denmark, 1998); Patients' Rights Act, No. 63 (Norway, 1999).
- ¹¹⁹ See e.g., Law on the Rights of Patients, (Belgium, 2002); Act on the Status and Rights of Patients, No.785/1992 (Finland, 1992); Patient's Rights Act (Romania, 1996).
- ¹²⁰ Law No. 20120 on scientific research on human beings, the human genome, and the prohibition of cloning (Chile, 2007); see also, Human Genome Research Act, RT I 2000, 104, 685 (Estonia, 2000); Rights of Users of Genetic Services Act (France, 2004); Law 14/2007 on Biomedical Research (Spain, 2007); Human Genetic Examination Act (Germany, 2009); Genetic Information Law, 5761-2000 (Israel, 2000).
- ¹²¹ See e.g., Law on the Rights of Patients, No. 283-IIS (Georgia, 2000); Human Genetic Examination Act (Germany, 2009).
- ¹²² Genetic Information Nondiscrimination Act (GINA), 122 Stat. 881-922 (2008).
- ¹²³ Kang, P.B. (2011). Presymptomatic and early symptomatic genetic testing. *Neurogenetics*, 2, 343-6.
- ¹²⁴ Kostecka, B.E. (2009). *GINA Will Protect You, Just Not From Death: The Genetic Information Nondiscrimination Act and Its Failure to Include Life Insurance within Its Protections*, 34 Seton Hall Legis. J. 93 (2009).
- ¹²⁵ HIPAA, 29 U.S.C. §§1181-82, 42 U.S.C. §§300gg-41.
- ¹²⁶ GINA, 122 Stat. 881-922 (2008).
- ¹²⁷ *Ibid.*

- ¹²⁸Rothstein, M.A. (n.d.). GINA's beauty is only skin deep. *Genewatch*. Retrieved from <http://www.councilforresponsiblegenetics.org/GeneWatch/GeneWatchPage.aspx?pageId=184>.
- ¹²⁹Kostecka, op cit.
- ¹³⁰See Appendix IV: State Law Tables. Also see Genome Statute and Legislation Database. (2010). Retrieved from <http://www.genome.gov/PolicyEthics/LegDatabase/pubsearch.cfm>
- ¹³¹See e.g., Del. Code Ann. §§ 12.2.1220 to 12.2.1227.
- ¹³²See e.g., Ariz. Rev. Stat. Ann. §12-2801-4, §20-448.02.
- ¹³³See e.g., Colo. Rev. Stat. 10-3-1104.7.
- ¹³⁴*Whalen v. Roe* 429 U.S. 589 (1977); *Sorrell v. IMS Health Inc.*, 131 S. Ct. 2653 (2011).
- ¹³⁵For a discussion on how Courts have generally disfavored a property-rights analysis in identifying information, see Farahany, op cit.
- ¹³⁶*Moore v. Regents of the Univ. of Cal.*, 51 Cal. 3d 120 (Cal. 1990).
- ¹³⁷Rodriguez, L.L., Director, Office of Policy, Communications, and Education, National Human Genome Research Institute, NIH. (2012). Presentation to PCSBI, August 1, 2012. Retrieved from <http://bioethics.gov/cms/node/749>.
- ¹³⁸Knoppers, B., Director, Centre of Genomics and Policy, Canada Research Chair in Law and Medicine, McGill University. (2012). Presentation to PCSBI, August 1, 2012. Retrieved from <http://bioethics.gov/cms/node/740>.
- ¹³⁹HITECH, Pub. L. 111-5, 123 Stat. 115 (2009); E-Government Act, Pub. L. 107-347, 116 Stat. 2899 (2002); HIPAA, Pub. L. 104-191, 110 Stat. 1936 (1996); Privacy Act, 5 U.S.C. § 552a; GINA, Pub. L. No. 110-233, 122 Stat. 881 (2008); PCSBI. (2012). Analysis of Responses to Common Rule Agency Data Call; See Appendix IV: U.S. State Genetic Laws.
- ¹⁴⁰An example of a third-party storage and analysis provider is cloud computing services, which include web-based systems of virtual servers.
- ¹⁴¹Rachels, J. (1975). Why privacy is important. *Philosophy & Public Affairs*, 4(4), 323-333.
- ¹⁴²Sweeney, L.S., Visiting Professor and Scholar, Computer Science; Director, Data Privacy Lab, Harvard University. (2012). How technology is changing views of privacy. Presentation to PCSBI, August 1, 2012. Retrieved from <http://bioethics.gov/cms/node/748>.
- ¹⁴³Harmon, A. (2010, April 21). Indian tribe wins fight to limit research of its DNA. *New York Times*. Retrieved from <http://www.nytimes.com/2010/04/22/us/22dna.html?pagewanted=all>.
- ¹⁴⁴*Bearder v. State*, 788 N.W.2d 144 (2010); Settlement Agreement and Release, *Beleno v. Tex. Dept. of State Health Servs.*, No. SA-09-CA-188-FB (W.D. Tex. 2009). Document obtained from plaintiffs' attorney, Jim Harrington, at the Texas Civil Rights Project.
- ¹⁴⁵NIH. (2008). Policy for Sharing of Data Obtained in NIH Supported or Conducted Genome-Wide Association Studies (GWAS). Retrieved from <http://grants.nih.gov/grants/guide/notice-files/NOT-OD-07088.html>.
- ¹⁴⁶GINA, Pub. L. No. 110-233, 122 Stat. 881 (2008).
- ¹⁴⁷Baruch, S., and K. Hudson. (2008). Civilian and military genetics: Nondiscrimination policy in a postGINA world. *American Journal of Human Genetics*, 83, 435-444; Greenbaum, D., et al. (2011). Genomics and privacy: Implications of the new reality of closed data for the field. *PLoS Computational Biology*, 7(12), e1002278; Hayden, E.C. (2012). A broken contract. *Nature*, 486, 312-314.
- ¹⁴⁸The Affordable Care Act of 2010 helps mitigate concerns about obtaining insurance with its prohibition on discriminating against individuals with pre-existing conditions.
- ¹⁴⁹Sweeney, L.S., Visiting Professor and Scholar, Computer Science Director, Data Privacy Lab, Harvard University. (2012). How Technology is Changing Views of Privacy. Presentation to PCSBI, August 1, 2012. Retrieved from <http://bioethics.gov/cms/node/748>.
- ¹⁵⁰Genetics and Public Policy Center. (2009, January 21). State laws pertaining to surreptitious DNA testing. Retrieved from http://www.dnapolicy.org/resources/State_law_summaries_final_all_states.pdf.
- ¹⁵¹Genomics Law Report. (2010). Surreptitious genetic testing: WikiLeaks highlights gap in genetic privacy law. Retrieved from <http://www.genomicslawreport.com/index.php/2010/12/09/surreptitious-genetic-testing-wikileaks-highlights-gap-in-genetic-privacy-law/>.
- ¹⁵²Lake Research Partners and American Viewpoint. (2006). National Survey on Electronic Personal Health Records, conducted by the Markle Foundation. Retrieved from http://www.markle.org/sites/default/files/research_doc_120706.pdf.
- ¹⁵³Aldhous, P., and M. Reilly. (2009). Special investigation: How my genome was hacked. *New Scientist*. Retrieved from <http://www.newscientist.com/article/mg20127013.800-special-investigation-how-my-genome-was-hacked.html?page=1>.
- ¹⁵⁴Green, R.C., and G.J. Annas. (2008). The genetic privacy of presidential candidates. *New England Journal of Medicine*, 359(21), 2192-2193.
- ¹⁵⁵HHS Health Information Privacy, HIPAA Breach Notification Rule, Breaches Affecting 500 or More Individuals. Retrieved from <http://www.hhs.gov/ocr/privacy/hipaa/administrative/breachnotificationrule/breachtool.html>.
- ¹⁵⁶Memorial Sloan-Kettering Cancer Center. (2012, June 15). Privacy Alert. Retrieved from <http://www.mskcc.org/public-notices/privacy-alert>.

- ¹⁵⁷ Sweeney, L.S., Visiting Professor and Scholar, Computer Science Director, Data Privacy Lab, Harvard University. (2012). How Technology is Changing Views of Privacy. Presentation to PCSBI, August 1, 2012. Retrieved from <http://bioethics.gov/cms/node/748>.
- ¹⁵⁸ Terry, S.F., and P.F. Terry. (2001). A consumer perspective on informed consent and third party issues. *Journal of Continuing Education in the Health Professions*, 21, 256-264.
- ¹⁵⁹ HHS Health Information Privacy, HIPAA Breach Notification Rule, Breaches Affecting 500 or More Individuals. Retrieved from <http://www.hhs.gov/ocr/privacy/hipaa/administrative/breachnotificationrule/breachtool.html>.
- ¹⁶⁰ Dissemination of Information (AHRQ), 42 USC §299C-3; General Provisions Respecting Effectiveness, Efficiency, and Quality of Health Services (CDC), 42 USC §242M; Justice System Improvement Administrative Provisions, 42 USCS § 3789g. Confidentiality of Information; The Public Health and Welfare Act, 42 U.S.C. § 46.
- ¹⁶¹ NIH. (2002). Notice NOT-OD-02-037, NIH Announces Statement on Certificates of Confidentiality. March 15. Retrieved from <http://grants.nih.gov/grants/guide/notice-files/NOT-OD-02-037.html>.
- ¹⁶² Cooper, Z.N., Nelson, R.M., and L.F. Ross. (2004). Certificates of confidentiality in research: Rationale and usage. *Genetic Testing*, 8(2), 214-220. This study sampled three NIH institutes: The National Human Genome Research Institute (NHGRI); the National Heart, Lung, and Blood Institute (NHLBI); and the National Institute of Neurological Disorders and Stroke (NINDS).
- ¹⁶³ Angrist, M. (2010). Urge overkill: Protecting deidentified human subjects at what price? *Health Privacy in Research*, 10(9), 17-18; Catania, J.A., et al. (2007). Research participants' perceptions of the certificate of confidentiality's assurances and limitations. *Journal of Empirical Research on Human Research Ethics*, 2(4), 53-59; Cooper, Z.N., Nelson, R.M., and L.F. Ross. (2004). Certificates of confidentiality in research: Rationale and usage. *Genetic Testing*, 8(2), 214-220; Hudson, K., and S. Devaney. (2008). Novel forensic technique highlights need for greater privacy protection for research participants [News release]. Retrieved from http://www.dnapolicy.org/news.release.php?action=detail&pressrelease_id=104; Lo, B., and M. Barnes. (2011). Protecting research participants while reducing regulatory burdens. *Journal of the American Medical Association*, 306(20), 2260-2261; Wolf, L.E., and J. Zandeki. (2006). Sleeping better at night: Investigators' experiences with certificates of confidentiality. *IRB*, 28(6), 1-7.
- ¹⁶⁴ Liang, B.A., and T. Mackey. (2011). Reforming direct-to-consumer advertising. *Nature Biotechnology*, 29(5), 397-400.
- ¹⁶⁵ Dissemination of Information (AHRQ), 42 USC §299c-3(c). The confidentiality statute that is part of AHRQ's authorizing legislation, grounded in judicially recognized public policies intended to foster participation in and the conduct of research, provides a respected form of federal statutory protection for all identifiable data submitted to the Agency, its grantees and contractors, for research purposes and permits no disclosures or uses of it, other than those consented to by the suppliers of the data or by the research subjects. Memorandum from Susan Greene Merewitz, Senior Attorney, AHRQ to Nancy Foster, Coordinator for Quality Activities, AHRQ. (April 16, 2001). Statutory Confidentiality Protection of Research Data. Retrieved from <http://www.ahrq.gov/fund/datamemo.htm>.
- ¹⁶⁶ Office of Science and Technology Policy (OSTP). (2012, March 29). Obama administration unveils "Big Data" initiative: Announces \$200 million in new R&D investments [Press release]. Retrieved from http://www.whitehouse.gov/sites/default/files/microsites/ostp/big_data_press_release_final_2.pdf.
- ¹⁶⁷ National Human Genome Research Institute (NHGRI). (2012). 1000 Genomes Project data available on Amazon Cloud [Press release]. Retrieved from <http://www.nih.gov/news/health/mar2012/nhgri-29.htm>.
- ¹⁶⁸ Breach Notification for Unsecured Protected Health Information, 74 *Fed. Reg.* 42,740 (Aug. 24, 2009) (codified at 45 C.F.R. §§ 160, 164); Health Breach Notification Rule; Final Rule, 74 *Fed. Reg.* 42,962 (Aug. 25, 2009) (codified at 16 C.F.R. § 318); HIPAA Administrative Simplification: Enforcement; Final Rule, 71 *Fed. Reg.* 8,390 (Feb. 16, 2006) (codified at 45 C.F.R. §§ 160, 164); Standards for Privacy of Individually Identifiable Health Information; Final Rule, 65 *Fed. Reg.* 82,462 (Dec. 28, 2000) (codified at 45 C.F.R. §§ 160, 164); Health Insurance Reform: Security Standards; Final Rule, 68 *Fed. Reg.* 8,334 (Feb. 20, 2003) (codified at 45 C.F.R. §§ 160, 162, and 164).
- ¹⁶⁹ HHS. (2010). HHS Strengthens Health Information Privacy and Security through New Rules [Press Release]. Retrieved from <http://www.hhs.gov/news/press/2010pres/07/20100708c.html>; National Institutes for Standards and Technology (NIST). (2012). Guidelines on Security and Privacy in Public Cloud Computing. Retrieved from <http://csrc.nist.gov/publications/nistpubs/800-144/SP800-144.pdf>.
- ¹⁷⁰ OHRP. (2008). Guidance on Research Involving Coded Private Information or Biological Specimens. October 16. Retrieved from <http://www.hhs.gov/ohrp/policy/cdebiol.html>.
- ¹⁷¹ Paasche-Orlow, M.K., Taylor, H.A., and F.L. Brancati. (2003). Readability standards for informed-consent forms as compared to actual readability. *New England Journal of Medicine*, 348(8), 721-726.
- ¹⁷² PRIM&R. (McGuire, A.L.) (2012). Data sharing in genomic research: Participant attitudes and ethical issues. [Webinar].
- ¹⁷³ Kaufman, D.J., et al. (2009). Public opinion about the importance of privacy in biobank research. *American Journal of Human Genetics*, 85, 643-654.

- ¹⁷⁴The Family Educational Rights and Privacy Act (FERPA), 20 U.S.C. § 1232g; 34 C.F.R. § 99; Children's Online Privacy Protection Act of 1998, 18 U.S.C. §§ 6501-6506; Parent Initiated Alternatives Act of 2005, HB10582005-06 (Washington); Hickey, K. (2007). Minors' rights in medical decision making. *Journal of Nursing Administration Health Care Law, Ethics, and Regulation*, 9(3), 100-104.
- ¹⁷⁵AMA. (1996). AMA Code of Medical Ethics, Opinion 2.138: Genetic testing of children. Retrieved from <http://www.ama-assn.org/ama/pub/physician-resources/medical-ethics/code-medical-ethics/opinion2138.page>; American Academy of Pediatrics. (2001). Ethical issues with genetic testing in pediatrics. *Pediatrics*, 107(6), 1451-1455; Donley, G., Hull, S.C., and B.E. Berkman. (2012). Prenatal Whole Genome Sequencing: Just Because We Can, Should We? *Hastings Center Report*, 42(4), 28-40.
- ¹⁷⁶Gostin, L.O. (2009). Privacy: Rethinking health information technology and informed consent. In M. Crowley (Ed.), *Connecting American Values with Health Reform*. Garrison, NY: The Hastings Center, pp. 15-17.
- ¹⁷⁷Kaye, J., et al. (2012). From patients to partners: Participant-centric initiatives in biomedical research. *Nature Reviews Genetics*, 13(5), 371-376; Nietfeld, J.J., Sugarman, J., and J.E. Litton. (2011). The Bio-PIN: A concept to improve biobanking. *Nature Reviews Cancer*, 11, 303-308; Time to open up. (2012). *Nature*, 486, 293.
- ¹⁷⁸Consent to Research (n.d.). About Us. Retrieved from <http://weconsent.us/about>.
- ¹⁷⁹Brothers, K.B., Morrison, D.R., and E.W. Clayton. (2011). Two large-scale surveys on community attitudes toward an opt-out biobank. *American Journal of Medical Genetics*, Part A, 155, 2982-2990.
- ¹⁸⁰Kaufman, D.J., et al. (2009). Public opinion about the importance of privacy in biobank research. *American Journal of Human Genetics*, 85, 643-654.
- ¹⁸¹Valle-Mansilla, J.I., Ruiz-Canela, M., and D.P. Sulmasy. (2010). Patients' attitudes to informed consent for genomic research with donated samples. *Cancer Investigation*, 28(7), 726-734.
- ¹⁸²Human Subjects Research Protections: Enhancing Protections for Research Subjects and Reducing Burden, Delay, and Ambiguity for Investigators, 76 *Fed. Reg.* 143, 44,513 (July 26, 2011).
- ¹⁸³Colditz, G.A. (2009). Constraints on data sharing: Experience from the Nurses' Health Study. *Epidemiology*, 20(2), 169-171.
- ¹⁸⁴One potential result of this sharing of data between physicians and researchers is the publication of findings in academic journals. In the interest of making data sharing an element of the scholarly publication process, some journals require that, if public repositories have been established for a particular type of data (including whole genome sequence data), all data from which results were drawn should be deposited in open access databases before publication.
- ¹⁸⁵Knoppers, B., Director, Centre of Genomics and Policy, Canada Research Chair in Law and Medicine, McGill University. (2012). Consent and Return of Findings. Presentation to PCSBI, August 1, 2012. Retrieved from <http://bioethics.gov/cms/node/740>.
- ¹⁸⁶Johnson, E.J., and D. Goldstein. (2003). Do defaults save lives? *Science*, 302, 1338-1339.
- ¹⁸⁷Donate Life America. (2012). 2012 National Donor Designation Report Card Released [Press Release]. Retrieved from <http://donatelife.net/2012-national-donor-designation-report-card-released/>; Goldman, R. (2012, May 2). States see instant spike in organ donors following Facebook push. *ABC News*. Retrieved from <http://abcnews.go.com/Health/states-instant-spike-organ-donors-facebook-push/story?id=16255979>.
- ¹⁸⁸Sunstein, C.R., and R.H. Thaler. (2003). Libertarian paternalism is not an oxymoron. *University of Chicago Law Review*. Retrieved from http://www.law.uchicago.edu/files/files/185.crs_paternalism.pdf; Thaler, R.H., and C.R. Sunstein. (2008). *Nudge: Improving Decisions About Health, Wealth, and Happiness*. New Haven, CT: Yale University Press.
- ¹⁸⁹Human Subjects Research Protections: Enhancing Protections for Research Subjects and Reducing Burden, Delay, and Ambiguity for Investigators, 76 *Fed. Reg.* 143, 44,513 (July 26, 2011).
- ¹⁹⁰McGuire, A.L., and L.M. Beskow. (2010). Informed consent in genomics and genetic research. *Annual Review of Genomics and Human Genetics*, 11, 361-381.
- ¹⁹¹Bollinger, J.M., et al. (2012). Public preferences regarding the return of individual genetic research results: Findings from a qualitative focus group study. *Genetics in Medicine*, 14(4), 451-457; Murphy, J., et al. (2008). Public expectations for return of results from large-cohort genetic research. *American Journal of Bioethics*, 8(11), 36-34; Terry, S.F., and P.F. Terry. (2011). Power to the people: Participant ownership of clinical trial data. *Science Translational Medicine*, 3(69), 1-3.
- ¹⁹²Murphy, J., et al. (2008). Public expectations for return of results from large-cohort genetic research. *American Journal of Bioethics*, 8(11), 36-43.
- ¹⁹³Maschke, K.J. (2012). Returning genetic research results: Considerations for existing no-return and future biobanks. *Minnesota Journal of Law, Science, and Technology*, 13(2), 559-573.
- ¹⁹⁴Clinical Laboratory Improvement Amendments of 1988 (CLIA), 42 U.S.C. § 263a (2006).
- ¹⁹⁵Sharp, S.E., and B.L. Elder. (2004). Competency assessment in the clinical microbiology laboratory. *Clinical Microbiology Review*, 17(3), 681-694.
- ¹⁹⁶CLIA, 42 U.S.C. § 263a (2006).
- ¹⁹⁷Hudson, K., et al. (2006). Oversight of U.S. genetic testing laboratories. *Nature Biotechnology*, 24(9), 1083-1091; Ledbetter, D.H., and W.A. Faucett. (2008). Issues in genetic testing for ultra-rare diseases: Background and introduction. *Genetics in Medicine*, 10(5), 309-313.

- ¹⁹⁸ Wolf, S.M. (2012). The past, present, and future of the debate over return of research results and incidental findings. *Genetics in Medicine*, 14(4), 355-357.
- ¹⁹⁹ Green, R., et al. (2012). Exploring concordance and discordance for return of incidental findings from clinical sequencing. *Genetics in Medicine*, 14(4), 1-6.
- ²⁰⁰ GARNET. (2012). GARNET Statement on Incidental Findings and Potentially Clinically Relevant Genetic Results. May 18. Retrieved from https://www.garnetstudy.org/sites/www/content/files/subcom/if/GARNET_RORstatement_final.pdf.
- ²⁰¹ Kohane, I.S., et al. (2007). Reestablishing the researcher-patient compact. *Science*, 316, 836-837.
- ²⁰² Wolf, S.M., et al. (2012). Managing incidental findings and research results in genomic research involving biobanks and archived data sets. *Genetics in Medicine*, 14(4), 361-384.
- ²⁰³ Berg, J.S., Khoury, M.J., and J.P. Evans. (2011). Deploying whole genome sequencing in clinical practice and public health: Meeting the challenge one bin at a time. *Genetics in Medicine*, 13(6), 499-504.
- ²⁰⁴ Kohane, I.S., and P.L. Taylor. (2010). Multidimensional results reporting to participants in genomic studies: Getting it right. *Science Translational Medicine*, 2(37), 1-4.
- ²⁰⁵ Angrist, M. (2011). You never call, you never write: Why return of 'omic' results to research participants is both a good idea and a moral imperative. *Personalized Medicine*, 8(6), 651-657; Kohane, I.S., et al. (2007), op cit.; Terry, S., and R. Cook-Deegan. (2012, June 8). Your genome belongs to you. *Health Affairs Blog*. Retrieved from <http://healthaffairs.org/blog/2012/06/08/your-genome-belongs-to-you/>; Time to open up. (2012). *Nature*, 486, 293.
- ²⁰⁶ Genomics Law Report. (2011, March 11). The FDA and DTC genetic testing: Setting the record straight. Retrieved from <http://www.genomicslawreport.com/index.php/2011/03/11/the-fda-and-dtc-genetic-testing-setting-the-record-straight/>.
- ²⁰⁷ Vorhaus, D., MacArthur, D., and L. Jostins. (2011, June 16). DTC genetic testing and the FDA: Is there an end in sight on regulatory uncertainty? [Blog]. Retrieved from <http://www.genomesunzipped.org/2011/06/dtc-genetic-testing-and-the-fda-is-there-an-end-in-sight-to-the-regulatory-uncertainty.php#more-3681>.
- ²⁰⁸ The Commission plans to look into incidental findings in a future report.
- ²⁰⁹ Beecher, H.K. (1966). Ethics and clinical research. *New England Journal of Medicine*, 274(24), 1354-1360; Jones, J.H. (1993) *Bad Blood: The Tuskegee Syphilis Experiment*. New York: The Free Press; Advisory Committee on Human Radiation Experiments (ACHRE). (1996). *Final Report of the Advisory Committee on Human Radiation Experiments*. New York: Oxford University Press.
- ²¹⁰ Friedman, C.P., Wong, A.K., and D. Blumenthal. (2010). Achieving a nationwide learning health system. *Science Translational Medicine*, 2(57), 1-3.
- ²¹¹ IOM. (2007). *The Learning Health Care System: Workshop Summary (IOM Roundtable on Evidence-Based Medicine)*. Washington, DC: The National Academies Press.
- ²¹² Kass, N., Faden, R., and S. Tunis. (2012). Addressing low-risk comparative effectiveness research in proposed changes to U.S. federal regulations governing research. *Journal of the American Medical Association*, 307(15), 1589-1590.
- ²¹³ Friedman, C.P., op cit.
- ²¹⁴ IOM. (2011). *Integrating Large-scale Genomic Information into Clinical Practice: Workshop Summary*. Washington, DC: National Academies Press; IOM. (2012). *Digital Data Priorities for Continuous Learning in Health and Health Care: An Institute of Medicine Workshop*. National Academies Press: Washington, DC; IOM. (2011). *Digital Infrastructure for the Learning Health System: The Foundation for Continuous Improvement in Health and Health Care*. Washington, DC: National Academies Press; Nass, S.J., Levit, L.A., and L.O. Gostin. (2009). *Beyond the HIPAA Privacy Rule: Enhancing Privacy, Improving Health Through Research*. Washington, DC: National Academies Press.
- ²¹⁵ President's Council of Advisors on Science and Technology. (2010). *Report to the President Realizing the Full Potential of Health Information Technology to Improve Health Care for Americans: The Path Forward*. Retrieved from <http://www.whitehouse.gov/sites/default/files/microsites/ostp/pcast-health-it-report.pdf>.
- ²¹⁶ CDC. (2012). Next generation sequencing: Standardization of clinical testing (Nex-StoCT) working group. Retrieved from http://www.cdc.gov/osels/lspppo/Genetic_Testing_Quality_Practices/Nex-StoCT.html.
- ²¹⁷ Kaye, J., Director of the Centre for Law, Health and Emerging Technologies, Oxford University. (2012). Privacy II – Control, Access and Human Genome Sequence Data. Presentation to PCSBI, February 2, 2012. Retrieved from <http://bioethics.gov/cms/node/659>.
- ²¹⁸ Kaye, J., et al. (2012). From patients to partners: Participant-centric initiatives in biomedical research. *Nature Reviews*, 13, 371-376.
- ²¹⁹ The Patient-Centered Outcomes Research Institute (PCORI). (2012). Retrieved from <http://www.pcori.org/>.
- ²²⁰ Wetterstrand, K.A. (n.d.). DNA Sequencing Costs: Data from the NHGRI Large-Scale Genome Sequencing Program Retrieved from www.genome.gov/sequencingcosts.
- ²²¹ Brothers K.B., Morrison D.R., and E.W. Clayton. (2011). Two large-scale surveys on community attitudes toward an opt-out biobank. *American Journal of Medical Genetics Part A*, 155, 2982-2990; Human Subjects Research Protections: Enhancing Protections for Research Subjects and Reducing Burden, Delay, and Ambiguity for

- Investigators, 76 *Fed. Reg.* 143, 44,513 (July 26, 2011); Kaye, J., et al., op cit; HITECH, Pub. L. 111-5, 123 Stat. 115 (2009).
- ²²² Battelle. (n.d.). \$3.8 billion investment in Human Genome Project drove \$796 billion in economic impact creating 310,000 jobs and launching the genomic revolution. Retrieved from [http://www.battelle.org/media/news/2011/05/11/\\$3.8-billion-investment-in-human-genome-project-drove-\\$796-billion-in-economic-impactcreating-310-000-jobs-and-launching-the-genomic-revolution](http://www.battelle.org/media/news/2011/05/11/$3.8-billion-investment-in-human-genome-project-drove-$796-billion-in-economic-impactcreating-310-000-jobs-and-launching-the-genomic-revolution).
- ²²³ Bustamante, D.D., Burchard, E.G., and F.M. De la Vega. (2011). Genomics for the world. *Nature*, 475, 163-165.
- ²²⁴ Length of a human DNA molecule. In *The Physics Factbook*. Elert, G. (Ed.). Retrieved from <http://hypertextbook.com/facts/1998/StevenChen.shtml>.
- ²²⁵ NIH. (2012). Clinical sequencing exploratory research coordinating center (U01). RFA-HG-12-008. Retrieved from <http://grants.nih.gov/grants/guide/rfa-files/RFA-HG-12-008.html>.
- ²²⁶ Gimelbrant, A., et al. (2007). Widespread monoallelic expression on human autosomes. *Science*, 318(5853), 1136-1140; Mardis, E.R. (2011). A decade's perspective on DNA sequencing technology. *Nature*, 470, 198-203; Lander, E. (2011). Initial impact of the sequencing of the human genome. *Nature*, 470, 187-197; Mason, C.E., and E. Onull. (2012). Faster sequencers, larger datasets, new challenges. *Genome Biology*, 13, 314.
- ²²⁷ Choia, M., et al. (2010). Genetic diagnosis by whole exome capture and massively parallel DNA sequencing. *PNAS*, 106, 19096-19101; Ng, S.B., et al. (2009). Targeted capture and massively parallel sequencing of 12 human exomes. *Nature*, 461, 272-276.
- ²²⁸ Ramos, E., and C. Rotimi. (2009). The A's, G's, C's, and T's of health disparities. *BMC Medical Genomics*, 2, 29.

Chapter 95

GENETIC TESTING: SCIENTIFIC BACKGROUND FOR POLICYMAKERS*

Amanda K. Sarata

ABSTRACT

Congress has considered, at various points in time, numerous pieces of legislation that relate to genetic and genomic technology and testing. These include bills addressing genetic discrimination in health insurance and employment; personalized medicine; the patenting of genetic material; and the quality of clinical laboratory tests, including genetic tests. The focus on these issues signals the growing importance of the public policy issues surrounding the clinical and public health implications of new genetic technology. As genetic technologies proliferate and are increasingly used to guide clinical treatment, these public policy issues are likely to continue to garner considerable attention. Understanding the basic scientific concepts underlying genetics and genetic testing may help facilitate the development of more effective public policy in this area.

Most diseases have a genetic component. Some diseases, such as Huntington's Disease, are caused by a specific gene. Other diseases, such as heart disease and cancer, are caused by a complex combination of genetic and environmental factors. For this reason, the public health burden of genetic disease is substantial, as is its clinical significance. Experts note that society has recently entered a transition period in which specific genetic knowledge is becoming critical to the delivery of effective health care for everyone. Therefore, the value of and role for genetic testing in clinical medicine is likely to increase significantly in the future.

INTRODUCTION

Virtually all disease has a genetic component.¹ The term "genetic disease" has traditionally been used to refer to rare monogenic (caused by a single gene) inherited disease, for example,

* This is an edited, reformatted and augmented version of a Congressional Research Service publication, CRS Report for Congress RL33832, prepared for Members and Committees of Congress, from www.crs.gov, dated December 19, 2011.

cystic fibrosis. However, we now know that many common complex human diseases, including common chronic conditions such as cancer, heart disease, and diabetes, are influenced by several genetic and environmental factors.² For this reason, they could all be said to be “genetic diseases.” Considering this broader definition of genetic disease, the public health burden of genetic disease can be seen to be substantial. In addition, an individual patient’s genetic make-up, and the genetic make-up of his disease, will help guide clinical decision making. Experts note that “(w)e have recently entered a transition period in which specific genetic knowledge is becoming critical to the delivery of effective health care for everyone.”³ This sentiment is still broadly shared, despite the fact that the translation to practice has perhaps been slower than anticipated due to the lack of a comprehensive evidence base to inform clinical validity and utility determinations for many genomic technologies. Experts in the field note that, “[d]espite dramatic advances in the molecular pathogenesis of disease, translation of basic biomedical research into safe and effective clinical applications remains a slow, expensive, and failure-prone endeavor.”⁴ Over time, as translational obstacles are addressed, the value of and role for genetic testing in clinical medicine is likely to increase significantly. As the role of genetics in clinical medicine and public health continues to grow, so too will the importance of public policy issues raised by genetic technologies.

Science is beginning to unlock the complex nature of the interaction between genes and the environment in common disease, and their respective contributions to the disease process. The information provided by the Human Genome Project is helping scientists and clinicians to identify common genetic variation that contributes to disease, primarily through genome-wide association studies (GWAS).⁵ However, researchers have identified a significant translational gap between genetic discoveries and application in clinical and public health practice and note that “the pace of implementation of genome-based applications in health care and population health has been slow.”⁶ Efforts are underway to close this gap and expedite translation into practice, specifically the recent development of the NIH-CDC collaborative Genomic Applications in Practice and Prevention Network.⁷ Experts note that the moderate effect of many common variants, uncovered by GWAS, has helped to underscore the multifactorial etiology of complex disease, and that substantially greater research efforts will be required to detect “missing” genetic influences.⁸ GWAS efforts have identified 1,100 well-validated genetic risk factors for common disease; however, the potential for many of these factors to serve as drug targets is unknown.⁹ In addition, research conducted utilizing large population databases that collect health, genetic, and environmental information about entire populations will likely provide more information about the genetic and environmental underpinnings of common diseases. Many countries have established such databases, including Iceland, the United Kingdom, and Estonia. The knowledge of the potential relevance of genetic information to the clinical management of nearly all patients coupled with the lack of complete information about the genetic and environmental factors underlying disease creates a challenging climate for public policymaking.

In many cases, the results of genetic testing may be used to guide clinical management of patients, and a particularly prominent role is anticipated in the realm of preventive medicine.¹⁰ For example, more frequent screening may be recommended for individuals at increased risk of certain diseases by virtue of their genetic make-up, such as colorectal and breast cancer. In some cases, prophylactic surgery may even be indicated. Decisions about courses of treatment and dosing may also be guided by genetic testing, as might reproductive decisions (both clinical and personal). However, many diseases with an identified molecular cause do not have any

treatment available; specifically, therapies exist only for approximately 200 of the more than 4,000 conditions with a known molecular cause.¹¹ In these cases, the benefits of genetic testing lie largely in the information they provide an individual about his or her risk of future disease or current disease status. The value of genetic information in these cases is personal to individuals, who may choose to utilize this information to help guide medical and other life decisions for themselves and their families. The information can affect decisions about reproduction; the types or amount of health, life, or disability insurance to purchase; or career and education choices. As genetic research continues to advance rapidly, it will often be the case that genetic testing may be able to provide information about the probability of a health outcome without an accompanying treatment option. This situation again creates unique public policy challenges, for example, in terms of decisions about the coverage of genetic testing services and education about the value of testing.

Policymakers may need to balance concerns about the potential use and misuse of genetic information, as well as issues of genetic exceptionalism¹² and genetic determinism¹³, with the potential of genetics and genetic technology to improve care delivery, for example by personalizing medical care and treatment of disease. In addition, policymakers face decisions about the extent of federal oversight and regulation of genetic tests, patients' safety, and innovation in this area. Finally, the need for and degree of federal support for research to develop a comprehensive evidence base to facilitate the integration of genetic testing into clinical practice (for example, to support coverage decisions by health insurers) may be debated. This report will summarize basic scientific concepts in genetics and will provide an overview of genetic tests, their main characteristics, and the key policy issues they raise.

FUNDAMENTAL CONCEPTS IN GENETICS

The following section explains key concepts in genetics that are essential for understanding genetic testing and issues associated with testing that are of interest to Congress.

Cells Contain Chromosomes

Humans have 23 pairs of chromosomes in the nucleus of most cells in their bodies. These include 22 pairs of autosomal chromosomes (numbered 1 through 22) and one pair of sex chromosomes (X and Y). One copy of each autosomal chromosome is inherited from the mother and from the father, and each parent contributes one sex chromosome.

Many syndromes involving abnormal human development result from abnormal numbers of chromosomes (such as Down Syndrome). Other diseases, such as leukemia, can be caused by breaks in or rearrangements of chromosome pieces.

Chromosomes Contain DNA

Chromosomes are composed of deoxyribonucleic acid (DNA) and protein. DNA is comprised of complex chemical substances called bases. Strands made up of combinations of

the four bases (adenine (A), guanine (G), cytosine (C) and thymine (T)) twist together to form a double helix (like a spiral staircase). Chromosomes contain almost 3 billion base pairs of DNA that code for about 20,000-25,000 genes (this is a current estimate, although it may change and has changed several times since the publication of the human genome sequence).¹⁴

DNA Codes for Protein

Proteins are fundamental components of all living cells. They include enzymes, structural elements, and hormones. Each protein is made up of a specific sequence of amino acids. This sequence of amino acids is determined by the specific order of bases in a section of DNA. A gene is the section of DNA which contains the sequence which corresponds to a specific protein. Changes to the DNA sequence, called mutations, can change the amino acid sequence. Thus, variations in DNA sequence can manifest as variations in the protein which may affect the function of the protein. This may result in, or contribute to the development of, a genetic disease.

Genotype Influences Phenotype

Though most of the genome is very similar between individuals, there can be significant variation in physical appearance or function between individuals. In other words, although we share most of the genetic material we have, we can see that there are significant differences in our physical appearance (height, weight, eye color, etc.). Humans inherit one copy (or allele) of most genes from each parent. The specific alleles that are present on a chromosome pair constitute a person's genotype. The actual observable, or measurable, physical trait is known as the phenotype. For example, having two brown-eye color alleles would be an example of a genotype and having brown eyes would be the phenotype.

Many complex factors affect how a genotype (DNA) translates to a phenotype (observable trait) in ways that are not yet clear for many traits or conditions. Study of a person's genotype may determine if a person has a mutation associated with a disease, but only observation of the phenotype can determine if that person actually has physical characteristics or symptoms of the disease. Generally, the risk of developing a disease caused by a single mutation can be more easily predicted than the risk of developing a complex disease caused by multiple mutations in multiple genes and environmental factors. Complex diseases, such as heart disease, cancer, immune disorders, or mental illness, for example, have both inherited and environmental components that are very difficult to separate. Thus, it can be difficult to determine whether an individual will develop symptoms, how severe the symptoms may be, or when they may appear.

GENETIC TESTS

What Is a Genetic Test?

Scientifically, a genetic test may be defined as:

an analysis performed on human DNA, RNA, genes, and/or chromosomes to detect heritable or acquired genotypes, mutations, phenotypes, or karyotypes that cause or are likely to cause a specific disease or condition. A genetic test also is the analysis of human proteins and certain metabolites, which are predominantly used to detect heritable or acquired genotypes, mutations, or phenotypes.¹⁵

Once the sequence of a gene is known, looking for specific changes is relatively straightforward using the modern techniques of molecular biology. In fact, these methods have become so advanced that hundreds or thousands of genetic variations can be detected simultaneously using a technology called a microarray.¹⁶

Policy Issues

The way genetic test is defined can be very important to the development of genetics-related public policy. For example, the above scientific definition is very broad, including both predictive and diagnostic tests and analyses on a broad range of material (nucleic acid, protein, and metabolites), but this may not be the best way to achieve certain policy goals. It may sometimes be desirable to limit the definition only to predictive, and not diagnostic, genetic testing because often, predictive tests raise public policy concerns that diagnostic tests do not (see “What Are the Different Types of Genetic Tests?”). In other cases, it may be desirable to limit the definition to only analysis of specific material, such as DNA, RNA, and chromosomes, but not metabolites or proteins, for example, to help avoid capturing certain types of tests, such as some newborn screening tests, in the scope of a proposed law. Policies extending protection against discrimination may aim to be as broad as possible, whereas policies addressing the stringency of oversight of genetic tests may aim to be more limited (to predictive or probabilistic tests only, or to those for conditions with no treatment, for example).

How Many Genetic Tests are Available?

As of December 2011, genetic tests are available for 2,492 diseases. Of those tests, 2,238 are available for clinical diagnosis, while 254 are available for research only.¹⁷ The majority of these tests are for single-gene rare diseases.

What Are the Different Types of Genetic Tests?

Most clinical genetic tests are for rare disorders, but increasingly, tests are becoming available to determine susceptibility to common, complex diseases and to predict response to medication.

With respect to health-related tests (i.e., excluding tests used for forensic purposes, such as “DNA fingerprinting”), there are two general types of genetic testing: diagnostic and predictive. Genetic tests can be utilized to identify the presence or absence of a disease (diagnostic). Predictive genetic tests can be used to predict if an individual will definitely get a disease in the future (presymptomatic) or to predict the risk of an individual getting a disease in the future (predispositional). For example, testing for mutations in the BRCA1 and/or BRCA2 genes provides probabilistic information about how likely an individual is to develop

breast cancer in his or her lifetime (predispositional). The genetic test for Huntington's Disease provides genetic information that is predictive in that it allows a physician to predict with certainty whether an individual will develop the disease, but does not allow the physician to determine when the onset of symptoms will actually occur (presymptomatic). In both of these examples, the individual does not have the clinical disease at the time of genetic testing, as they would with diagnostic genetic testing.

Within this broader framework of diagnostic and predictive genetic tests, several distinct types of genetic testing can be considered. Reproductive genetic testing can identify carriers of genetic disorders, establish prenatal diagnoses or prognoses, or identify genetic variation in embryos before they are used in *in vitro* fertilization. Reproductive genetic testing, such as prenatal testing, may be either diagnostic or predictive in nature. Newborn screening is a type of genetic testing that identifies newborns with certain metabolic or inherited conditions (although not all newborn screening tests are genetic tests). Traditionally, most newborn screening has been diagnostic, but some states have chosen to add certain predictive genetic testing to their newborn screening panels (for example, Maryland includes testing for cystic fibrosis).¹⁸ Finally, pharmacogenomic testing, or testing to determine a patient's likely response to a medication, may be considered either diagnostic or predictive, depending on the context in which it is being utilized (i.e., before administration of a medication to determine potential effectiveness, dosing levels, or potential adverse interactions or events vs. after administration and manifestation of a clinical event, for use in determining the basis of the specific event or outcome in the particular patient).

Policy Issues

Generally, predictive genetic testing (both presymptomatic and predispositional), rather than diagnostic testing, raises more complex ethical, legal, and social issues. For example, issues surrounding insurance coverage and reimbursement for this type of test, especially if no treatment is available, are more complex than with diagnostic genetic testing. A private insurer may feel that paying for a test that predicts the onset of a disease with no treatment is not cost-effective. Even more complicated are cases where the test only shows an increased probability of getting a disease.

Another issue is the oversight of genetic tests. Decisions about the need for oversight of genetic testing may be based on whether the information they provide is probabilistic rather than diagnostic, and whether a treatment is available. Additionally, stronger regulation of direct-to consumer marketing of genetic tests, or direct access testing,¹⁹ may be desirable in cases where a test is probabilistic rather than diagnostic.

Finally, issues of genetic discrimination may be different with predictive testing than they are with diagnostic testing. Genetic discrimination may be defined as differential treatment in either health insurance coverage or employment based upon an individual's genotype. Discriminatory action based on the possibility of something happening in the future, or even the certainty of it happening in the future, might raise more concern than would action taken based upon diagnostic information. With probabilistic genetic information (generated by predictive testing, see above), the health outcome at issue may never manifest, or if it is certain to, may not manifest for decades into the future. An individual's concern about the privacy of her genetic information may also be different if the information is probabilistic. For example, an individual who tests positive for being at increased risk of developing breast cancer in the future might believe unfavorable insurance or employment decisions based on this information

in the present (when she does not have breast cancer) would be unfair. If this were in fact her belief, this individual may have heightened concern with keeping this information private from health insurers or employers.

The Genetic Test Result

Genetic tests can provide information about both *inherited* genetic variations, that is, the individual's genes that were inherited from their mother and father, as well as about *acquired* genetic variations, such as those that cause some tumors. Acquired variations are not inherited, but rather are acquired in DNA due to replication errors or exposure to mutagenic chemicals and radiation (e.g., UV rays). In contrast with most other medical tests, genetic tests can be performed on material from a body, and may continue to provide information after the individual has died, as a result of the stability of the DNA molecule.

DNA-based testing of inherited genetic variations differs from other medical testing in several ways. These test results can have exceptionally long-range predictive powers over the lifespan of an individual; can predict disease or increased risk for disease in the absence of clinical signs or symptoms; can reveal the sharing of genetic variants within families at precise and calculable rates; and, at least theoretically, have the potential to generate a unique identifier profile for individuals.

Genetic changes to inherited genes can be acquired throughout a person's life (acquired genetic variation). Tests that are performed for acquired genetic variations that occur with a disease have implications only for individuals with the disease, and not family members. Tests for acquired genetic variations are also usually diagnostic rather than predictive, since these tests are generally performed after the presentation of symptoms.

Pharmacogenomic testing may be used to determine *both* acquired genetic variations in disease tissue (i.e., acquired variations in a tumor) or may be used to determine inherited variations in an individual's drug metabolizing enzymes. For example, with respect to determining acquired genetic variations in disease tissue, a tumor may have acquired genetic variations that render the tumor susceptible or resistant to chemotherapy. With respect to inherited genetic variation in drug metabolizing enzymes, an individual may, for example, be found to be a slow metabolizer of a certain type of drug (e.g., statins) and this information can be used to guide both drug choice and dosing.

Policy Issues

In some cases, people feel differently about their genetic information than they do about other medical information, a sentiment embodied by the concept of genetic exceptionalism. This viewpoint may be based on actual differences between genetic testing and other medical testing, but also may be based on personal belief that genetic information is powerful and different than other medical information. For example, genetic information about an individual may reveal things about family members, and therefore decisions by an individual to share her own genetic information can potentially also affect her family. Partially as a result of these considerations, the 110th Congress passed the Genetic Information Nondiscrimination Act of 2008 (P.L. 110-233), and many states, beginning in the early 1990s, enacted laws addressing genetic discrimination in health insurance, employment, and life insurance. Whether genetic

information is in fact different from other medical information and whether it deserves special protection are important public policy issues.²⁰

Pharmacogenomic testing is important because it will help provide the foundation for personalized medicine. Personalized medicine is healthcare based on individualized diagnosis and treatment for each patient determined by information at the genomic level. Many public policy issues are associated with personalized medicine. For example, there is some uncertainty currently as to how health insurers will assess and choose to cover pharmacogenomic testing as it becomes available. In addition, there are issues surrounding the regulation of pharmacogenomic testing and the encouragement of the co-development of drugs and diagnostic genetic tests (companion diagnostics). Companion diagnostics guide the use of the drug in a given individual.

Characteristics of Genetic Tests

Genetic tests function in two environments: the laboratory and the clinic. Genetic tests are evaluated based primarily on three characteristics: analytical validity, clinical validity, and clinical utility.

Analytical Validity

Analytical validity is defined as the ability of a test to detect or measure the analyte it is intended to detect or measure.²¹ This characteristic is critical for all clinical laboratory testing, not only genetic testing, as it provides information about the ability of the test to perform reliably at its most basic level. This characteristic is relevant to how well a test performs in a laboratory.

Clinical Validity

The clinical validity of a genetic test is its ability to accurately diagnose or predict the risk of a particular clinical outcome. A genetic test's clinical validity relies on an established connection between the DNA variant being tested for and a specific health outcome. Clinical validity is a measure of how well a test performs in a clinical rather than laboratory setting. Many measures are used to assess clinical validity, but the two of key importance are clinical sensitivity and positive predictive value. Genetic tests can be either diagnostic or predictive and, therefore, the measures used to assess the clinical validity of a genetic test must take this into consideration. For the purposes of a genetic test, positive predictive value can be defined as the probability that a person with a positive test result (i.e., the DNA variant tested for is present) either has or will develop the disease the test is designed to detect. Positive predictive value is the test measure most commonly used by physicians to gauge the usefulness of a test to clinical management of patients. Determining the positive predictive value of a predictive genetic test may be difficult because there are many different DNA variants and environmental modifiers that may affect the development of a disease. In other words, a DNA variant may have a known association with a specific health outcome, but it may not always be causal. Clinical sensitivity may be defined as the probability that people who have, or will develop a disease, are detected by the test.

Clinical Utility

Clinical utility takes into account the impact and usefulness of the test results to the individual and family and primarily considers the implications that the test results have for health outcomes (for example, is treatment or preventive care available for the disease). It also includes the utility of the test more broadly for society, and can encompass considerations of the psychological, social, and economic consequences of testing.

Policy Issues

These three above-mentioned characteristics of genetic tests, or analytic validity, clinical validity, and clinical utility, have important ties to public policy issues. For example, although the analytical validity of genetic tests is regulated by the Centers for Medicare and Medicaid Services (CMS) through the Clinical Laboratory Improvement Amendments (CLIA) of 1988 (P.L. 100- 578), the clinical validity of the majority of genetic tests is not regulated at all. This has raised concerns about direct-to-consumer marketing of genetic tests where the connection between a DNA variant and a clinical outcome has not been clearly established. Marketing of such tests to consumers directly may mislead consumers into believing that the advice given them based on the results of such tests could improve their health status or outcomes when in fact there is no scientific basis underlying such an assertion. This issue was the subject of a July 2006 hearing by the Senate Special Committee on Aging. In addition, clinical utility and clinical validity both figure prominently into coverage decisions by payers, but a lack of data often hinders coverage decisions, potentially leaving patients without coverage for these tests.

Coverage by Health Insurers

Health insurers are playing an increasingly large role in determining the availability of genetic tests by deciding which tests they will pay for as part of their covered benefit packages. Many aspects of genetic tests, including their clinical validity and utility, may complicate the coverage decision-making process for insurers. While insurers require that, where applicable, a test be approved by the Food and Drug Administration (FDA), they also want evidence that it is “medically necessary;” that is, evidence demonstrating that a test will affect a patient’s health outcome in a positive way. This additional requirement of evidence of improved health outcomes underscores the importance of patient participation in long-term research in genetic medicine. Particularly for genetic tests, data on health outcomes may take a very long time to collect.

Policy Issues

Decisions by insurers to cover new genetic tests have a significant impact on the utilization of such tests and their eventual integration into the health care system. The integration of personalized medicine into the health care system will be significantly affected by coverage decisions. Although insurers are beginning to cover pharmacogenomic tests and treatments, the high cost of such tests and treatments often means that insurers require stringent evidence that they will improve health outcomes. As mentioned previously, this evidence is often lacking.

In addition, coverage of many genetic tests and services, which may be considered preventive in some cases, might be affected by the passage of the Patient Protection and

Affordable Care Act of 2010 (ACA, P.L. 111-148). The ACA requires private health insurers, Medicare, and Medicaid to cover clinical preventive services (as specified in the law) and outlines cost-sharing requirements in some cases for these services.²² However, the ACA provisions in some cases tie coverage of clinical preventive services to determinations by the U.S. Preventive Services Task Force (USPSTF, located in the Agency for Healthcare Research and Quality), and these determinations are based on the quality of the evidence available to support a given clinical preventive service. For this reason, coverage of genetic tests and services (that are determined to be preventive clinical services) might be negatively affected by a lack of high-quality evidence to support their use.

Regulation of Genetic Tests by the Federal Government

Genetic tests are regulated by the Food and Drug Administration (FDA) and CMS, through CLIA. FDA regulates genetic tests that are manufactured by industry and sold for clinical diagnostic use. These test kits usually come prepackaged with all of the reagents and instructions that a laboratory needs to perform the test and are considered to be products by the FDA. FDA requires manufacturers of the kits to ensure that the test detects what the manufacturer says it will, in the intended patient population. With respect to the characteristics of a genetic test, this process requires manufacturers to prove that their test is clinically valid. Depending on the perceived risk associated with the intended use promoted by the manufacturer, the manufacturer must determine that the genetic test is safe and effective, or that it is substantially equivalent to something that is already on the market that has the same intended use.

Most genetic tests, however, are performed not with test kits, but rather as laboratory testing services (referred to as either laboratory-developed or “homebrew” tests), meaning that clinical laboratories themselves perform the test in-house and make most or all of the reagents used in the tests. Laboratory-developed tests (LDTs) are not currently regulated by the FDA in the way that test kits are and, therefore, the clinical validity of the majority of genetic tests is not regulated. The FDA does regulate certain components used in LDTs, known as Analyte Specific Reagents (ASRs), but only if the ASR is commercially available. If the ASR is made in-house by a laboratory performing the LDT, the test is not regulated at all by the FDA. This type of test is sometimes referred to informally as a “homebrew-homebrew” test.

Any clinical laboratory test that is performed with results returned to the patient must be performed in a CLIA-certified laboratory. CLIA is primarily administered by CMS in conjunction with the Centers for Disease Control and Prevention (CDC) and the FDA.²³ FDA determines the category of complexity of the test so the laboratories know which requirements of CLIA they must follow. As previously noted, CLIA regulates the analytical validity of a clinical laboratory test only. It generally establishes requirements for laboratory processes, such as personnel training and quality control or quality assurance programs. CLIA requires laboratories to prove that their tests work properly, to maintain the appropriate documentation, and to show that tests are interpreted by laboratory professionals with the appropriate training. However, CLIA does not require that tests made by laboratories undergo any review by an outside agency to see if they work properly. Supporters of the CLIA regulatory process argue that regulation of the testing process gives the laboratories optimal flexibility to modify tests

as new information becomes available. Critics argue that CLIA does not go far enough to assure the accuracy of genetic tests since it only addresses analytical validity and not clinical validity.

End Notes

- ¹ Collins, F.S. and V.A. McCusick. (2001) "Implications of the Human Genome Project for Medical Science." *Journal of the American Medical Association* 285:540-544.
- ² Manolio, T.A. et al. (2009) "Finding the missing heritability of complex diseases." *Nature* 461(8): 747-753.
- ³ Guttmacher, A.E. and F.S. Collins. (2002) "Genomic Medicine - A Primer." *New England Journal of Medicine* 347(19): 1512-1520.
- ⁴ Collins F.S. (2011) "Reengineering Translational Science: The Time Is Right." *Sci Transl Med.* 3(90):90cm17.
- ⁵ Genome-wide association studies are defined by the National Human Genome Research Institute as "...an approach used in genetics research to associate specific genetic variations with particular diseases. The method involves scanning the genomes from many different people and looking for genetic markers that can be used to predict the presence of a disease." <http://www.genome.gov/glossary/index.cfm?id=91>
- ⁶ Khoury M.J. et al. (2009) "The Genomic Applications in Practice and Prevention Network." *Genetics in Medicine* 11(7): 488-494.
- ⁷ For more information about the Genomic Applications in Practice and Prevention Network, see <http://www.cdc.gov/genomics/translation/GAPPNet/index.htm>.
- ⁸ See note 2 at page 751.
- ⁹ Collins F.S. (2011) "Reengineering Translational Science: The Time Is Right." *Sci Transl Med.* 3(90):90cm17.
- ¹⁰ Collins F.S. (2010) "Opportunities for Research and NIH." *Science* 327: 36-37.
- ¹¹ Collins F.S. (2011) "Reengineering Translational Science: The Time Is Right." *Sci Transl Med.* 3(90):90cm17.
- ¹² Genetic exceptionalism is the concept that genetic information is inherently unique, should receive special consideration, and should be treated differently from other medical information. For more information about genetic exceptionalism in public policy, see CRS Report RL34376, Genetic Exceptionalism: Genetic Information and Public Policy, by Amanda K. Sarata.
- ¹³ Genetic determinism is the concept that our genes are our destiny and that they solely determine our behavioral and physical characteristics. This concept has mostly fallen out of favor as the substantial role of the environment in determining characteristics has been recognized.
- ¹⁴ National Research Council, *Reaping the Benefits of Genomic and Proteomic Research: Intellectual Property Rights, Innovation, and Public Health*. Washington, DC: National Academies Press (2006); p. 19. The National Human Genome Research Institute at the National Institutes of Health reports that the estimated number of human genes is closer to 25,000. See <http://www.genome.gov/11508982>.
- ¹⁵ Report of the Secretary's Advisory Committee on Genetic Testing (SACGT), "Enhancing the Oversight of Genetic Tests: Recommendations of the SACGT," July 2000, at http://oba.od.nih.gov/oba/sacgt/reports/oversight_report.pdf.
- ¹⁶ Microarray technology is defined as "...a developing technology used to study the expression of many genes at once. It involves placing thousands of gene sequences in known locations on a glass slide called a gene chip. A sample containing DNA or RNA is placed in contact with the gene chip. Complementary base pairing between the sample and the gene sequences on the chip produces light that is measured. Areas on the chip producing light identify genes that are expressed in the sample." See <http://ghr.nlm.nih.gov/glossary/microarraytechnology>.
- ¹⁷ See <http://www.genetests.org> for information on disease reviews, an international directory of genetic testing laboratories, an international directory of genetics and prenatal diagnosis clinics, and a glossary of medical genetics terms.
- ¹⁸ Newborn Screening Home, Maryland Department of Health and Mental Hygiene. <http://dhmh.maryland.gov/labs/html/nbs.html>.
- ¹⁹ For more information about direct-to-consumer genetic testing, see <http://ghr.nlm.nih.gov/handbook/testing/directtoconsumer>.
- ²⁰ For more information about characteristics of genetic information that may be viewed as unique and public perspectives on the differences between genetic and other medical information, see CRS Report RL34376, Genetic Exceptionalism: Genetic Information and Public Policy, by Amanda K. Sarata.
- ²¹ An analyte is defined as a substance or chemical constituent undergoing analysis.

²² For more information about requirements relating to the coverage of clinical preventive services under the ACA, see CRS Report R41278, Public Health, Workforce, Quality, and Related Provisions in PPACA: Summary and Timeline, coordinated by C. Stephen Redhead and Erin D. Williams. ²³ See <http://www.cms.hhs.gov/CLIA/>.

Chapter 96

EVOLUTION OF HUMAN GENOME ANALYSIS: IMPACT ON DISEASES DIAGNOSIS AND MOLECULAR DIAGNOSTIC LABS

***Julie Gauthier^{1,*}, Isabelle Thiffault¹,
Virginie Dormoy-Raclet¹ and Guy A. Rouleau^{1,2}***

¹Medical Biological Unit, Molecular Diagnostic Laboratory, Sainte-Justine University
Hospital Center, Montreal, QC, Canada

²Montreal Neurological Institute and Hospital,
Montréal (Québec), Canada

ABSTRACT

The advent of massively parallel sequencing has changed the interrogation process of the human genome and now provides a high resolution and global view of the genome which is beyond research applications. Together with powerful bioinformatics tools, these next generation sequencing technologies have revolutionized fundamental research and have important consequences for clinically actionable tests, diagnosis and treatment of rare diseases and cancers. Today, molecular testing is commonly used to confirm clinical diagnosis of specific diseases; it requires that a clinician specify the gene or mutation to test and, in return, will receive information only about this sequence. Despite relative successes, a large number of patients receive no accurate diagnosis, even after many expensive molecular investigations. A clear paradigm shift has taken place in the health network with the introduction of the exome sequencing in molecular diagnostic lab. In this chapter, the impact of the implementation of high throughput sequencing technologies on molecular diagnosis and on the practice of medicine, with an emphasis in paediatrics, is reviewed. We compared well-established genetic tests, using examples from our molecular diagnostic lab, to the recent exome sequencing applications. The genetic tests can fall into three main categories: 1) Mendelian Single Gene Disorder tests that include targeted mutation and targeted gene approaches 2) Genetic Disease Panels which are composed of a few to a dozen genes and 3) Exome or Genome approaches, which interrogate either the

* Corresponding Author's Email: julie.gauthier@recherche-ste-justine.qc.ca.

entire coding sequences of the 22,333 human genes or the entire human genome. For each of these categories, advantages and limitations are discussed. We devoted a section on the future of molecular diagnosis and discuss which tests will subsist and which one may be soon abandoned. Massively parallel sequencing is transforming the molecular diagnostic field: it offers personalized genetic tests and generates new ethical challenges. Important questions like incidental findings and possible forms of discrimination are addressed. Finally, we conclude with a section on the future directions surrounding the application of these multimodal molecular approaches in general and their putative applications in neonatal intensive care units.

1. INCREASED HUMAN GENOME KNOWLEDGE THROUGH PROGRESS IN SEQUENCING TECHNOLOGY

Sequencing the DNA of an organism is feasible since 1977 when the "dideoxy" chain-termination method for sequencing DNA molecules was introduced by Frederick Sanger and colleagues [1]. This technique allows for the sequencing of a single DNA fragment up to 1000 base pairs long. The original method used radiolabeled dNTP and the reading was done manually. Sanger sequencing has been the basis of several major gene discoveries transforming the field of molecular biology. For example, before DNA sequencing, proteins were sequenced directly, this is a laborious technique. Now, this is easily accomplished by sequencing a cDNA and translating the DNA sequence into the amino acid sequence of the protein. In the early 1990s, DNA sequencing was automated using a 4-channel capillary approach (basically the current Life Technologies ABI 3730 system) and ddNTP labeled with different color fluorescent dyes which allow fast DNA sequencing, with hundreds of fragments simultaneously. These sequencing systems were the first generation of sequencing technologies. Their high-throughput yield allows a single lab to sequence millions of base pairs, compared to thousands before their introduction. Commonly called the Sanger method, this sequencing technique is the gold standard in research and clinical diagnostic laboratories for genetic testing. This advance in technology led to science's greatest achievement, the Human Genome Project. This project aimed to determine the sequence of the 3 billion nucleotide base pairs that constitute the human genome and to identify all 22,333 genes [2] in human DNA [3]. This 13-year project gave rise to the first draft of the human genome sequence in 2001. Decoding the human genome has opened unprecedented new avenues in research and increased our understanding of human health. Information from the human genome sequence enables us to understand how the genetic information drives the development and function of the human body. More importantly, the Human Genome Project accelerated the exploration of genetic variations predisposing to disease and thus revolutionizes medical practice and biological research. Since then, the availability of genomic information (gene and chromosome structure, polymorphisms, disease causing mutations, etc.), has continuously increased.

Sanger sequencing helped the discovery of new genetic mechanisms. For example, by sequencing more than 400 synaptic genes in affected probands with sporadic schizophrenia or autism (that is, with no history of psychiatric disorders in the parents or the extended family), we and others have observed a significant excess of potentially deleterious *de novo* mutations in affected individuals [4]. A similar observation was done in intellectual disability cases [5].

The most recent DNA sequencing development is the advent of massively parallel sequencing platforms leading to the “next generation sequencing” technologies. The development of ultra-high throughput sequencing technologies pushed forward at an unprecedented speed the identification of DNA variations associated with diseases. Ultrasequencing platforms can be combined with microarray technologies to capture and amplify the functional genome that encodes proteins, also known as the “exome”, or to capture more comprehensively a subset of related genes. As technology progressed, next-generation DNA sequencing techniques have improved significantly and are now much faster, and to some extent less expensive than Sanger sequencing. What took over 10 years like the Human Genome Project can now be done in less than a month. It’s now feasible to decode multiple human genomes at once. All of these progresses gave birth to larger scale genetics.

A remarkable increase of genetic and genomic data occurred in the last decade notably through the introduction of microarrays and next-generation sequencing. In fact, human genome array-comparative genomic hybridization and SNP arrays were developed to detect genetic aberrations or copy number variants (CNVs) at a higher resolution than traditional karyotyping. Comparative genomic hybridization, also called array CGH, is now currently used in diagnostic labs to identify small chromosomal deletions and duplicated regions. For example, conventional karyotyping has a resolution of 5-10 million bases compare to 100 000 for array-CGH. Therefore, submicroscopic chromosomal alterations can now be detected. These technological progresses lead researchers to major discoveries in the understanding of genetic mechanisms of different neuropsychiatric conditions such as intellectual disabilities, autism and schizophrenia. De Vries et al. used array based comparative genomic hybridization (array CGH) to study 100 patients with unexplained intellectual disability. They were able to detect a potential causative chromosomal anomaly in 10% of their patients which represents a diagnostic yield of at least twice as high as that of conventional method [7]. In addition to the identification of the causative genetic factors implicated in pediatric and childhood complex disorders, microarrays initiated the discovery of new genetic mechanisms for a number of complex disorders, namely autism, intellectual disability and schizophrenia. Using these technologies, studies have highlighted the involvement of rare (<1% frequency) point mutations and CNVs in the genetic etiology of autism, schizophrenia, intellectual disability, attention deficit disorder and other disorders [6-9].

Molecular diagnostics has become an essential tool for the evaluation and management of patients, particularly of children with genetic diseases. According to the Online Mendelian Inheritance in Man (OMIM), an online catalog of human genes and genetic disorders, more than 3,807 diseases (the majority >2,500 being rare) have been characterized at the molecular level, and for over 3,500 of these the genetic causes are still unknown. Some genetic conditions are related to a single gene (e.g., cystic fibrosis), but most genetic diseases are characterized by a great heterogeneity, each of which can be caused by mutations in one of several genes (Intellectual disability > 300 genes, Deafness > 60 genes, Mitochondrial diseases >300 genes). The current molecular diagnosis of clinical phenotypic heterogeneity in genetic conditions is laborious and expensive because it involves the sequential analysis of several genes.

2. THE GENETIC DIAGNOSIS: A MULTI TECHNICAL APPROACH FOR NEURODEVELOPMENTAL AND METABOLIC DISORDERS

The advance of modern genomics has changed the health economic decisions concerning genetic screening, with costs- per-megabase for DNA sequencing falling at a faster rate than predicted. Although individually rare, Mendelian diseases collectively account for a significant percentage of infant mortality and pediatric hospitalizations [10]. The speed with which genomics is becoming clinically relevant and the increasing power of the new sequencing technologies led to its rapid implementation in clinical settings. Assuming that new technology automatically translates into improved patient care is idealistic.

The ability to amplify DNA by polymerase chain reaction (PCR) has revolutionized our ability to test for genetic mutations. Many different assay systems are based on PCR for analyzing the amplified DNA or RNA. These techniques are sensitive, reliable and can be performed easily on different material: blood, skin or muscle biopsies, tumour, foetus, and even single cells from blastomeres and polar bodies. The aims of adapting new PCR-based strategies in clinical settings are improving the accuracy, speeding up the diagnosis, the time-consuming and costing without decreasing the sensitivity or specificity. Understanding the type of mutations, the incidence of *de novo* mutations or genetic rearrangements is essential to accurately select the best molecular technique for the detection of carrier in pre-natal diagnosis. Techniques involving PCR-based amplification are successfully used for mutation screening in the majority of diseases. It can be combined for the detection of the presence or absence of restriction sites, electrophoretic mobility shift or sizing analysis, as in single strand conformation polymorphism (SSCP) or in denaturing gradient gel electrophoresis (DGGE). Computer-assisted highly sensitive mutation detection is also performed, for the above techniques, by means of fluorescent PCR for allelic specific discrimination. Recent advances in the development of quantitative real-time PCR (qPCR)-based diagnostic tools allow detection and quantification of gene or exon dosage.

An alternative procedure to mutation-directed protocols for complex genetic mutations is the use of fluorescent multiplex PCR. Indirect diagnosis performed by the use of polymorphic microsatellite markers, allowing identification of the pathogenic haplotype instead of the mutation [11, 12] as for instance in some spinal muscular atrophy or Duchenne muscular dystrophy families. For diseases involving a heterogeneous spectrum of identified mutations, such as cystic fibrosis, autosomal non-syndromic intellectual disability (*SYNGAP1*) or Rett Syndrome (*MECP2*), the development of a mutation-based strategy is not practical and sequencing of the entire coding sequence is recommended to facilitate mutation detection.

Multiplex ligation-dependent probe amplification (MLPA) is a variation of the multiplex polymerase chain reaction that permits multiple targets (exons) to be amplified with only a single primer pair [13]. Specific fluorescent probes consist of two unique oligonucleotides which recognise adjacent target sites on the DNA and each amplicon generates a fluorescent peak which can be detected by a capillary sequencer. Comparing the peak pattern obtained on a given sample with those obtained on various control samples, allow the relative quantity of each target to be determined. This technique is commonly used to detect chromosomal anomalies in cell and tissue samples, [14] detection of gene copy number [13], detection of duplications and deletions in disease-related genes such as *DMD*, *MECP2*, *BRCA1*, *BRCA2*, etc. It has replaced the need to use of southern blot in many diseases [15-17].

Expansion repeat disorders such as of the glutamine codon (CAG) in spinocerebellar ataxias, or the trinucleotide repeat in non-coding regions, as the CGG in Fragile-X syndrome, the GAA in Friedreich ataxia or the CTG in myotonic dystrophy type I, are a special type of mutation. The unstable dynamic nature of those mutations requires the use of a multimodal approach [18]. Molecular tests for the diagnosis and carrier detection often combine PCR for the repeat expansion and triplet repeat primed PCR (TP-PCR) analysis of genomic DNA, as well as southern blot.

The detection rate of the screening test, as well as the frequency of the mutation in the study population and the technical limitations of the procedure, will determine the usefulness of a positive or negative result. Preimplantation testing provides a paradigm for the ease of use of PCR-based testing, yet also underscores the problems encountered with genetic screening because of the multitude of possible mutations and the possible misinterpretation of results [12].

Molecular diagnostic testing is currently available for only a certain number of disorders and with the increasing number of new genes associated to human diseases, it is becoming more and more challenging to provide cost effective molecular testing [10]. Preconception screening, together with genetic counselling of carriers, has resulted in remarkable declines in the incidence of several severe recessive diseases in populations at risk [19, 20]. However, extensive preconception screening and molecular diagnostic testing has been limited to targeted population or family history-directed individual and is still impractical for many disorders [20].

2.1. Mendelian Single Gene Disorders: At the Mutation or Gene Level

Single-gene disorders have a straight forward inheritance pattern, and the genetic causes can be traced to changes in specific individual genes. A particular disorder may be rare; however, as a group of disease-causing genes, single-gene disorders are responsible for a significant percentage of pediatric diseases [21]. About 1% of the approximately 4 million annual live births in the United States will have a single gene disorder that requires intensive clinical investigation, specific medical treatment and hospitalization [22, 23]. Based on the location of the relevant genes, single-gene traits can be divided into autosomal or sex-linked inheritance. Autosomal inheritance, depending on whether one or two mutant alleles are required to cause the disease phenotype, can be classified as autosomal dominant or autosomal recessive. Each of these single gene disorders, called Mendelian traits or diseases, are relatively uncommon. The frequency often varies with ethnic background, with each ethnic group having one or more Mendelian traits in higher frequency when compared to the other ethnic groups. For example, cystic fibrosis has a frequency of about one in 2,000 births in Americans descended from western European Caucasians [10] but is much rarer in African-Americans descent while sickle cell anemia has a frequency of about 1/600 births in African-American, but is rare in Caucasians [24]. Just to name a few, Mediterranean descent have a high frequency of thalassemia [25]; Eastern European Jews have a high frequency of Tay-Sachs disease [26, 27]; French Canadians from Quebec have a high frequency of tyrosinemia [28], all when compared to other ethnic groups. It has been estimated, regardless of the ethnicity, that each healthy individual is carrying between 1 and 8 mutations which, if found in the homozygous state would result in the expression of a Mendelian recessive disease [10]. Since each human genome has 22,333 genes it is unlikely that any two unrelated individuals would be carrying

the same mutations, even if they are from the same ethnic background. This explains why most Mendelian diseases are rare, affecting about 1/10,000 to 1/100,000 live births [21, 29].

In Quebec, the distribution of Mendelian diseases is due to local founder effects caused by the stratification of the contemporary French Canadian gene pool. The migration of a small number of French individuals from France to Quebec created a founder effect. Subsequent inland migrations have created smaller regional founder effects [30, 31]. The limited size of the population favoured genetic drift, and the social context encouraged endogamy, only few unions were reported between French Canadians with English and other immigrants [31]. The French-Canadian population of Quebec, currently about 6 million people, descends from about 7,798 immigrant founders who arrived in Quebec between 1608 and 1759 [31]. Recent studies showed that the Quebec population structure through the analysis of the genetic contribution of the first French settlers can be partitioned in eight regions, and they contributed to over 90% of gene pools in seven out of those eight regions [32]. This particular local genetic effect highlights the importance of considering the geographic origin of samples in the design of genetic testing in Quebec [33]. The conditions under which the peopling of Quebec was made, have favoured changes in the frequency of certain alleles in comparison with the French original population. As a result, certain genetic diseases are specific or more prevalent to the Quebec population. The prevalence and distribution of genetic diseases in Quebec is an essential factor to consider in clinical practice and particularly in differential diagnostic to prioritize molecular investigations [20].

This founder effect has impacted our molecular diagnostic testing system and is still a key factor when developing new diagnostic test for a genetic disease. The prevalence of the disease and the nature of the mutations found in the Quebec population need to be taken into account. The performance of the test depends on how well it accounts for the particularities of the disease in the French Canadians but more specifically depending on the regional founder effect [31]. The current changes in the immigration and the increased admixture is bringing new challenges in the differential diagnosis and amelioration of our molecular genetic testing system.

Over thirty Mendelian diseases have a high prevalence in the Quebec population [31, 34, 35]. Before the advances in sequencing technologies, it was believed that some of the disease were almost exclusive to the French Canadians, as autosomal recessive spastic ataxia of Charlevoix-Saguenay (ARSACS; MIM 270550), hereditary motor and sensory neuropathy with agenesis of the corpus callosum (ACCPN; MIM 218000) or French-Canadian-type Leigh syndrome (MIM 220111) [36-38]. Taking ARSACS as an example, after extensive resequencing of the responsible gene, SACS, it became evident that ARSACS was not limited to Quebec, and more than 100 different pathogenic mutations have now been identified worldwide [39, 40]. ARSACS is believed to be underdiagnosed in patients with atypical phenotypes and recent data on exome sequencing suggest a presumably high frequency of allele carriers around the world [39-42].

Admixture has more and more impact on the genetic characteristics of disease in French Canadians. Other ethnic groups have deep roots in the province. Many immigrant communities that are established mainly in and near Montreal now account for over 15% of the Quebec population. First Nations groups in Quebec have specific genetic diseases with three autosomal recessive conditions that have been well documented: Cree leukoencephalopathy [MIM 603896], Cree encephalitis [MIM 608505], North American Indian Childhood Cirrhosis [MIM 604901]. In Cree communities, the carrier frequency of Cree leukoencephalopathy is estimated at 1/10, 1/30 for Cree encephalitis and, 9/100 for the North American Indian Childhood

Cirrhosis [43, 44]. However, universal screening is currently offered only in the neonatal period for phenylketonuria, tyrosinaemia type I, medium-chain acyl-coenzyme A dehydrogenase deficiency [MCAD] and congenital hypothyroidism in addition to selected inborn errors of metabolism by urine screening [45, 46]. Effective targeted carrier-screening programs are provided in some specific communities but are often limited to individuals with a family history of recessive diseases. For public health, the genetic structure of Quebec presents major challenge for genetic screening but brings also opportunities for gene identification studies, clinical genetics research and practice. In the next section, we will discuss our experience over time with the implementation of effective targeted genetic testing and carrier-screening programs.

2.1.1. Targeted Mutations

The French-Canadian population of Quebec evolved as a mosaic of layered founder effects which has stimulated the development and feasibility of population-based carrier screening for at-risk individuals. For most of the Mendelian diseases in Quebec, the mutant founder alleles are characteristic and often unique. Usually one or two founder mutations account for 90% of French-Canadian alleles, depending on the region. However, founder mutations panels do exist for some disorders in Quebec such as for familial hypercholesterolemia, hereditary breast cancer, etc.

A good example of a mutant founder allele in Quebec is the Cree encephalitis, a severe early-onset progressive neurological disorder in an inbred Canadian Aboriginal community [MIM 608505]. The symptoms appear within the first few weeks of life and children usually die in infancy or early childhood. The main neurological symptoms are acquired microcephaly, mental retardation, cerebral atrophy with white matter changes, cerebral calcification, and chronic cerebrospinal fluid lymphocytosis. Cree encephalitis shows phenotypic overlap with Aicardi-Goutières syndrome with elevated levels of IFN- α in cerebrospinal fluid [47, 48]. The gene responsible, named *TREX1* (3-prime repair exonuclease 1, MIM 606609), has only one coding exon and is located on 3p21.31. The mutation causing Cree encephalitis is the p.Arg164Ter (c.490C>T) [49]. The molecular genetic diagnosis consists of determining the presence or absence of p.Arg164Ter in symptomatic patients. CREE encephalitis is among the leading causes of death of Cree infants. Cree carrier rates of this mutation are estimated to be 2-3 individual out of twenty meaning that about one in 300 births will be affected.

In 2006, a genetic screening program was developed and managed by the Cree Health Board to identify carriers of the mutation p.Arg164Ter in the *TREX1* gene. A second autosomal recessive condition is included in the Cree population genetic screening, Cree leukoencephalopathy [MIM 603896]. All patients are homozygous for the mutation p.Glu584Ala in the translation-initiation factor *EIF2B5* gene causing childhood ataxia with central hypomyelination and vanishing white matter disease [CACH/VWM] [50]. Both tests are performed in our molecular diagnostic laboratory at CHU Sainte-Justine. To date, >500 individuals have been tested. Pregnant woman and teenagers in High School are the targeted population for this genetic test. In fact, the rate of teenage pregnancy is very high in Cree population (23% among those aged <20, compared to 4.7% for the rest of Quebec).

Another autosomal recessive disorder caused by a founder mutation in Quebec is the Spastic ataxia of Charlevoix-Saguenay, more commonly known as ARSACS [MIM 270550]. This condition was first observed in people of the Charlevoix-Saguenay region of Quebec, Canada [51]. ARSACS is a debilitating progressive childhood neurological disorder affecting

the spinal cord and peripheral nerves. Most patients become wheelchair-bound but cognitive function is usually not affected. The incidence of ARSACS in the Charlevoix-Saguenay region of Quebec is estimated to be 1 in 1,500 to 2,000 individuals. Genetically confirmed cases of ARSACS have now been described in Italy, Spain, Tunisia, France, Belgium, Hungary, Morocco, Turkey, Serbia, other provinces of Canada, Netherlands, United Kingdom, Algeria and Japan. Therefore, ARSACS has a worldwide distribution but its prevalence in most countries is still unknown, except in the Netherlands where it may be as frequent as Friedreich Ataxia [52]. ARSACS is caused by mutations in the gene encoding the SACSIN protein (SACS) located on 13q12.12. Two major mutations were described in the Charlevoix-Saguenay region of Quebec representing 96,3% of the cause of ARSACS. Other types of mutation such as large deletion of 1.54 MB have been reported in some cases [53].

For carrier status genetic testing at the mutation level is the most appropriate method. However, with the growing number of clinically presumed ARSACS cases with a single mutation or no mutations identified, a systematic search for novel SACS point mutations and genomic rearrangements for the diagnosis of ARSACS is recommended. Recent reports demonstrated the clinical and genetic heterogeneity among ARSACS patients, now considered to be a common form of spastic ataxia worldwide, the challenge required a multimodal approach to rapidly screen larger numbers of samples and enable detection of both point mutations and deletions. Testing at the mutation level is no longer the optimal method with the exception of individuals with a clear genealogy to the Charlevoix Saguenay region of Quebec.

2.1.2. Targeted Gene

Rett Syndrome (RTT) is one of the most common causes of severe intellectual disability in females, with an incidence of 1:12,500 female live births. Most RTT cases are sporadic with no familial history of the disease. We now know that the majority of RTT patients have a *de novo* mutation in the X-linked gene methyl CpG binding protein 2 (*MECP2*) [54]. Very few families with multiple affected individuals, typical siblings, with Rett Syndrome have been described (for examples [55, 56]) where asymptomatic transmitting mothers have skewed X-chromosome inactivation [56], or parental gonadal mosaicism [57]).

Based on the RettBASE IRSF *MECP2* Variation Database, a database merging mutations and polymorphism data, there are more than 4,000 reported different variants in *MECP2* gene. Given that mutation in any of the 4 exons of *MECP2* can lead to the disease, the targeted gene screening approach is necessary for molecular diagnosis purpose. The same approach is used for autosomal non-syndromic intellectual disability and *SYNGAP1* gene [58].

2.2. Genetic Disease Panels: The Intermediate Decision

In recent years, there has been increasing recognition of the importance of inherited gene mutations as a cause of neurodevelopmental and metabolic diseases. Substantial progress has been made in elucidating the molecular defects underpinning these diseases, but the impact of these discoveries on clinical management is often unclear and the increasing number of genes associated with these features causes limitations of testing. The clinical spectrum of these diseases is vast and non-specific phenotype presentations are frequent.

In 1975, at the initiative of Dr Charles Scriver, the Province of Quebec implemented an integrated program for the diagnosis, counseling and treatment of hereditary metabolic diseases. Similar programs were implemented worldwide in order for physicians not to miss a treatable disorder. Clinical expression can be acute or systemic or can involve a specific organ, and can strike in the neonatal period or later and intermittently from infancy to late adulthood. During the last half century, many new disorders have been discovered and many therapeutic procedures have been tried. Some have been life-saving; others are still experimental. The long-term outcome of patient with metabolic patients is variable and requires multidisciplinary management and treatment [59, 60]. Several factors can impact on the availability and utility of genetic panel testing, such as access to testing, cost of testing and the mutation detection rate. There are over 100 different inborn errors of metabolism, including all the disorders detected by expanded newborn screening for which genetic causes have been identified. Traditionally the inherited metabolic diseases were categorized as disorders of carbohydrate metabolism, amino acid metabolism, organic acid metabolism or lysosomal storage diseases [61]. In recent decades, hundreds of new inherited disorders of metabolism have been discovered and the categories have proliferated. Genetic testing is also indicated in some multisystem disorders with frequent metabolic involvement, however genetic testing is not so straightforward for many of the other conditions. The overall incidence of the inborn errors of metabolism were estimated to be 70 per 100,000 live births or 1 in 1,400 [62]. As the number of inherited metabolic diseases that are included in state-based or province newborn screening programs continues to increase, ensuring the quality and delivery of testing services remain a continuous challenge. The vast majority of inborn errors of metabolism are inherited in an autosomal recessive manner. In the case of diseases carried on the X chromosome, males are usually more severely affected because they have only one copy. Females who have one X chromosome with the gene defect and one without may or may not develop symptoms.

Multi-exon or gene panel testing has proven efficient in cystic fibrosis (mutation panels; focussing on exon 6, 12, 23, 40, 97, 102 etc.) in the form of Ashkenazi disease panels and in the Charlevoix-Saguenay disease panel as described in the previous section. Now, molecular diagnostic laboratories offer panels for Comprehensive Carrier Screen Panels (comprising 80 – 100 conditions, between 350 – 500 variants). This multi-gene panel testing is becoming more and more advantageous for heterogeneous disorders such as the inborn inherited metabolic disorders. Historically, the development of carrier panels or diagnostic testing panels was subject to the recommendation of professional committees such as the American College of Medical Genetics and Genomics (ACMG) and the College of American Pathologists (CAP). Disease selection for the development of a diagnostic testing panel must take several points into consideration; 1) the disorder is clinically severe, 2) there is a high frequency of carriers in the screened population, 3) the availability of a reliable test with a high specificity and sensitivity, 4) the availability of prenatal diagnosis & access to genetic counselling, 5) genes recommended for screening by ACMG, 6) Autosomal recessive, X-linked diseases or with high de novo mutation rate [63-65].

As another example of multi-gene testing panel, in the past years, several comparative genomic hybridization studies have suggested recurrent rearrangements in synaptic and neurodevelopment genes. Rare CNVs at numerous loci are involved in the cause of mental retardation, autism, and schizophrenia [66-69]. In addition to those, Angelman syndrome and RTT have significant phenotypic overlap, including acquired microcephaly, loss of purposeful movements, developmental delay or intellectual disability and autistic behaviours. Additional

features include scoliosis, epilepsy, poor growth or obesity, and irregular breathing. There is also a broad clinical variability in the severity of the disease [70, 71]. *MECP2* mutations are present in 70-90% of females with classic RTT and approximately 20% of females with atypical RTT [72]. Partial deletions of *MECP2* are found in approximately 16% of girls with classic or atypical RTT. *CDKL5* mutations have been demonstrated in a broad spectrum of phenotypes including atypical RTT with infantile spasms [73]. Mutations of the *MEF2C* gene have been identified in patients with severe mental retardation, stereotypic movements, hypotonia, and epilepsy [74]. Abnormalities of the *FOXG1* gene have been identified in patients with the congenital variant of RTT [75]. Patients with the congenital variant of RTT have clinical features similar to classic RTT, but hypotonia and severe developmental delay starts in the first months of life [75]. Phenotypic overlap exists between patients with RTT/Angelman or atypical RTT led to a recommendation for sequencing all coding exons and the intron/exon boundaries of *MECP2*, *CDKL5*, *MEF2C*, and *FOXG1* as well as deletion/duplication analysis for all four genes by array-CGH or MLPA.

In the 1990's, gene testing was mostly done a research-based activity. After the Human Genome Project, companies started offering gene sequencing (Beginning of commercially availability for gene sequencing, Ambry Genetics, City of Hope, GeneDx, Mayo, Baylor, etc.). In the middle of 2000, Sanger sequencing for individual genes or a small group of related genes were offered. Mini-sequencing technologies emerged and let the promise of a more flexible platform to add new mutations or genes to the sequencing chip. However, the cost remains high and to lower the cost, pooling sample became the only feasible way to offer gene testing panel [23]. This reality caused was problematic to small diagnostic laboratories and more importantly for rare orphan diseases.

In the late 2000's, next-generation sequencing started to be used for sequencing medium and large gene panels. These sequencing technologies have benefits (faster diagnosis, reduced costs) and limitations (do not include all genes, incidental findings, variant of unknown significance). In an universal health care system, like in Quebec, diagnostic utility and yields are as important as cost efficiency. Current panels in use are a brief glimpse into the future of molecular diagnostic. Variants of unknown significance will continue to decrease over time and the next generation sequencing will perform better for the detection of copy number and complex rearrangements.

2.3. The “Omic” Approach, a Comprehensive Interrogation

The introduction of massively parallel sequencing platforms impacted on the cost of sequencing which in turn created, together with progresses in bioinformatics tools, a novel era of study in life sciences. These studies focus on large-scale data and can be divided into three main categories: genomics, proteomics and metabolomics hence the term “omic”. Thanks to this technological growth there came new clinical opportunities. These include: identification of genetic mechanism of heterogeneous disorders (those with dozen and hundreds of genes involved), targeted therapy (especially in cancer), targeted treatment (pharmacogenomics), prenatal screening (example, detection of Trisomy 21 through circulating cell free fetal DNA) and population screening for disease risk.

Genetic variations are a significant determinant of health and response to healthcare. Indeed, 70% of medical decisions rely on clinical laboratory results, and genetic innovations

will be a major source of diagnostic and prognostic information in this century. Although the number of genes linked to diseases is continually growing, the number of diagnostic tests has unfortunately lagged behind. Furthermore, complex diseases are expected to be caused by rare variations in a large number of functionally linked genes. For example, non-syndromic intellectual disability (NSID), the most frequent form of intellectual disability representing up to 2/3 of all cases, is caused by mutations in at least 34 genes. Together, these still explain less than 10% of sporadic intellectual disability cases, suggesting that the total number of NSID genes is likely to be above 100. Other well-documented examples of diseases showing high genetic heterogeneity include the spinocerebellar ataxias (35 genes), cardiomyopathies (150 genes), retinitis pigmentosa (78 genes), and deafness (160 genes). Comprehensive gene screening assays for each of these diseases is beyond the capabilities of the majority of diagnostic laboratories. The “omics” era holds the potential to solve the current problem of expanding molecular diagnosis by sequencing at once all the genes in the patient. Exome and genome sequencing are becoming cost-effective in a clinical setting as proven by the multiplication of providers offering clinical exome sequencing in the help for diagnosis (ex: Baylor College of Medicine, Ambry Genetics, GeneDx, UMC St Radboud, etc.).

Establishing an accurate diagnosis is the keystone on which medicine is based. However, an accurate diagnosis can only be accomplished through the use of diagnostic tools that are thoroughly validated for many properties (analytical validity, clinical validity, clinical utility, cost/effectiveness and cost-utility) and supported by the appropriate evidence-based data. A definitive diagnosis greatly facilitates the genetic counseling of families with diseases of early onset, leads to timely intervention, improved outcomes, and provides great reassurance to individuals and their families that they suffer from a known and definable disease.

Sequencing the exome of a patient in a clinical basis is of great value in two obvious situations: 1) finding a new mutation in a known gene underlying an “atypical presentation” (many examples can be found in the FORGE project, *Finding of Rare Disease Genes*: <http://www.cpgdsconsortium.com/>) and 2) identifying mutations in a novel gene. This could only be possible through a comprehensive approach. Moreover, exome sequencing is a faster and cheaper diagnostic method, it decreases the use of tests and additional consultations, is more accurate and offers personalized treatment. For example, a disease (intellectual disabilities) caused by mutations in ~100 genes will require sequencing of each gene, one at a time, at a cost ~ \$ 1,000/gene, and will take several years, for a total cost of over 100,000\$. With exome sequencing, screening of these 100 genes now costs ~\$ 2,000-\$ 7,000 (this cost and analysis turn-around time will decrease rapidly over the next few years). An “omic” approach can also have a positive economic impact. As an example, a 7 year old patient who presented with a neurodegenerative clinical picture characterized by spasticity was referred for exome sequencing. Prior to the sequencing of its entire genes, three pediatric neurologists assessed the child and a plethora of clinical examinations, biochemical tests and medical tests were completed. The three neurologist offered different diagnoses such as spastic paraparesis (> 45 known genes), spinocerebellar ataxia (> 15 known genes) and mitochondrial disease (> 300 known genes). The exploration of these assumptions was simply prohibitive as proposed by the clinical genetic testing targeted 20 genes and would have cost > \$ 40,000. Exome sequencing, with a cost of ~\$ 7000 rapidly led to the causative mutation/gene.

The major hurdle of exome sequencing is the missing or suboptimal exons coverage mainly due to high GC content. One good example is the aortopathies characterized by aortic dilation, which can lead to life threatening aneurysms and/or dissections. An early diagnosis is critical,

since timely initiation of pharmacological treatment can slow dilation and prophylactic surgery can prevent aortic dissection or rupture. Mortality rate is 20% for aortic dissections. Mutations in twelve different genes (for a total of 360 exons) are known to be responsible for these diseases. The commercial kit of exome capture SureSelect from Agilent is missing (or suboptimal) 11% exons of these genes. Fortunately, these weaknesses will be overcome with promised technological update.

3. THE FUTURE OF MOLECULAR DIAGNOSTIC LAB: WHAT WILL BE KEPT AND WHAT WILL BE DROPPED

Routine DNA-based diagnostic sequencing for neurodevelopmental or metabolic disorders currently targets well-defined genes or sets of genes that address very specific clinical questions. Today, available services vary greatly depending on the diagnostic laboratory, including Sanger sequencing of complete genes harboring mutations with known, well-described clinical phenotypes or multi-gene testing panels or next-generation sequencing-based disease/phenotype panels. Test ordering depends on the ability of a physician to apply a differential diagnosis; it follows a paradigm where a patient with phenotype A will have a mutation in a known gene causing phenotype A, which will be sequenced. Today's reality is that several genes are known to cause almost any particular phenotype, so gene panels can be used to sequence multiple genes simultaneously, or in a linear approach where the most commonly mutated genes are sequenced first, and if a negative result is obtained, the remainder are sequenced. The recent advances of next-generation sequencing enable the laboratory to simultaneously sequence a large number of genes at a significantly decreasing cost per gene. This raises multiple questions - is there a point at which additional sequencing diminishes the clinical utility of the resulting data? Is the proliferation of DNA testing panels enabled by next-generation sequencing beneficial for clinical diagnostics? What will be kept and what will be dropped?

This type of linear strategy is generally very satisfying for the diagnosing clinician as each test in specific population, for example in micro founder effect of Cree subpopulation or Saguenay, has a known value for diagnostic purpose. It also utilizes a focused approach, which tends to prioritize the most common gene involved in a certain phenotype. The genetics field is still young, but the actual genes and associated casual mutations are often incrementally discovered. Newly discovered genes are reported weekly, but it is generally difficult to systematically add them to a laboratory test menu. The diagnostic value from a single, well-characterized gene is relatively high since it is easier to interpret and it minimizes the likelihood that a variant of unknown significance will be discovered [64]. However, if a gene is reported only for a few cases, the diagnostic test may not be offered. Offering genetic testing for rare orphan disease has a high relative cost of test development and a low diagnostic yield [41, 63, 64].

As described in the previous section, some specific mutation panels will remain for at-risk population until the cost of next-generation sequencing decreases further. In addition, triplet-repeat diseases because of the unstable dynamic nature of those mutations will continue to require the use of a multimodal approach.

Next-generation sequencing is really moving the field of towards comprehensive gene panels. Expanded comprehensive diagnostic gene panels have several advantages. First of all, the design of an expanded diagnostic panel is quite straightforward and begins by identifying all of the genes involved in the disease(s) being targeted. Adding genes to the list is simple and does not require additional cost, unless enrichment is needed [63]. More comprehensive gene panel tests will simplify test ordering by consolidating all candidate genes into a single diagnostic test. By taking a more comprehensive approach, the sensitivity of the test increases and the rate of molecular under diagnosis decreases by including genes with low-frequency. It will also allow the identification of the causative gene in patients with an atypical phenotypic presentation. Moreover, with the advent of next-generation sequencing, an expanded neurodevelopmental and metabolic molecular diagnostic sequencing panel becomes economically feasible, allowing diagnosis of extremely rare orphan disease in one genetic test.

Next-generation sequencing introduces some interpretative challenges which are not new to diagnostic testing, but the scale of incidental findings or number of variants of unknown significance will be far greater than any previous test. Sanger sequencing typically identifies three categories of mutations: known non-pathogenic or benign polymorphisms; known pathogenic mutations; and variants of unknown significance. Variants identified as known pathogenic mutations are straightforward to interpret as in most cases, there are clearly indicated for clinical action [76]. With time, and better variant databases, the experience gained using next-generation sequencing testing panels will provide more data to improve variant interpretation and, over time, reduce the total number of variants of unknown significance encountered during next-generation sequencing diagnostic test [76]. The molecular clinical approach should focus on immediate clinical benefit of next-generation sequencing, such as the ability to offer comprehensive gene panels, to provide a first-pass diagnostic yield, and to limit gene list to genes of known diagnostic value [64]. This will aid the interpretation of the data set in a way that the healthcare provider will be able to understand and utilize.

4. THE NON-TARGETED GENETIC STRATEGY GENERATES NEW ETHICAL CHALLENGES

Particular emphasis has been placed in recent years on the identification of rare disease genes due to the availability of new genomic sequencing technologies. As seen earlier, these technologies greatly facilitate the search for causative disease mutations since they allow the analysis of the 22,333 genes at once in a short period of time for a reasonable cost. As a result, more genetic mechanisms involved in rare diseases are known, and a portion of these findings are already translated into diagnostic tools. Indeed, the clinical sequencing (referring to exome or genome sequencing) is rapidly being integrated into the practice of medicine. Although technically feasible and proven to be successful [77, 78], several challenges remain surrounding the use of clinical sequencing. Nonetheless molecular diagnosis using genomic and genetic methods such as microarray CGH, exome and genome sequencing have in common several ethical, legal and social concerns with other methods of medical investigation. Because of the unprecedentedly large amount of information generated by these comprehensive tests on an individual, the concerns (e.g., content of the consent form), incidental findings, returning results to the patient, etc. are amplified.

4.1. Consent Form

Consent forms need to be adapted to the new generation of sequencing clinical test. The main addition to the current medical consent form is the extent to which the patients would have control/access to the whole data (exome/genome), what results might be returned to the patient and what are the potential risks. The consent form should also inform the patient about the possible risk of discovering unwanted findings (unrelated to the original medical investigation). For instance, the exome analysis may reveal that the patient is likely to get a serious and untreatable disease.

4.2. Data Storage and Sharing

Next-generation sequencing technologies generate terabytes of data, and storing this amount of data will constitute a challenge in itself but also raised ethical issues. The genetic information of individual genome that is derived from the clinical sequencing is an accessible and robust “login” into a patient identity. It can give information on the patient’s relatives, ethnic groups, and more. The Wellcome Trust and the National Institute of Health, two respective Institutions whose mission is to operate open access genetic and genomic databases for the benefits of researchers and patients, now restricted and controlled their web database access. They even formed specific committees to oversee the circulation of the data included in their databases. This was effective following the conclusion from the study of Homer and colleagues in 2008 demonstrating that genome-wide association derived data is sufficient to re-identify an individual that have participated in a study [79]. They concluded that anonymizing data was unsatisfactory to protect the confidentiality of research participants. Because the understanding, diagnosed and treatment of complex diseases such as cancer or pediatric disorders required comparison of genetic information from several hundreds and often thousands of individuals, genetic data sharing is critical for research and molecular diagnosis. To further illustrate the significance of protecting information derived from whole genome sequencing of individual, the Presidential Commission for the Study of Bioethical Issues published a 150-page report, on the Privacy and progress in the era of whole genome sequencing. They concluded *“to realize the enormous promise that whole genome sequencing holds for advancing clinical care and the greater public good, individual interests in privacy must be respected and secured. As the scientific community works to bring the cost of whole genome sequencing down from millions per test to less than the cost of many standard diagnostic tests today, the Commission recognizes that whole genome sequencing and its increased use in research and the clinic could yield major advances in health care. However it could also raise ethical dilemmas. The Commission offers a dozen timely proactive recommendations that will help craft policies that are flexible enough to ensure progress and responsive enough to protect privacy.”* Finally, researchers and policy-makers need to find methods to protect individuals’ genomic data while still being able to share information.

4.3. Incidental Findings

As mentioned in earlier sections, clinical exome and genome sequencing present several advantages for clinical practice and for the patient. The main advantage discussed is the nondiscrimination on the selection of candidate's genes that will be analyzed. However, these methods can yield incidental medical information not related to the principal medical target. For instance, the genome of a child investigated for a rare form of deafness can lead to the discovery of a known mutation leading to a late onset neurodegenerative disorder. In addition, the patient can be found to be carrier of a recessive lethal disease. These "secondary" findings have been the subject of many discussions and recently the American College of Medical Genetics and Genomics whose mission is to improve health through medical genetics released its highly anticipated "Recommendations on Incidental Findings in Clinical Exome and Genome Sequencing".

In addition to fortuitous findings, whole genome sequencing raises great social concerns that still need to be resolved, such as possible forms of discrimination. Efforts have been made to overcome this phenomenon since the Genetic Information Nondiscrimination Act was signed in the USA in 2008 to protect individuals from improper use of genetics information with regard to health insurance and employment [69]. Emerging discoveries of susceptibility genes and gene variants associated with major neurological or psychiatric disorders are likely to challenge the existing ethical guidelines. It is up to researchers, health professionals and experts in the Ethical, Economic, Environmental, Legal, and Social aspects of genomics [GE3LS] to manage this growing scientific knowledge in a way that prioritizes protection of research participants and improves patient care.

CONCLUSION AND FUTURE

The successful use of whole-genome and exome sequencing for diagnosis has been amply confirmed by numerous studies. Whether targeted gene, gene panel approaches and exome sequencing will be entirely replaced by whole-genome sequencing is still unknown. Compared to traditional methods, it is now well accepted that exome sequencing has fewer false positives, and a greater sensitivity due to the higher coverage achieved when focusing only on a small fraction of the genome. Exome and whole-genome sequencing allow the discovery of point mutations and small deletion. With advanced bioinformatics tool it is now feasible to detect larger insertions and deletions (CNVs), currently detected using the microarrays technologies. We think that these "cytogenetic" methods, including karyotyping and molecular diagnosis, will merge. We also think that the current modalities of next-generation sequencing will eventually be replaced by genome sequencing. Recently, Saunders and colleague showed the feasibility of using whole-genome sequencing in neonatal intensive care units to screen for genetic disease diagnosis [77]. They described a 50-hour delay in the diagnosis of genetic disorders using whole genome sequencing.

Some challenges will subsist: mosaicism, balanced translocation and complex diseases (unknown mode of transmission). However, an important challenge in using whole-genome sequencing will be the interpretation of the data linking a genetic variation to the disease/phenotype and all ethical issues around it.

The rapid development of sequencing is now positively impacting prenatal diagnosis. Since the discovery of cell-free fetal nucleic acids circulating in the blood of pregnant women, [80] combined with next-generation DNA sequencing technologies, it is now possible to do early, noninvasive prenatal genetic testing [81]. Clinical applications of these methods already include fetal sex determination and blood group typing [82]. Ongoing research is currently evaluating the use of this approach for noninvasive detection of trisomies [83]. Other uses being explored are the detection of single-gene disorders, chromosomal abnormalities and inheritance of parental polymorphisms across the whole fetal genome.

Use of preimplantation genetic diagnosis and preimplantation genetic screening using next-generation sequencing can provide blastocyst preimplantation genetic diagnosis (PGD) results with high level of consistency with established diagnostic methods. Furthermore, single-gene disorder screening by next-generation sequencing could be performed in parallel with qPCR-based comprehensive chromosome screening or array SNP-CGH. Several studies showed that next-generation sequencing could serve as an essential PGD tools for further development of this important and emerging field [84]. Next-generation sequencing data provides a unique opportunity to evaluate multiple genomic loci and multiple samples on one experiment (e.g foetus can be tested in parallel with the parents). Next-generation sequencing might also be useful for simultaneous evaluation of aneuploidy, single-gene disorders, and translocations from the same biopsy without the need for multiple technological platforms [85]. Clearly, the fertility field has seen intensive efforts to significantly improve and validate better state-of-the-art in vitro fecundation techniques. Martin J. et al. suggested that IVF techniques coupled with deep comprehensive diagnosis/screening methods using NGS should result in high implantation and live birth rates [85]. Nevertheless, there is always room for improvements; it is important to be prudent, recognize the limits of sequence depth necessary to maintain accuracy, and variation in sequencing depth across different genomic loci that is critical to its clinical application in PGD.

We expect that whole-genome sequencing will allow the identification of genetic variants that will determine an individual's risk for developing diseases, including neurological or psychiatric disorders. In fact, the continuing reduction in sequencing costs may lead to replacement of most of the other currently used approaches.

ACKNOWLEDGMENTS

We are very grateful for the support of our Institution. We wish to thank our colleagues, Mélanie Lafleur, Françoise Couture, Sylvie Filiatrault, Josée Gauthier, Janique Ladouceur, Caroline Deschaînes, Marie-Josée Lassonde, Julie Ménard, Li Fan and Marie-Lyne Darveau for their technical and administrative work in the Molecular Diagnostic Lab. We also acknowledge the work/input of our collaborative clinicians: Drs. Jacques Michaud, Grant Mitchell, Jean-François Soucy, Dorothée Dal Soglio, Luc Oligny, George-Etienne Rivard and Sonia Cellot for their help in the discovery, development and for working in close collaboration to improve the use of genetics and genomic knowledge and tools in today's medicine.

REFERENCES

- [1] Sanger, F., S. Nicklen, and A.R. Coulson, DNA sequencing with chain-terminating inhibitors. *Proc Natl Acad Sci U S A*, 1977. 74(12): p. 5463-7.
- [2] Pruitt, K.D., et al., NCBI Reference Sequences: current status, policy and new initiatives. *Nucleic Acids Res*, 2009. 37(Database issue): p. D32-6.
- [3] *The Human Genome Project*. <http://genomics.energy.gov.>, 1999.
- [4] Awadalla, P., et al., Direct measure of the de novo mutation rate in autism and schizophrenia cohorts. *Am J Hum Genet*, 2010. 87(3): p. 316-24.
- [5] Hamdan, F.F., et al., Excess of de novo deleterious mutations in genes associated with glutamatergic systems in nonsyndromic intellectual disability. *Am J Hum Genet*, 2011. 88(3): p. 306-16.
- [6] Griswold, A.J., et al., Evaluation of Copy Number Variations Reveals Novel Candidate Genes in Autism Spectrum Disorder Associated Pathways. *Hum Mol Genet*, 2012.
- [7] Elia, J., et al., Genome-wide copy number variation study associates metabotropic glutamate receptor gene networks with attention deficit hyperactivity disorder. *Nat Genet*, 2012. 44(1): p. 78-84.
- [8] Pfundt, R. and J.A. Veltman, Structural genomic variation in intellectual disability. *Methods Mol Biol*, 2012. 838: p. 77-95.
- [9] Malhotra, D., et al., High frequencies of de novo CNVs in bipolar disorder and schizophrenia. *Neuron*, 2011. 72(6): p. 951-63.
- [10] Kingsmore, S., Comprehensive carrier screening and molecular diagnostic testing for recessive childhood diseases. *PLoS Curr*, 2012: p. e4f9877ab8ffa9.
- [11] Dreesen, J.C., et al., Multiplex PCR of polymorphic markers flanking the CFTR gene; a general approach for preimplantation genetic diagnosis of cystic fibrosis. *Mol Hum Reprod*, 2000. 6(5): p. 391-6.
- [12] Moutou, C., N. Gardes, and S. Viville, Multiplex PCR combining deltaF508 mutation and intragenic microsatellites of the CFTR gene for pre-implantation genetic diagnosis (PGD) of cystic fibrosis. *Eur J Hum Genet*, 2002. 10(4): p. 231-8.
- [13] Schouten, J.P., et al., Relative quantification of 40 nucleic acid sequences by multiplex ligation-dependent probe amplification. *Nucleic Acids Res*, 2002. 30(12): p. e57.
- [14] Hochstenbach, R., et al., Rapid detection of chromosomal aneuploidies in uncultured amniocytes by multiplex ligation-dependent probe amplification (MLPA). *Prenat Diagn*, 2005. 25(11): p. 1032-9.
- [15] Marzese, D.M., et al., Detection of deletions and duplications in the Duchenne muscular dystrophy gene by the molecular method MLPA in the first Argentine affected families. *Genet Mol Res*, 2008. 7(1): p. 223-33.
- [16] Engert, S., et al., MLPA screening in the BRCA1 gene from 1,506 German hereditary breast cancer cases: novel deletions, frequent involvement of exon 17, and occurrence in single early-onset cases. *Hum Mutat*, 2008. 29(7): p. 948-58.
- [17] Makrythanasis, P., et al., De novo duplication of MECP2 in a girl with mental retardation and no obvious dysmorphic features. *Clin Genet*, 2010. 78(2): p. 175-80.
- [18] Thompson, T., *Genetics in Medicine*. 7th ed, ed. S. Elsevier. 2007, Philadelphia: Elsevier. 567.

- [19] Andermann, A., I. Blancquaert, and V. Dery, Genetic screening: a conceptual framework for programmes and policy-making. *J Health Serv Res Policy*, 2010. 15(2): p. 90-7.
- [20] Laberge, A.M., et al., Population history and its impact on medical genetics in Quebec. *Clin Genet*, 2005. 68(4): p. 287-301.
- [21] Rodelsperger, C., et al., Identity-by-descent filtering of exome sequence data for disease-gene identification in autosomal recessive disorders. *Bioinformatics*, 2011. 27(6): p. 829-36.
- [22] Yu, J.H., et al., Self-guided management of exome and whole-genome sequencing results: changing the results return model. *Genet Med*, 2013.
- [23] Jamal, S.M., et al., Practices and policies of clinical exome sequencing providers: analysis and implications. *Am J Med Genet A*, 2013. 161A(5): p. 935-50.
- [24] Section on Hematology/Oncology Committee on, G. and P. American Academy of, Health supervision for children with sickle cell disease. *Pediatrics*, 2002. 109(3): p. 526-35.
- [25] Atanasovska, B., et al., Efficient detection of Mediterranean beta-thalassemia mutations by multiplex single-nucleotide primer extension. *PLoS One*, 2012. 7(10): p. e48167.
- [26] Natowicz, M.R. and E.M. Prenc, Heterozygote screening for Tay-Sachs disease: past successes and future challenges. *Curr Opin Pediatr*, 1996. 8(6): p. 625-9.
- [27] Rozenberg, R. and V. Pereira Lda, The frequency of Tay-Sachs disease causing mutations in the Brazilian Jewish population justifies a carrier screening program. *Sao Paulo Med J*, 2001. 119(4): p. 146-9.
- [28] Bergeron, P., C. Laberge, and A. Grenier, Hereditary tyrosinemia in the province of Quebec: prevalence at birth and geographic distribution. *Clin Genet*, 1974. 5(2): p. 157-62.
- [29] Guillemin, L., [The national plan for orphan rare diseases: nearly 10 years on]. *Rev Neurol (Paris)*, 2013. 169 Suppl 1: p. S9-11.
- [30] Bouchard, G., C. Laberge, and C.R. Scriver, [Demographic reproduction and genetic transmission in the north-east of the province of Quebec (18th-20th centuries)]. *Eur J Popul*, 1988. 4(1): p. 39-67.
- [31] Laberge, A.M., [Prevalence and distribution of genetic diseases in Quebec: impact of the past on the present]. *Med Sci (Paris)*, 2007. 23(11): p. 997-1001.
- [32] Roy-Gagnon, M.H., et al., Genomic and genealogical investigation of the French Canadian founder population structure. *Hum Genet*, 2011. 129(5): p. 521-31.
- [33] Bherer, C., et al., Admixed ancestry and stratification of Quebec regional populations. *Am J Phys Anthropol*, 2011. 144(3): p. 432-41.
- [34] Scriver, C.R., et al., Feasibility of chemical screening of urine for neuroblastoma case finding in infancy in Quebec. *CMAJ*, 1987. 136(9): p. 952-6.
- [35] Scriver, C.R., Human genetics: lessons from Quebec populations. *Annu Rev Genomics Hum Genet*, 2001. 2: p. 69-101.
- [36] Engert, J.C., et al., ARSACS, a spastic ataxia common in northeastern Quebec, is caused by mutations in a new gene encoding an 11.5-kb ORF. *Nat Genet*, 2000. 24(2): p. 120-5.
- [37] Debray, F.G., et al., LRPPRC mutations cause a phenotypically distinct form of Leigh syndrome with cytochrome c oxidase deficiency. *J Med Genet*, 2011. 48(3): p. 183-9.

- [38] Howard, H.C., et al., The K-Cl cotransporter KCC3 is mutant in a severe peripheral neuropathy associated with agenesis of the corpus callosum. *Nat Genet*, 2002. 32(3): p. 384-92.
- [39] Thiffault, I., et al., Diversity of ARSACS mutations in French-Canadians. *Can J Neurol Sci*, 2013. 40(1): p. 61-6.
- [40] Synofzik, M., et al., Autosomal recessive spastic ataxia of Charlevoix Saguenay (ARSACS): expanding the genetic, clinical and imaging spectrum. *Orphanet J Rare Dis*, 2013. 8: p. 41.
- [41] Chen, Z., et al., Using next-generation sequencing as a genetic diagnostic tool in rare autosomal recessive neurologic Mendelian disorders. *Neurobiol Aging*, 2013.
- [42] Tzoulis C, J.S., Haukanes BI, Boman H, Knappskog PM, et al., Novel SACS Mutations Identified by Whole Exome Sequencing in a Norwegian Family with Autosomal Recessive Spastic Ataxia of Charlevoix-Saguenay. *PLoS One*, 2013. 8(6): p. e66145.
- [43] Black, D.N., et al., Leukoencephalopathy among native Indian infants in northern Quebec and Manitoba. *Ann Neurol*, 1988. 24(4): p. 490-6.
- [44] Drouin, E., et al., North American Indian cirrhosis in children: a review of 30 cases. *J Pediatr Gastroenterol Nutr*, 2000. 31(4): p. 395-404.
- [45] Laberge, C., et al., [The genetic medicine network in Quebec: An integrated program for diagnosis, counseling and treatment of hereditary metabolic diseases]. *Union Med Can*, 1975. 104(3): p. 428-32.
- [46] Grenier, A. and C. Laberge, [Detection of hereditary metabolic diseases in Quebec]. *Union Med Can*, 1974. 103(3): p. 453-6.
- [47] Crow, Y.J., et al., Cree encephalitis is allelic with Aicardi-Goutieres syndrome: implications for the pathogenesis of disorders of interferon alpha metabolism. *J Med Genet*, 2003. 40(3): p. 183-7.
- [48] Black, D.N., et al., Encephalitis among Cree children in northern Quebec. *Ann Neurol*, 1988. 24(4): p. 483-9.
- [49] Crow, Y.J., et al., Mutations in the gene encoding the 3'-5' DNA exonuclease TREX1 cause Aicardi-Goutieres syndrome at the AGS1 locus. *Nat Genet*, 2006. 38(8): p. 917-20.
- [50] Fogli, A., et al., Cree leukoencephalopathy and CACH/VWM disease are allelic at the EIF2B5 locus. *Ann Neurol*, 2002. 52(4): p. 506-10.
- [51] Bouchard JP, R.A., Melancon SB, Mathieu J, Michaud J., Autosomal recessive spastic ataxia (Charlevoix– Saguenay). *Handbook of ataxia disorders*, Klockgether T, New York: Marcel Dekker, 2000: 311–324., 2000.
- [52] Vermeer, S., et al., ARSACS in the Dutch population: a frequent cause of early-onset cerebellar ataxia. *Neurogenetics*, 2008. 9(3): p. 207-14.
- [53] Breckpot, J., et al., A novel genomic disorder: a deletion of the SACS gene leading to spastic ataxia of Charlevoix-Saguenay. *Eur J Hum Genet*, 2008. 16(9): p. 1050-4.
- [54] Amir, R.E., et al., Rett syndrome is caused by mutations in X-linked MECP2, encoding methyl-CpG-binding protein 2. *Nat Genet*, 1999. 23(2): p. 185-8.
- [55] Curtis, A.R., et al., X chromosome linkage studies in familial Rett syndrome. *Hum Genet*, 1993. 90(5): p. 551-5.
- [56] Schanen, N.C., et al., A new Rett syndrome family consistent with X-linked inheritance expands the X chromosome exclusion map. *Am J Hum Genet*, 1997. 61(3): p. 634-41.

- [57] Wan, M., et al., Rett syndrome and beyond: recurrent spontaneous and familial MECP2 mutations at CpG hotspots. *Am J Hum Genet*, 1999. 65(6): p. 1520-9.
- [58] Hamdan, F.F., et al., Mutations in SYNGAP1 in autosomal nonsyndromic mental retardation. *N Engl J Med*, 2009. 360(6): p. 599-605.
- [59] Saudubray, J.M., F. Sedel, and J.H. Walter, Clinical approach to treatable inborn metabolic diseases: an introduction. *J Inherit Metab Dis*, 2006. 29(2-3): p. 261-74.
- [60] Saudubray, J.M., Neurometabolic disorders. *J Inherit Metab Dis*, 2009. 32(5): p. 595-6.
- [61] Fernandes, J.S., J.M.; van den Berghe, G.; Walter, J.H., *Inborn Metabolic Diseases: Diagnosis and Treatments*. 4th ed, ed. Springer. 2006: Springer.
- [62] Applegarth, D.A., J.R. Toone, and R.B. Lowry, Incidence of inborn errors of metabolism in British Columbia, 1969-1996. *Pediatrics*, 2000. 105(1): p. e10.
- [63] Ku, C.S., et al., Exome sequencing: dual role as a discovery and diagnostic tool. *Ann Neurol*, 2012. 71(1): p. 5-14.
- [64] Klee, E.W., N.L. Hoppman-Chaney, and M.J. Ferber, Expanding DNA diagnostic panel testing: is more better? *Expert Rev Mol Diagn*, 2011. 11(7): p. 703-9.
- [65] Levenson, D., The tricky matter of secondary genomic findings: ACMG plans to issue recommendations. *Am J Med Genet A*, 2012. 158A(7): p. ix-x.
- [66] Lee, J.A. and J.R. Lupski, Genomic rearrangements and gene copy-number alterations as a cause of nervous system disorders. *Neuron*, 2006. 52(1): p. 103-21.
- [67] Kusenda, M. and J. Sebat, The role of rare structural variants in the genetics of autism spectrum disorders. *Cytogenet Genome Res*, 2008. 123(1-4): p. 36-43.
- [68] Guilmatre, A., et al., Recurrent rearrangements in synaptic and neurodevelopmental genes and shared biologic pathways in schizophrenia, autism, and mental retardation. *Arch Gen Psychiatry*, 2009. 66(9): p. 947-56.
- [69] Cantor, R.M. and D.H. Geschwind, Schizophrenia: genome, interrupted. *Neuron*, 2008. 58(2): p. 165-7.
- [70] Williams, C.A., et al., Angelman syndrome: mimicking conditions and phenotypes. *Am J Med Genet*, 2001. 101(1): p. 59-64.
- [71] Marschik, P.B., et al., Changing the perspective on early development of Rett syndrome. *Res Dev Disabil*, 2013. 34(4): p. 1236-9.
- [72] Neul, J.L., The relationship of Rett syndrome and MECP2 disorders to autism. *Dialogues Clin Neurosci*, 2012. 14(3): p. 253-62.
- [73] Maortua, H., et al., CDKL5 gene status in female patients with epilepsy and Rett-like features: two new mutations in the catalytic domain. *BMC Med Genet*, 2012. 13: p. 68.
- [74] Novara, F., et al., MEF2C deletions and mutations versus duplications: A clinical comparison. *Eur J Med Genet*, 2013. 56(5): p. 260-5.
- [75] Guerrini, R. and E. Parrini, Epilepsy in Rett syndrome, and CDKL5- and FOXP1-gene-related encephalopathies. *Epilepsia*, 2012. 53(12): p. 2067-78.
- [76] Richards, C.S., et al., ACMG recommendations for standards for interpretation and reporting of sequence variations: Revisions 2007. *Genet Med*, 2008. 10(4): p. 294-300.
- [77] Saunders, C.J., et al., Rapid whole-genome sequencing for genetic disease diagnosis in neonatal intensive care units. *Sci Transl Med*, 2012. 4(154): p. 154ra135.
- [78] de Ligt, J., et al., Diagnostic exome sequencing in persons with severe intellectual disability. *N Engl J Med*, 2012. 367(20): p. 1921-9.

- [79] Homer, N., et al., Resolving individuals contributing trace amounts of DNA to highly complex mixtures using high-density SNP genotyping microarrays. *PLoS Genet*, 2008. 4(8): p. e1000167.
- [80] Lo, Y.M., et al., Presence of fetal DNA in maternal plasma and serum. *Lancet*, 1997. 350(9076): p. 485-7.
- [81] Sayres, L.C. and M.K. Cho, Cell-free fetal nucleic acid testing: a review of the technology and its applications. *Obstet Gynecol Surv*, 2011. 66(7): p. 431-42.
- [82] Lo, Y.M., et al., Prenatal diagnosis of fetal RhD status by molecular analysis of maternal plasma. *N Engl J Med*, 1998. 339(24): p. 1734-8.
- [83] Chiu, R.W., et al., Non-invasive prenatal assessment of trisomy 21 by multiplexed maternal plasma DNA sequencing: large scale validity study. *BMJ*, 2011. 342: p. c7401.
- [84] Treff, N.R., et al., Evaluation of targeted next-generation sequencing-based preimplantation genetic diagnosis of monogenic disease. *Fertil Steril*, 2013. 99(5): p. 1377-1384 e6.
- [85] Martin, J., et al., The impact of next-generation sequencing technology on preimplantation genetic diagnosis and screening. *Fertil Steril*, 2013. 99(4): p. 1054-61 e3.

Chapter 97

IMPROVEMENTS IN HSV-1- DERIVED AMPLICON VECTORS FOR GENE TRANSFER

**Matias E. Melendez^{1,2,*}, Alejandra I. Aguirre^{3,*},
Maria V. Baez³, Carlos A. Bueno^{1,4}, Anna Salvetti⁵,
Diana A. Jerusalinsky^{†,*},³ and Alberto L. Epstein^{†,*},^{1,5}**

¹Centre de Génétique et de Physiologie Moléculaire et Cellulaire,
CNRS UMR5534, Université Claude Bernard Lyon 1,
Villeurbanne, France

²Molecular Oncology Research Center,
Barretos Cancer Hospital, SP, Brazil

³Instituto de Biología Celular y Neurociencia "Prof. Eduardo De Robertis",
CONICET-UBA, Facultad de Medicina, Universidad de Buenos Aires,
(1121) Ciudad de Buenos Aires, Argentina

⁴Laboratorio de Virología, Departamento de Química Biológica,
Facultad de Ciencias Exactas y Naturales, Universidad de Buenos Aires, Argentina

⁵Centre International de Recherche en Infectiologie (CIRI) INSERM U1111 – CNRS
UMR5308 Université Lyon 1, Lyon, France

ABSTRACT

Viral vectors engineered to carry transgenic sequences can be delivered into discrete tissues or anatomical structures to express specific transgenes into the transduced cells. Therefore, they are useful tools to produce specific, transient and localized knockout, knockdown, ectopic expression or overexpression of a gene, leading to the possibility of analyzing both *in vitro* and *in vivo* molecular basis of relevant functions. Replication-incompetent helper-dependent amplicon vectors, derived from herpes simplex virus type-

* Equivalent contribution to this work.

† Corresponding Author's Email: alberto.epstein@ens-lyon.fr (CIRI INSERM U1111 – CNRS UMR5308. Université Lyon 1, ENS de Lyon. 46 Allée d'Italie. 69364 LYON cedex 07, France).

† Corresponding Author's Email: djerusal@gmail.com (IBCN - School Medicine. University of Buenos Aires. 2155 Paraguay Street, 3rd floor. (1121) City of Bs. As., Argentina).

1 (HSV-1) are devoid of viral genes. Thus, these vectors have great advantages as tools for *in vitro* and *in vivo* gene transfer and, in particular (i) minimal toxicity or induction of adaptive immune responses, and (ii) large transgene capacity, being able to carry up to 150 kbp of foreign DNA. In addition, these vectors have (iii) widespread cellular tropism: amplicons can experimentally infect several cell types, either quiescent or not, though naturally HSV-1 infects mainly neurons and epithelial cells, and (iv) absence of insertional mutagenesis, since the viral genome does not integrate into the host cell genome. These vectors have been used both on basic and applied research, and they have revealed as most suitable tools to study complex functions involving the nervous system, such as anxiety, sexual behavior, learning and memory. In addition, amplicon vectors are being used for the development of new experimental gene therapy approaches, both for inherited and acquired diseases affecting the nervous system, including neuro-degenerative diseases. Although several technological improvements have been achieved in the last decade, some difficulties regarding these appealing vectors remain still unresolved, such as the inability to generate large amounts of high-titer fully helper-free vectors and the fact that expression from the transgenic sequence delivered by the vectors is generally unstable, often leading to a complete silencing of expression after a few weeks. To overcome these obstacles and to improve these vectors, we have recently modified (a) the amplicon genome, in order to fully delete bacterial sequences and (b) developed novel complementing cell lines, in order to improve helper-free vector production and to render amplicon stocks compatible with clinical trials. In this review article we briefly review data supporting the potential of HSV-1-based amplicon vector model for gene delivery in primary cultures of neural cells and into the brain of living animals.

Keywords: HSV-1-derived vectors, Amplicons, nervous system, neurodegenerative disorders, cancer, vaccines, experimental gene therapy

HSV-1 BACKGROUND

HSV-1 Epidemiology and Clinical Features

HSV-1 is one of the most common human pathogens, infecting 40–80% of people worldwide. Primary, productive infections of HSV-1 typically occur in regions of the orofacial mucosa, causing a mild gingivo-stomatitis. The virus particles produced during this primary infection enter the peripheral sensory neurons, where the virus genome often remains into a latent for the lifetime of the host. Reactivation, occurring generally following stress, usually leads to recurrent lesions in the vicinity of the primary infected area and are typically limited to cold sores of the mouth. Although recrudescence herpes may affect any site along the involved sensory division of the fifth cranial nerve, the most frequent location is the muco-cutaneous junction of the lips [1]. Classic lesions are characterized by virally induced epithelial damage and go through different clinical stages. Initially, a transient cluster of micro-vesicles appears at the site of recrudescence and breaks open to form irregular, superficial erosions that crust over and heal without scarring over a period of 1–2 weeks. Shedding of the virus is often present for several days after resolution of clinical signs and symptoms [1]. HSV-1 infection can also cause ocular herpes and, much more infrequently, it can cause encephalitis. In neonates and in immune-depressed individuals HSV-1 can cause severe disseminated infection with neurological impairment and high mortality [2].

HSV-1 Virion Structure

The architecture of this enveloped double-stranded DNA virus, of 220 nm in diameter, is highly complex [3]. The mature HSV-1 virion (Fig. 1) consists of the following four components:

- A core of double-stranded (ds) DNA
- An icosadeltahedral capsid
- The tegument, which is a layer of proteins located between the envelope and the underlying capsid; these proteins play a critical role in virion morphogenesis, and they also help the virus to take the control of the expression machinery of the cell very early following infection
- A lipid envelope from cellular origin, where viral glycoproteins and other membrane proteins are embedded, several of which are involved in receptor-mediated cellular entry.

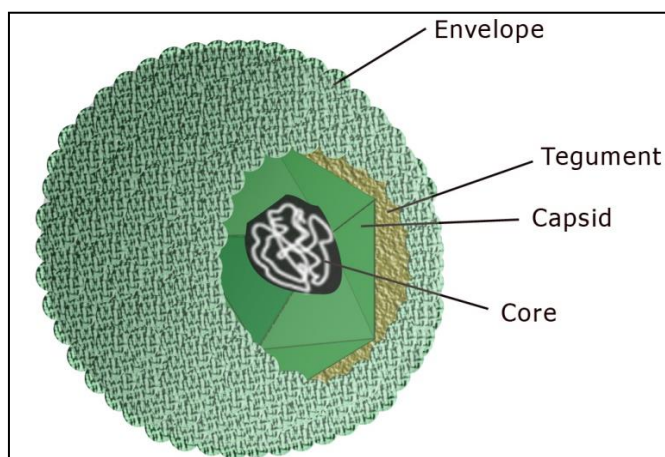


Figure 1. HSV-1 Virion.

HSV-1 Genome

The linear 152-kbp dsDNA virus genome (Fig. 2) is densely packaged within the capsid cavity [4] and is devoid of histone proteins at this stage [5]. This genome is arranged as long (UL) and short (US) unique segments, of 126 and 26 kbp, respectively. The UL and US segments are flanked by repeated sequences, designated ab, b'a' a'c' and ca (or TRL, IRL, IRS and TRS, for Terminal Repeat L, Internal Repeat L, Internal Repeat S and Terminal repeat S) [6].

The HSV-1 genome, which is 68% G+C rich, has been fully sequenced [7-9]. At least eighty-four viral genes are encoded, and these may be divided into essential and non-essential genes, according to whether their expression is necessary or not for viral growth in permissive cultured cells. HSV-1 genes do not contain intron sequences, with the exception of those encoding ICP0, ICP22, ICP47, UL15 and LAT. Nonessential genes often encode functions that

are important for specific virus-host interaction *in vivo*, such as immune evasion, replication in non-dividing cells or shutdown of host protein synthesis. Approximately half of the viral genes have been shown to be dispensable for replication of the virus in cultured cells, and thus, these genes could be replaced by exogenous genetic material, which has been the premise for the development of HSV-1-based vectors for gene therapy [10]. In contrast, most genes involved in virus entry, DNA replication, capsid assembly and DNA packaging are essential.

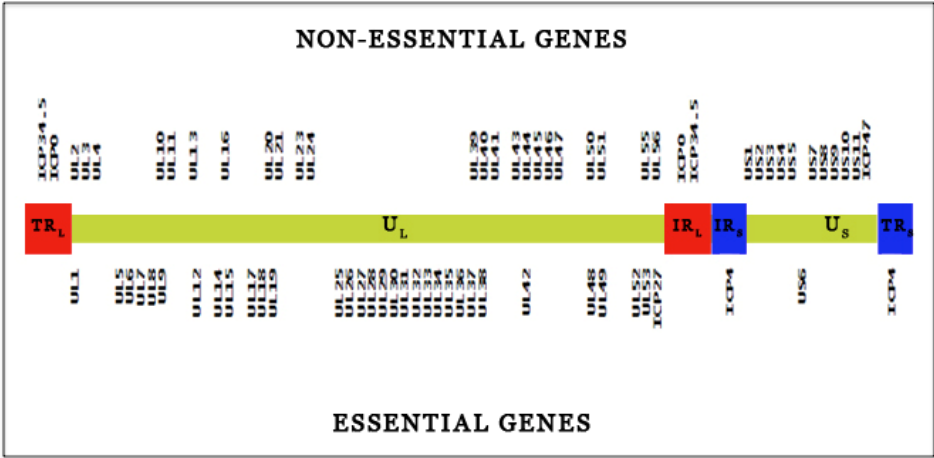


Figure 2. HSV-1 genome.

The HSV-1 genome contains two *cis*-acting sequences that are essential for virus multiplication: the viral DNA replication origins and the cleavage/packaging signals. HSV-1 harbors three lytic origins of replication, with two located within the repeated segments surrounding US (oriS) and one in the UL segment (oriL) [11]. Mutant viruses lacking either oriL or both copies of oriS are replication competent, suggesting that all origins are functionally competent [6]. The oriS is located within the shared promoter regulatory region between the divergently transcribed immediate-early genes ICP4 and ICP22/ICP47, whereas oriL lies in the regulatory region between two divergently transcribed early genes that are essential for viral DNA replication: the major DNA-binding protein and the HSV-1 DNA polymerase [12].

The cleavage/packaging signals of the HSV-1 genome are located within the *a* sequences, generally found in tandem repeats at both ends of the genome, as well as at the L/S junction [13, 14]. In different HSV-1 strains, the *a* sequence ranges from 250 to 500 bp. An approximately 200-bp fragment (Uc-DR1-Ub) spanning the junction between tandem *a* sequences has been shown to contain all the essential *cis*-acting sequences necessary for DNA cleavage and packaging [15]. The specific signals for DNA cleavage and packaging, termed pac1 and pac2, are located within the Ub and Uc regions, respectively [14-16].

HSV-1-DERIVED VECTORS

The improvement of methods for efficient delivery of genetic material in mammalian cells has been a major objective of molecular and cellular biology, gene therapy and vaccine development during the last 30 years and is still a subject of increasing interest. Viral-derived

vectors are the most promising gene transfer tools as viruses have naturally evolved as molecular carriers to specific tissues. As already quoted, HSV-1 is a naturally epithelial-neurotropic virus, highly adapted to the nervous system environment of the host organism, and neurons and glia are thus the most common target cells for HSV-1-derived vectors. Coding more than 80 genes, HSV-1 is a complex virus, which can be engineered to exquisitely design viral vectors for fundamental research and gene therapy of neurological disorders. In general terms, HSV-1 possesses several features that make it especially interesting as a vector:

- Although essentially neurotropic, under experimental conditions HSV-1 has a broad host cell range, due to the fact that its cellular receptors are widely distributed. In addition, HSV-1 infects and replicates in cells from different mammals, easily allowing preclinical evaluation.
- It can infect dividing and non-dividing cells.
- It generally causes nonthreatening diseases.
- It has a very large transgene capacity, allowing to deliver multiple or large transgenes (approximately half of the genome is nonessential in cultured cells and can be deleted without compromising viral replication, allowing to lodge considerable amounts of transgenic DNA). Moreover, HSV-1-derived amplicon vectors are devoid of almost the totality of their genome (i.e., 152-kbp DNA), which can therefore be replaced by exogenous DNA.
- Safe replication-defective HSV-1 vectors may be prepared to high titers.
- The ability to establish a latent state in neurons may be very useful for stable long-term expression of therapeutic transgenes.
- HSV-1 genome does not integrate into host chromosomes, remaining episomal and reducing the risk of insertional mutagenesis.

To exploit the versatility of HSV-1 for gene transfer, three types of HSV-1-based vectors have been developed: **(i)** attenuated replication-competent vectors, **(ii)** replication-defective recombinant vectors, and **(iii)** amplicon vectors. Attenuated replication-competent HSV-1 vectors can replicate only in certain cell types and tissues *in vivo* due to deletion of non-essential viral genes. Such vectors have been typically used for the development of tumor therapies, and are usually referred to as oncolytic HSV-1 vectors. However, other uses of attenuated vectors are possible, for example to deliver genes to the peripheral sensory ganglia following infection at peripheral tissues. Replication-defective recombinant HSV-1 vectors are devoid of one or more viral genes that are essential for lytic replication (such as the immediate early genes ICP4 and ICP27), but they can retain their ability to establish latency [17, 18]. The production of these vectors requires complementing cell lines expressing the deleted viral genes. This type of vectors can be used as genetic vaccines or for gene therapy of neurological disorders.

Lastly, Spaete and Frenkel [19] reported the natural generation of fully defective HSV-1 genomes, in which almost all the transacting virus genes were absent. These defective genomes were formed mainly by repeats of the two non-coding HSV-1 sequences, i.e., the origins of DNA replication and the cleavage/packaging signals. Moreover, they have showed that these defective genomes could be replicated and packaged into particles in the presence of a helper virus that supplied all the virus functions *in trans*, thus demonstrating that the origins of virus replication and the cleavage/packaging signals were the only two *cis*-acting sequences required for replication and packaging of a defective virus genome. These *cis*-acting sequences from the

HSV-1 genome were then incorporated into a standard bacterial plasmid, which they have termed the amplicon plasmid [19], and they demonstrated that the amplicon plasmid could also be replicated and packaged into viral particles in the presence of a helper system, thus generating what they have termed amplicon vectors. This chapter will be devoted to the biology and applications of this very appealing type of vectors.

Amplicon Vectors

A conventional HSV-1 amplicon plasmid consists of (i) a plasmid backbone harboring a bacterial origin of DNA replication and an antibiotic resistance gene for propagation in bacteria, (ii) an HSV-1 origin of DNA replication (generally oriS), (iii) an HSV-1 cleavage/packaging sequence (pac or *a*), (iv) a transcription unit expressing a reporter gene, generally encoding the green fluorescence protein (GFP), which is useful for identifying the amplicon-infected cells and to facilitate the titration of amplicon stocks, and (v), a multiple cloning site (MCS) for introducing the transgene(s) of interest (Fig. 3). This basic structure of amplicon plasmids has remained almost unchanged from the beginning [19].

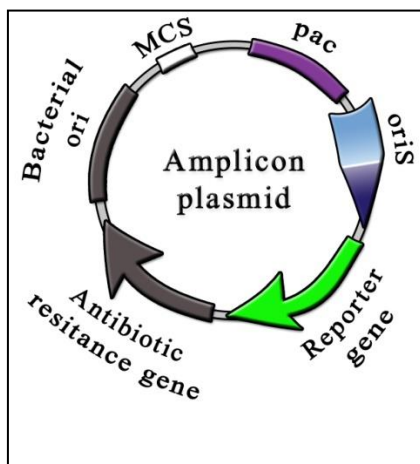


Figure 3. Structure of a typical amplicon plasmid.

Heterologous transcription units, either alone or in combination, can be cloned into amplicon plasmids using conventional molecular cloning techniques; the resulting construct is then packaged into viral particles to be then used for transduction of cells or tissues. From the structural point of view, amplicon vectors are virus particles structurally and immunological identical to wild-type HSV-1 particles, but bearing, instead of the virus genome, a concatemeric form of DNA amplified in a head-to-tail arrangement, which derives from the amplicon plasmid (Fig. 4).

Amplicon vectors have been used both on basic and applied research, and they have revealed as most suitable tools to study complex functions involving the nervous system, such as anxiety, sexual behaviour, learning and memory. In addition, amplicon vectors are being used for the development of new experimental gene therapy approaches (both for inherited and acquired diseases affecting the nervous system, including neurodegenerative diseases) [20-23],

vaccine development [24, 25], hybrid viral vectors (such as adeno-associated and lentivirus viruses) [26, 27], human artificial chromosome technologies [28, 29] and basic research. It is not possible to fully describe these applications in the short extent of this manuscript. For a more detailed description of amplicon vector applications, please refer to Cuchet *et al.* (2007) [30], Epstein (2009) [31], Fraefel (2007) [32], Manservigi *et al.* (2011) [33] and de Silva and Bowers (2009) [10].

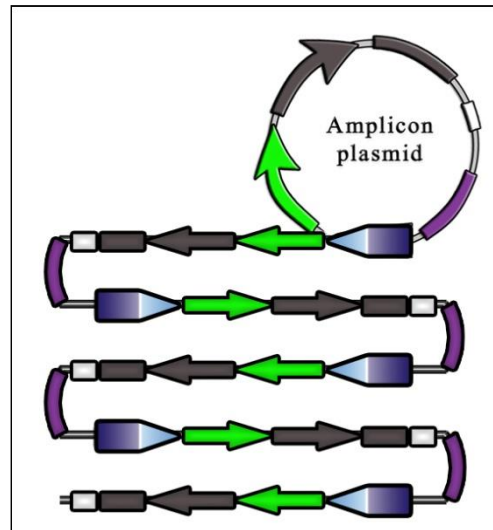


Figure 4. Concatemeric amplicon DNA replication.

Amplicon Vector Properties

The amplicon system has several properties that make it a reliable and efficient method for gene transfer.

1. The two HSV-1 elements required to support replication and packaging into virions, i.e., the *oriS* and *a* sequences, are smaller than 1-kbp. Since the HSV-1 particle can package up to 153-kbp, this means that the amplicon particle has the potential to accommodate very large fragments of foreign DNA, including multiple and/or large transgenes or large cell type-specific regulatory sequences. This is the most remarkable feature of amplicon vectors, as there is no other available viral vector system displaying the capacity to deliver such a large amount of foreign DNA (~150-kbp) to the nuclear environment of mammalian cells.
2. Depending on the size of the amplicon plasmid, several copies of the transgene can be packaged into a single vector particle [34], thereby increasing the transgene dose per infected cell. This happens because the amplicon genome is replicated via a mono-directional rolling circle-like mechanism, therefore generating long concatemers composed of tandem repeats of the amplicon plasmid, and because each HSV-1 amplicon particle always cleaves and packages a ~150-kbp DNA molecule, which represents the HSV-1 particle packaging capacity [35].

3. The HSV amplicon is not an 'alive' virus but an inert particle, bearing an inherently safe *in vivo* profile. Although amplicon genomes carry no virus genes and consequently do not induce synthesis of viral proteins, the approximately 40 different HSV-1 structural proteins that shape the HSV-1 virion (which are the same in the amplicon virion) are delivered into the cell during infection and can trigger cell signaling and cellular responses, probably having a transient impact on cell homeostasis and cellular gene expression. Nevertheless, these proteins soon disappear and the cells can resume their normal functions, including the ability to divide and to respond to physiological stimuli. Furthermore, the absence of viral genes in the amplicon genome strongly reduces the risk of reactivation, complementation or recombination with latent or resident HSV-1 genomes.
4. As already quoted, after primary HSV-1 infection of epithelial cells, the virus particles can enter sensory nerve endings and the capsid is moved via retrograde axonal transport to the neuronal cell bodies of the sensory ganglia, where the virus genome may establish a latent infection. Most of these mechanisms are certainly reproduced by amplicon vectors, since the mature amplicon particle is identical to that of the helper HSV-1 and they will thus behave as normal virion particles. However, it is not yet clear if the amplicon genome can establish a latency-like long-term expression in sensory neurons.
5. Once packaged into viral particles, the amplicon vector retains the ability to infect numerous cell types, and its genome is maintained in an episomal state within the nucleus of the infected cell. Due to the absence of viral gene expression, the amplicon is replication-defective, and its episomal existence results in stable maintenance in post-mitotic cells, but leads to unequal segregation in mitotically active cells. Since it does not integrate into the host cell genome, the conventional amplicon does not lead to insertional mutagenesis, thus increasing its safety profile as a gene therapy vector.

Amplicon Production Systems

Production of amplicon vectors involves the coordinated action of more than 50 viral proteins, required to allow replication and packaging of the amplicon plasmid into fully infectious virus particles. Amplicon plasmids are therefore dependent on helper HSV-1 virus proteins, and the genes encoding these proteins could in principle be supplied by a helper HSV-1, or by HSV-1 viral DNA.

The most important limitation for the use of amplicons in gene therapy trials is the difficulty to produce large, high-titer stocks of vector particles free of helper virus contamination. At present, there are two major methods used for producing amplicon vectors with low or even without HSV-1-helper-contamination (Fig. 5): one is based on transfection of the amplicon plasmid with HSV-1 helper genome (either as bacterial artificial chromosome (BAC) or as cosmid-based packaging systems), and the other, is based on transfection of the amplicon plasmid in cells expressing Cre recombinase, followed by infection using a defective helper HSV-1 carrying loxP sequences surrounding the packaging signals.

Helper DNA-Based Packaging Methods

DNA-based packaging methods for amplicon production are carried out by co-transfecting amplicon plasmids with HSV-1 genomes lacking their packaging signals. This system, initially derived from a set of five overlapping cosmids, each one carrying a large fragment of the virus

genome, which allows the reconstitution of a full virus genome by homologous recombination, in cells co-transfected with the full cosmid set [36]. By deleting the *a* sequences from the two cosmids carrying them, this system generates a virus genome that is able to express all the required transacting functions but which cannot itself be packaged into HSV-1 particles. The co-transfection of amplicon plasmids with the modified set of cosmids allows therefore the production amplicon vectors. Using this system, however, amplicons are produced in relative low amounts (10^5 - 10^6 TU/mL) [37], and the vector stocks can be barely contaminated with helper virus particles.

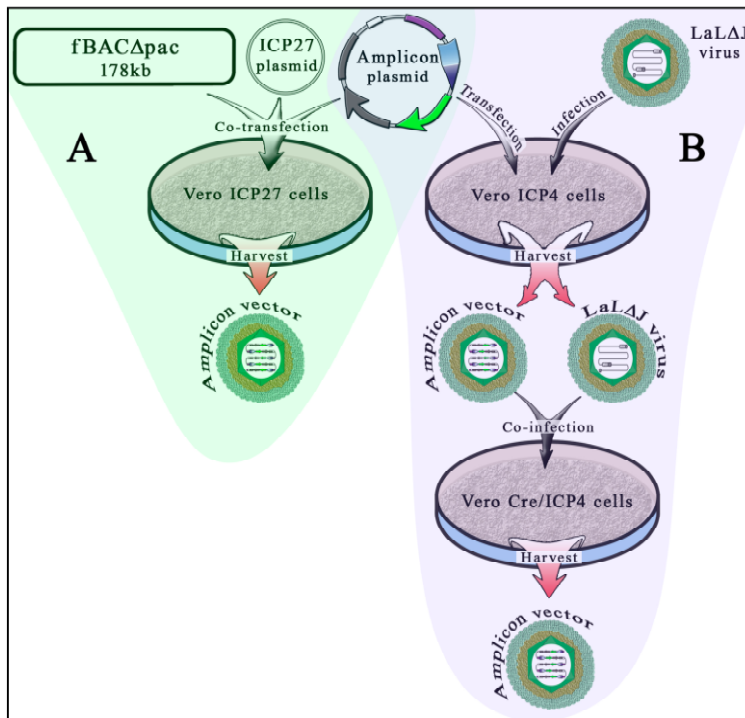


Figure 5. Amplicon production systems. (A) DNA-based packaging system: Vero ICP27-expressing cells are co-transfected with the amplicon plasmid, the fBACΔpac BAC (which carries a non-packageable HSV-1 genome) and an ICP27-expressing plasmid, and amplicon vectors are harvested from cells 2 or 3 days post-transfection. (B) Helper virus-based packaging system: Vero ICP4-expressing cells are transfected with an amplicon plasmid and super-infected with the HSV-1 LaLΔJ helper virus. At two days post-infection, the mixed population of viral particles (amplicons and helper particles) is used to infect ICP4/Cre recombinase-expressing cells. After 2 days, amplicon vectors, only barely contaminated with defective helper particles, are harvested from infected cells.

This approach was later simplified and significantly improved by cloning the entire HSV-1 genome, only devoid of *a* signals, into a bacterial artificial chromosome (BAC), thus generating a BAC-HSV-1 [38, 39]. In the last version of this system, the gene encoding the essential ICP27 protein was further deleted from the BAC-HSV-1 genome, which was then increased in size by adding non-coding DNA to further reduce the probability of being packed into newly assembled HSV-1 particles [39, 40]. The ICP27 protein is supplied *in trans*, both by a plasmid and by complementing cell lines expressing this protein (Fig. 5A). This system gave rise to the production of entirely helper-free amplicon stocks for the first time. Vector

titers obtained from helper virus-free amplicon packaging can range from 10^7 to 10^8 TU/ml, but the total amount of particles is somewhat limited by the fact that vector stocks produced in this way cannot be serially passaged.

Helper Virus-Based Packaging Methods

Historically, the first method to produce amplicon stocks was based in the transfection of cells with the amplicon plasmid, followed by super-infection with a helper HSV-1 virus, to supply *in trans* the necessary viral functions. One advantage of this system was that the vector/helper stock thus produced could be serially passaged, to produce as many amplicon particles as required. Although this method allows to produce large amounts of amplicon vectors, its major limitation is that it leads to a mixed population of particles, highly contaminated with HSV-1 helper particles, which induces strong cytotoxicity and immune responses upon infection of target cells or organisms, thus impeding their use in gene therapy and even in many fundamental studies. Due to the identical physical properties of both amplicon and HSV-1 helper virus particles, a selective purification of amplicon particles by physical treatment is not feasible.

Therefore, different strategies have been successively developed in order to reduce the toxicity of the helper-contaminated amplicon stocks, firstly by modifying the helper virus genome in order to limit its toxicity, and secondly, by limiting the production of contaminant particles. The first approach to improve this virus-based method was the use of a thermosensitive HSV-1 as helper [41]. Subsequently, defective helper HSV-1 viruses, carrying deletions in immediate early genes encoding one essential protein [42–45] were developed. However, and despite obtaining relatively high titers, these amplicon stocks were still contaminated with large amounts of defective HSV-1 helper particles, resulting in significant cytotoxicity and inflammation.

LaL and LaLΔJ HSV-1 Helper System

More recently, an attractive approach to produce amplicon vectors using the helper virus-based system, was developed to avoid or to limit helper genome packaging while producing high amounts of amplicon stocks. This system is based on the deletion of *a* sequences from the HSV-1 helper genomes by a Cre/loxP-based site-specific recombination, in order to abolish their packaging in the cells that are producing the amplicon vectors. The first of these helper systems, named HSV-1 *LaL* (for lox-*a*-lox) helper, carried a unique and ectopic *a* sequence flanked by two parallel loxP sites, located in the gC locus [46, 47]. This virus is therefore Cre-sensitive and cannot, in principle, be packaged in Cre-expressing cells due to deletion of the floxed *a* sequence. Nevertheless, some helper genomes could escape the action of the Cre recombinase, allowing the production of some helper particles, which are replication-competent and able to spread.

To further improve this helper system, the two genes surrounding the *a* sequence, encoding the virulence factor ICP34.5 and the essential protein ICP4, were deleted from the *LaL* genome, generating the *LaLΔJ* helper virus [48]. Although the amplicon stocks prepared with this helper, in a cell line trans-complementing both Cre and ICP4 proteins (Fig. 5B) still contain a very small amount of contaminating helper particles, the replication-incompetent *LaLΔJ* helper cannot spread upon infection of target cells or tissues. Use of the HSV-1 *LaLΔJ* helper virus generally results in the production of large stocks of amplicon vectors with titers often

exceeding 2×10^8 TU/mL, and only barely contaminated (about 1.0%) with defective, non-pathogenic helper particles. Nevertheless, the presence of contaminating helper virus, even at very low levels, could still be a limitation for the use of amplicon stocks in some gene therapy applications. Therefore, the approaches to produce amplicon vectors should still be improved in order to produce high amounts of high-titer of entirely non-toxic amplicon stocks.

Further Improvements of Amplicon Vectors Technology

Although several technological improvements for amplicon vectors production have been thus developed in the last decade, several difficulties remain still unresolved. These include: (i) cytopathic effects and immune responses induced by gene expression from potential helper virus contaminants; (ii) potential reversion of helper virus contaminants to the wild type HSV phenotype; (iii) potential interactions with residing HSV-1, eventually leading to complementation, reactivation and recombination; (iv) the inability of the vector genome to be maintained in proliferating cells; (v) the presence of bacterial DNA sequences, which can lead to inflammatory responses and to the silencing of transgene expression, and (vi) the already quoted inability to generate large amounts of high-titer vectors fully free of contaminant helper particles. To overcome some of these obstacles and to improve these vectors, we have recently introduced two modifications to the amplicon vector production system, corresponding to: (i) the amplicon genome, in order to fully delete bacterial sequences and (ii) the nature of the complementing cell lines where the amplicons are produced, in order to improve helper-free vector production. We briefly describe in the following paragraphs our ongoing program to improve the amplicon methodology.

Elimination of Bacterial DNA from the Amplicon Genome

It is known that the presence of bacterial sequences in transduced plasmids can induce silencing of the transgene cassette, as well as innate and inflammatory reactions. In addition, bacterial sequences are outlaw in gene therapy protocols. In the case of amplicons, it was recently demonstrated that the presence of bacterial sequences in the amplicon plasmid (and thus, also present in the amplicon vector genome) cause inflammatory responses and rapid transgene silencing, by forming inactive chromatin [49]. In this regard, it has been shown that infection with amplicon vectors lacking bacterial sequences (minicircle amplicon vectors) induced approximately 20-fold higher transgene expression due to transcriptional enhancement [49]. In addition, nude mice injected with minicircle amplicon vectors exhibited 10-fold higher luciferase expression than mice injected with conventional amplicons, detectable up to at least 28 days post-infection [49]. Elimination of these bacterial DNA sequences from the amplicon genome is therefore critical if the vectors will be used in gene therapy.

Our efforts to eliminate bacterial DNA from the amplicon genome are based on the modification of the amplicon plasmid. We have recently generated a new amplicon plasmid, named pOPNE (Fig. 6), which possesses a typical plasmidic backbone with a bacterial origin of DNA replication (*E. coli* ori) and an antibiotic resistance gene (AmpR). This plasmid also contains an amplicon cassette surrounded by two loxP (L) sites in parallel orientation. As usual, the amplicon cassette carries one HSV-1 oriS and one pac (or *a*) signal. Note that the oriS sequence is linked to the HSV-1 IE4/5 promoter (IE4/5), very close to one of the loxP sites. pOPNE also contains a Neomycin/Kanamycin resistance gene (Neo) bordered by two FRT sites (F). This FRT-bordered cassette will serve as an acceptor locus for the introduction of transgenes through FRT-flipase site-specific recombination. Finally, the amplicon cassette also

possesses a promoterless EGFP (enhanced-GFP) reporter gene, juxtaposed to the second loxP site.

Preliminary results (not shown here), indicate that when the pOPNE plasmid is transfected in Cre recombinase-expressing cell lines, the amplicon plasmid recombines at the loxP sites, thereby producing two molecules: one minicircle carrying the bacterial sequences present in the plasmid backbone (residual DNA) and a second minicircle carrying the complete amplicon module devoid of bacterial sequences (Fig. 7). In addition, the Cre recombinase-mediated deletion of the bacterial sequences brought the IE4/5 promoter upstream the promoterless-EGFP open reading frame (ORF), therefore leading to EGFP transcription. Following infection of these cells with helper HSV-1, the amplicon genome was then encapsidated into HSV-1 particles. These particles expressed EGFP upon infecting target cells, therefore demonstrating that they have lost the bacterial sequences between the promoter and the EGFP ORF. These vectors are therefore appropriate for further used in gene transfer approaches. We are currently investigating the biological properties of these bacterial DNA-free amplicon vectors.

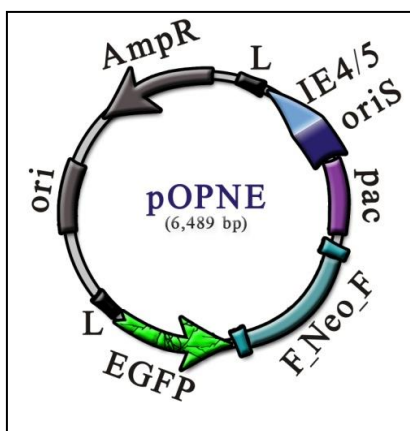


Figure 6. Structure of pOPNE amplicon plasmid. *ori*: bacterial origin of DNA replication; *AmpR*: Ampicillin resistance gene; *L*: loxP sites; *oriS*: HSV-1 origin of DNA replication; *pac*: HSV-1 cleavage/packaging-encapsidation sequence; *IE4/5*: HSV-1 IE4/5 promoter; *F_Neo_F*: Neomycin resistance gene bordered by two FRT sites (F); and *EGFP*: EGFP reporter gene, without promoter.

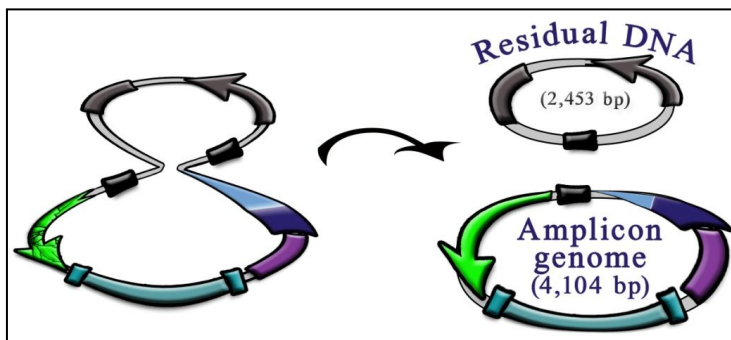


Figure 7. Cre recombinase-mediated cleavage and dissociation of pOPNE plasmid. Upon Cre recombinase expression, the amplicon plasmid pOPNE is dissociated into two DNA molecules: one carries the residual bacterial DNA and another carries the amplicon sequences devoid of bacterial DNA.

Modification of the Complementing Cell Lines

VCre4 Cell Line

As already quoted, when using HSV-1 *LaLΔJ* as helper virus, the helper particles are eliminated through a Cre/loxP-based site-specific recombination event that requires thus the expression of Cre-recombinase in the infected cells, in addition to the expression of the essential protein ICP4, whose gene is absent in this helper system. Previous cell lines expressing simultaneously ICP4 and Cre recombinase, that were used for the production of amplicon stocks with HSV-1 *LaLΔJ* helper, were derived from TE-671 cells (human rhabdomyosarcoma cells, ATCC CRL 8805) [46]. As these cells are cancer-derived cell lines, they cannot be used to produce vectors for gene therapy. In contrast, Vero and Vero-derived cell lines are approved for gene therapy approaches by the Food and Drug Administration (FDA - U.S. Department of Health and Human Services) [50].

Therefore, we have recently constructed another cell line, derived from Vero cells, which we named VCre4. These cells express Cre recombinase under the control of the HCMV promoter and the ~10.4-kbp ICP4 locus including its own promoter (HSV-1 bases 126,764-131,731). Preliminary results have shown that VCre4 cells produce higher levels of amplicon vector stocks than the previous system and these stocks contains lower than 1% of contaminant defective helper particles (data not shown here). Therefore, not only VCre4 cells provide a better packaging cell system but, in addition, the creation of this novel Vero-based cell line, brings amplicon vector technology closer to gene therapy protocols and represents a promising approach to treat neurodegenerative diseases.

AMPLICON VECTORS AND CANCER

In this section we present a very brief summary about some past and recent applications of HSV-1 derived vectors to cancer research and experimental therapy.

HSV-1-amplicon vectors have been tested for treatment of different experimental tumours, both in brain and in other organs and tissues, exploiting its inherent absence of toxicity and ability to infect many different cell types. Since these vectors can efficiently deliver transgenes to cancer cells, but they are diluted during successive cell divisions, most studies have used acute approaches by delivering pro-drug activating enzymes, cytotoxic proteins, apoptosis-inducing factors, fusogenic proteins and small interfering RNAs (siRNAs) for growth factor receptors. In addition, several studies have targeted amplicon vector entry or expression in order to improve treatment efficacy and selectivity.

Targeting tumour cells using HSV-1 vectors is a complicated issue as the process involves multiple interactions of viral envelope glycoproteins and cellular host surface proteins. In order to improve the selectivity of amplicon vectors tropism for tumour cells, such as glioma cells, the heparan sulfate-binding domain of glycoprotein C (gC) has been replaced with a human glioma-specific peptide sequence (MG11). Preferential homing of these virions in glioma cells was confirmed in a xenograft glioma mouse model, following intravascular delivery [51]. Another important issue is the inefficient distribution of vector particles *in vivo*, which may limit their therapeutic potential in patients with gliomas and, in this context, it has been recently

demonstrated that vector-mediated gene expression in gliomas was strongly dependent on vector application-injection pressure and injection time [52].

In order to improve efficiency and safety of cancer gene therapies, efforts have been made at specifically targeting proliferating cells in glioma models. The HSV-1 immediate-early gene ICP0 possesses E3-ubiquitin ligase activity [53] and it can induce the degradation of centromeric proteins [54]. Amplicons expressing the HSV-1 ICP0 were used to infect human glioblastoma Gli36 cells and well-established models of non-dividing cells, such as primary cultures of either rat cardiomyocytes or brain cells. ICP0 induced a strong cytostatic effect and significant cell death in Gli36 cells. In contrast, neither cell death nor any evidence of ICP0-induced toxicity was evident in both primary cultures of non-cycling cells. These observations suggest that ICP0 has gene therapy potential and would be the first member of a new family of cytostatic proteins that could be used to treat cancer [55].

A different approach used amplicons expressing siRNA in order to mediate post-transcriptional silencing of molecules involved in the pathogenesis of cancer or connected with the radiation resistance of tumour cells. Infected human glioblastoma cells with knockdown for the epidermal growth factor receptor (EGFR) expression displayed growth inhibition, both in culture and in athymic mice [56]. Besides, the combination of vector-mediated silencing of Rad51 expression (which is a key component of the homologous recombination repair of DNA double-strand breaks) and treatment with ionizing radiation, resulted in a pronounced reduction of human glioma cells survival in culture and in a significant decrease in tumour size in athymic mice [57].

Another tested strategy was to target tumour cells via transcriptional control of therapeutic genes. Ho and co-workers constructed a glioma-specific and cell cycle-regulated amplicon carrying the GFAP enhancer/promoter element, plus a cell cycle specific regulatory element from the cyclin A promoter [58]. Transgenic activity was mediated in a cell type-specific and cell cycle-dependent manner, both *in vitro* and *in vivo* in glioma-bearing animals [59]. Anti-tumour efficacy of this vector system was assessed using the pro-apoptotic proteins Fas ligand (FasL) and the Fas-associated death domain (FADD), inducing cell death in proliferating primary human glioma cells derived from patients, and resulting in prolonged survival of mice bearing orthotopic gliomas [60]. The authors pointed out that these vectors are stable, elicit minimal immune response and are not significantly hampered by chemotherapy or irradiation *in vivo* [58].

To achieve schwannomas regression without injury to associated neurons, it was generated an HSV-1 amplicon vector, in which the apoptosis-inducing enzyme caspase-1 (ICE), was placed under the Schwann cell-specific P0 promoter. Following direct intra-tumoral injection of the P0-ICE amplicon vector in an experimental xenograft mouse model (tumours were formed by subcutaneous injection of an immortalized human schwannoma cell line into one or both flanks of immunodeficient mice), there was a marked regression of schwannoma tumours. Injection of the same amplicon vector into the sciatic nerve produced no apparent injury to the associated dorsal root ganglia neurons or myelinated nerve fibers. Hence, the P0-ICE amplicon vector provides a potential means of 'knifeless resection' of schwannoma tumours by injection of the vector into the tumour, with low risk of damage to associated nerve fibers [61].

AMPLICON VECTORS AND VACCINES

In this section we present a brief summary about some applications of HSV-1 derived vectors as potential vaccines.

Wild-type HSV-1 is a potent immunogen that can elicit both host innate and adaptive immune responses [62], though is also able to evade host immunity due to immuno-modulatory gene products such as the ICP47 protein [63]. The biological characteristics of HSV-1 amplicons that make them an attractive candidate for vaccine delivery applications are: its safety profile, the lack of expression of viral immuno-evasive genes, large insert capacity and ability to infect a wide range of cells, including antigen-presenting cells. So far, immunization studies involving amplicons have been essentially directed against viral pathogens and neurological disease factors.

Virus-like particles (VLPs) are promising vaccine candidates, because they represent viral antigens in the authentic conformation of the virus particles and are, therefore, readily recognized by the immune system. As VLPs do not contain genetic material, they are safer than attenuated virus vaccines. In a first study, HSV-1 amplicon vectors were constructed to coexpress the rotavirus (RV) structural genes VP2, VP6, and VP7, and were used as platforms to launch the production of RV-like particles (RVLPs) in vector-infected mammalian cells. HSV-1 amplicon vectors launching the production of heterologous rotavirus-like particles were able to induce rotavirus-specific immune response in mice [25]. In a second study, amplicons expressing Foot-and-mouth disease (FMD) virus antigens were used to generate a genetic vaccine prototype, based in the *in situ* generation of FMD-VLPs [64].

Amplicons have been also used to express antigens and elicit immune responses in the context of veterinary diseases. In particular, specific serum antibody responses were detected in mice inoculated with amplicon vectors that expressed the glycoprotein D (gD) of bovine herpesvirus [24].

Most studies focused on immunization against HIV. Hocknell and colleagues showed that a single inoculation of 1×10^6 transducing units of amplicons expressing the HIV-1 envelope glycoprotein (gp120), was able to elicit strong, antigen-specific and long-lasting (≤ 5 months) cellular and humoral responses in mice [65]. Subsequent studies demonstrated that a combined heterologous prime-boost immunization approach using naked plasmid DNA prime and amplicon boost expressing the gp120 antigen could enhance by 15- to 20-fold the amplicon induced memory T-cell responses [66]. Gorantla and co-workers used non-obese diabetic/severe combined immunodeficient mice repopulated with human peripheral blood lymphocytes, as *in vivo* model for HIV immunization [67]. Injecting autologous dendritic cells transduced *ex vivo* with a gp120-expressing amplicon into these mice, resulted in primary HIV-1-specific humoral and cellular immune responses. The same authors demonstrated that there also was a partial protection against infectious HIV-1 challenge in the immunized mice [67]. Lastly, in an attempt to optimize amplicons for HIV vaccine delivery, Santos and collaborators showed that amplicons expressing the HIV-1 *gag* gene driven by two different HCMV hybrid promoters (HCMV/long-terminal repeat or HCMV/woodchuck post-regulatory element), elicited the strongest immune response in mice [68].

Other studies have explored the potential of amplicons in immunotherapy for prion disorders. Bowers and collaborators described the construction of amplicons expressing domains of PrP C fused to tetanus toxin Fragment C as adjuvant, as a base for prion disorders

immunotherapy [69]. Vaccination generated significant levels of prion protein (PrP)-specific antibodies in mice, but none of the constructions was able to alter disease progression.

Since amplicons can elicit strong antigen-specific immune responses to viral or mammalian proteins, they might be used in prophylactic vaccination against pathogens or neurodegenerative diseases. However, due to the high prevalence of HSV infection within the human population, one major concern about their use as vaccines is the possible impact of pre-existing antiviral immunity on vaccine efficiency. This question remains controversial. Some studies using replication-defective HSV-1 showed strong immune responses even in the presence of anti-HSV immunity [65, 70]. In contrast, another study with a replication-defective HSV-1, showed a substantial reduction in immune responses in animals previously vaccinated against HSV-1 [71].

AMPLICON VECTORS AND BEHAVIOUR

Amplicon vectors have been delivered into distinct brain regions to investigate complex aspects of the normal functioning of the central nervous system (CNS). Amplicon vectors designed to modify protein expression may carry sequences that allow overexpression or knockdown of an endogenous protein, or expression of an exogenous protein (e.g., dominant-negative mutants) relevant for CNS functions, like proteins directly involved in neurotransmission and in neuron signaling. A comprehensive review on the use of amplicons vectors to study behaviour can be found in Jerusalinsky and co-workers (2012) [20].

Different challenges to find causal relationships between neuronal molecular mechanisms and learning and memory processing, have been solved by the use of amplicon vectors. These vectors were used, for example, to investigate the involvement of NMDA glutamate receptor (NMDAR) in learning and memory. Amplicons carrying sequences in either sense or antisense orientations of the GluN1 subunit gene, in addition to EGFP as the reporter gene, were used to investigate the participation of hippocampal NMDAR by modifying the expression of that essential GluN1 subunit in the rat CNS. The ability to modify endogenous levels of GluN1 was first tested *in vitro* in primary cultures of neurons from rat embryo cerebral cortex [72, 73] and then assessed *in vivo*. Adult rats inoculated into the dorsal hippocampus with vectors expressing GluN1 antisense performed significantly worse than control rats in an inhibitory avoidance of a footshock task, and did not show habituation (decreased exploratory behaviour) by repeated exposure to an open field. Immunohistochemistry in serial brain slices from these animals, showed that the transduced cells represented approximately 6–7% of hippocampal pyramidal neurons in CA1 region and just about less than 1% of granule cells in the dentate gyrus, indicating that the knockdown of a single gene in a small number of those neurons significantly impaired memory [74].

Amplicon vectors expressing a constitutively active catalytic domain of the protein kinase C (PKC) β -II were used to transduce rat hippocampal dentate granule neurons. Activation of PKC pathways in a small percentage of these neurons was sufficient to enhance rat auditory discrimination reversal learning, suggesting an hippocampal auditory mediated learning in the rat [75].

Amplicon vectors have been used to elucidate the role of the alpha-amino-3-hydroxy-5-methyl-4-isoxazole-propionic acid receptor/channel (AMPA) in learning and memory

processes. Most endogenous AMPARs contain the GluR2 subunit and both inward and outward currents can pass through equally well. In contrast, AMPARs lacking GluR2 exhibit a profound inward rectification, what means that minimal outward current could pass through these channels when the cell depolarizes. Thus, incorporation of recombinant AMPARs into synapses can be monitored functionally [76]. To study the AMPARs trafficking in associative fear conditioning learning, Rumpel and co-workers injected amplicon vectors expressing recombinant AMPARs subunits into the rat basolateral amygdala [77]. An amplicon vector encoding GluR1 subunit fused with GFP was used to express homomeric AMPARs, which display greater inward rectification than endogenous AMPARs, allowing electrophysiological detection of synapses undergoing plasticity by incorporation of recombinant GluR1-AMPA. Another amplicon vector encoded the carboxyl cytoplasmic tail of GluR1 subunit fused with GFP, expressing a protein that acts as a dominant-negative construct to prevent synaptic incorporation of endogenous GluR1-receptors, and thereby blocks several forms of synaptic plasticity *in vitro* and *in vivo*. A third vector driving expression of GFP only, was used as control of infection. Using these vectors, they showed that fear conditioning drives AMPARs into the synapse of a large fraction of post-synaptic neurons in the basolateral amygdala, and that blocking GluR1-receptor trafficking in a few (~10 to 20%) neurons undergoing plasticity, was sufficient to impair memories depending on this structure [77].

The cyclic adenosine monophosphate (cAMP) response element-binding protein (CREB) plays an important role in learning and memory processes [78]. It was shown that changes in CREB function could influence the probability of individual lateral amygdala neurons to be recruited into a fear memory trace, suggesting a competitive model underlying memory formation. In such a model, eligible neurons are selected to participate in a memory trace as a function of their relative CREB activity in learning [79]. In this regard, several investigators have manipulated the function of CREB using amplicons vectors. An amplicon encoding wt-CREB was used to show that increasing CREB in the auditory thalamus enhanced memory and generalization of auditory conditioned fear, indicating that CREB-mediated plasticity in the thalamus plays a role in this cognitive process [80]. Using an amplicon vector encoding a dominant-negative mutant form of CREB (mCREB), Brightwell and collaborators demonstrated that hippocampal overexpression of mCREB can block long-term – though not short-term – memory for a socially transmitted food preference, therefore involving hippocampal CREB function in this type of memory [81]. They later showed that rats trained to make a consistent turning response in a water version of the plus maze, required CREB function in the dorsolateral striatum to form a long-term memory of the response strategy, and showed that it is independent on CREB function in the dorsal hippocampus [82]. Using a model of protracted social isolation in adult rats, Barrot et al. observed an increase in anxiety-like behaviour and in both the latency of the onset of sexual behaviour and the latency to ejaculate [83]. Then, in transgenic cAMP response element (CRE)-LacZ reporter mice, which express β -gal under control of CREs, they showed that protracted social isolation also reduced CRE-dependent transcription within the nucleus accumbens. This decrease in CRE-dependent transcription was mimicked in non-isolated animals by amplicon-based gene transfer of a dominant-negative mutant of CREB into the nucleus accumbens. In these animals, the local inhibition of CREB activity increased anxiety-like behaviour and delayed initiation of sexual behaviour, with no effect observed on the ejaculation parameters. In isolated animals, restoring CREB activity in the nucleus accumbens using amplicon vectors to overexpress wt-CREB, rescued the anxiety phenotype as well as the deficit in the latency to initiate sexual behaviour.

This study suggests a role for the nucleus accumbens in anxiety responses and in specific aspects of sexual behaviour, and provides novel insights into the molecular mechanisms by which social interactions affect brain plasticity and behaviour [83].

Interestingly, Liu and co-workers recently reported that an amplicon expressing siRNA for the γ -Aminobutyric acid A (GABA-A) receptor $\alpha 2$ subunit, infused into the central nucleus of the amygdala of alcohol-preferring rats, (*i*) caused profound and selective reduction of binge drinking associated with inhibition of $\alpha 2$ subunit expression, (*ii*) decreased GABA-A receptor density and (*iii*) inhibited Toll-like receptor 4 (TLR4) expression [84]. Furthermore, infusion of an amplicon expressing TLR4 siRNA into the central amygdala also inhibited binge drinking, but did not cause such changes when infused into the ventral pallidum. On the other hand, binge drinking was effectively inhibited by an amplicon expressing GABA-A receptor $\alpha 1$ subunit siRNA, infused into the ventral pallidum, showing that TLR4 contributes to binge drinking downstream to $\alpha 2$ subunit in the central amygdala, but not in the ventral pallidum, underscoring the relevance of TLR4 in specific neuroanatomical sites. Those data indicate that GABA-A $\alpha 2$ -regulated TLR4 expression in the central amygdala contributes to binge drinking and may be a key for early neuroadaptation in excessive drinking [84].

AMPLICON VECTORS AND NEURODEGENERATIVE DISEASES

Ataxias

Ataxias are a group of specific degenerative diseases of the nervous system characterized by the loss of movement coordination. Hereditary ataxias could be recessive or dominant. Nowadays, no treatment can effectively stop progression of ataxias.

The most prevalent form of hereditary ataxia is the Frederich's ataxia (FA) [85], where neurological symptoms result from neurodegeneration in the dorsal root ganglia (DRG), with loss of large sensory neurons and posterior columns, followed by degeneration in the spinocerebellar and corticospinal tracts of the spinal cord. This disease is caused by a mutation in the first intron of the frataxin gene (*frda wt*) which causes a large GAA repeat expansions (*frda mut*) and, in consequence, leads to a decrease in the level of the mitochondrial protein, frataxin.

Lim and collaborators generated a conditional frataxin transgenic mouse with the *frda wt* gene floxed (*loxP-frda*); then, they injected an amplicon vector carrying the CRE recombinase transgene into the brainstem of this transgenic *loxP-frda* mouse [86]. CRE expression causes the homologous recombination of loxP sites and the loss of *frda*, leading to a decrease in frataxin expression. These mice developed behavioral deficits after 4 weeks, resembling FA. In an attempt to generate a treatment for FA, the authors injected a second amplicon vector, bearing the cDNA of *frda wt*. These mice exhibit behavioral recovery as early as 4 weeks after the second vector injection [86].

Taking advantage of the large capacity of amplicon vectors, Gomez-Sebastian and collaborators used amplicons to deliver a 135-kb insert containing the entire *frda wt* human genomic locus, including long upstream and downstream regulatory sequences (~80-kb), to fibroblasts extracted from FA patients (which expressed low levels of frataxin) [87]. Synthesis of frataxin in these *frda wt*-transduced FA-deficient cells was confirmed by

immunofluorescence [87]. The fibroblasts of FA patients have been shown to exhibit biochemical deficits, including increased sensitivity to oxidative stress. Functional complementation studies demonstrated restoration of the wild type cellular phenotype in the *frda* wt-transduced cells, in response to H₂O₂ treatment (a classical stressor) [87].

To investigate the persistence of the transgene expression, the same group injected into the adult mouse cerebellum an iBAC-HSV-1-based vector, carrying the 135-kb *frda* wt genomic DNA locus [88]. As reporter, the authors constructed another vector, but with the *E. coli lacZ* gene inserted at the ATG start codon (iBAC-*frda-lacZ*). Direct intracranial injection of this vector into the adult mouse cerebellum resulted in a large number of cells expressing *lacZ*; this expression was driven by the *frda* wt locus and persisted for at least 75 days. In contrast, synthesis of GFP expressed from the same vector, but driven by the HSV-1 IE4/5 promoter, was strong but transient. This study demonstrated for the first time, a sustained transgene expression *in vivo*, by amplicon delivery of a very long genomic DNA locus. All together these results suggest the potential of the HSV-1 derived-*FRDA* vectors for gene therapy of FA.

Ataxia-telangiectasia is an autosomal recessive neurodegenerative disorder due to mutations in the A-T gene (*AT*). This gene is, in the AT wt form, a kinase responsible for recognizing and correcting errors in current DNA duplication before cell division. Several mutations led to an AT-mutated: (ATM) protein with different degrees of activity, generating a pleiotropic phenotype characterized by cerebellar degeneration, immunodeficiency, cancer predisposition, increased radiation sensitivity and premature aging, according to the residual kinase activity of the expressed ATM protein [89]. As the *AT* cDNA is too large (~9-kb) for most of the currently used vectors, it was difficult to rescue the phenotype of cells expressing ATM. Therefore, Cortes and co-workers used an amplicon vector codifying the *AT* cDNA, to express a non-mutated form of AT in human fibroblasts which express ATM [90]. The AT wt protein expression was confirmed by western blot and immunofluorescence. AT kinase function was tested by *in vitro* phosphorylation of p53; the *in vivo* functionality of AT was assessed by counting the accumulation of G2/M cells after ionizing radiation [90].

In a further study, the same group constructed an amplicon encoding both the EGFP and a human FLAG-tagged-AT protein; this vector was inoculated in the cerebellum of AT^{-/-} mice. The number and phenotype of infected cells were assessed by EGFP fluorescence and infection of thousands of cells at the inoculated cerebellum, including Purkinje cells, was confirmed. FLAG-tagged-AT expression was demonstrated at transcriptional (qRT-PCR, *in situ* hybridization) as well as at translational (immunoprecipitation of the full-length human protein) levels, 3 days post-inoculation [91].

In an attempt to achieve stable gene replacement, the same group generated an HSV/adeno-associated virus (AAV)-hybrid amplicon, carrying the expression cassette for the AT and EGFP cDNAs, flanked by AAV inverted terminal repeats (ITRs). In the presence of the AAV Rep proteins (proteins codified by AAV genome, required for AAV chromosome integration), this hybrid vector mediated a site-specific integration of the transgenic sequences into the AAV1 site of chromosome 19. To test functional activity of this AT vector, Cortes *et al.* exposed AT infected cells to ionizing radiation; then, AT activity was assessed by specific immunofluorescence using antibodies against the phosphorylated form of AT (pAT, activated form). An increase in pAT was observed only in AT-AAV/HSV infected cells and not in cells infected with a control vector [92]. These results showed that this AT HSV/AAV hybrid amplicon was able to integrate into the AAVS1 site and to achieve functional expression of human AT cDNA *in vivo*, in a mouse model of ATM.

Alzheimer Disease

One of the most studied neurodegenerative diseases is Alzheimer's disease (AD). In this pathology, the peptide known as amyloid beta ($A\beta$) acts as a neurotoxin that produces neurodegeneration. More precisely, a recently enunciated hypothesis states that soluble oligomers of $A\beta$ peptide (named ADDLs: $A\beta$ -derived diffusible ligands) bind to neurons, mainly at the post-synaptic side, and that this binding would be responsible for triggering toxic effects that ultimately lead to neuronal death [93, 94].

$A\beta$ peptide is generated by degradation of the Amyloid Precursor Protein (APP). Under physiological conditions, APP is first cleaved by α -secretase and subsequently cleaved by γ -secretase, resulting in a non-amyloidogenic soluble peptide. However, under certain abnormal conditions or by blocking the normal degradation pathway, APP is cleaved by the β -secretase BACE-1 (instead of α -secretase) and then by the γ -secretase, generating amyloidogenic peptides 1-40 and 1-42, (of 40 and 42 amino acids respectively, being $A\beta$ 1-42 more amyloidogenic than $A\beta$ 1-40) [95]. $A\beta$ initially aggregates in soluble oligomers of 2–14 monomers (ADDLs), which can bind to the post-synaptic densities; then, as the concentration rises up, they further aggregate into fibrils in the extracellular space and then form the typical amyloid plaques [94, 96-100].

Two studies have described the use of amplicons for " $A\beta$ vaccination" in mice as a possible therapeutic strategy for AD, aimed at preventing $A\beta$ fibrillogenesis and/or to enhance removal of parenchymal amyloid deposits. In the first study, transgenic Tg2576 mice, which overexpress APP with the Swedish mutation that results in enhanced generation and extracellular deposition of the $A\beta$ 1–42 peptide, were injected with amplicons expressing either $A\beta$ 1-42 (HSVA β) or $A\beta$ 1-42 fused to the molecular adjuvant tetanus toxin Fragment C (HSVA β /TtxFC). Peripheral administration of both "vaccines" augmented humoral responses to $A\beta$ and reduced CNS $A\beta$ deposition in this model of AD. However, HSVA β vaccination was found to be toxic, since it induced expression of pro-inflammatory transcripts within the mouse hippocampus [101].

Another amplicon vector -HSV(IE) $A\beta$ (CMV)IL-4-, was constructed to co-delivered $A\beta$ 1-42 and interleukin 4 (IL-4), a cytokine that promotes the generation of Th2 like T-cell responses. Triple transgenic AD (3XTg-AD) mice, which progressively develop both amyloid and neurofibrillary tangle pathology, were vaccinated with that vector or with a set of control amplicon vectors (one encoding $A\beta$ 1–42 but not IL-4, and one "empty" vector without $A\beta$ 1–42 or IL-4 transgenes). Prevention of AD-related amyloid and tau pathology progression were significantly more important in HSV(IE) $A\beta$ (CMV)IL-4 treated mice than in control groups. Furthermore, the expression of Th2-related $A\beta$ -specific antibodies appeared to improve learning and memory in the Barnes Maze spatial memory paradigm [102]. Therefore, these results underlined the potential of amplicons for $A\beta$ immune-therapy of AD [103].

Neurodegenerative pathologies as AD or Parkinson's disease (PD), as well as some forms of depression, have been associated to dysfunction of receptor-neurotransmitter systems. L-glutamate is the major excitatory neurotransmitter in the CNS. Therefore, glutamate receptors represent an attractive molecular target in the treatment of neurodegenerative diseases and also in epilepsy, schizophrenia and ischemia.

There is recent evidence showing that the transmembrane protein APP appears capable of interacting with NMDAR [104, 105]. These ionotropic receptors are tetramers made of two GluN1 subunits and different GluN2 (A-D), and/or GluN3 (A-B) subunits, with GluN1 being

essential for receptor assembly [106-108]. Nowadays, association of NMDAR with several neuropathologies has been continuously growing up. Thus, the generation of novel tools that modify expression and composition of NMDARs should help to understand both the normal functioning (as reported above in “*Amplicon vectors and behaviour*” section) and the physiopathology of these receptors. It was proposed that ADDLs binds to NMDAR or to post-synaptic complexes containing it, acting as gain of function ligands [93, 94, 109]. By targeting such post-synaptic complexes, ADDLs activate a cascade of signals that lead to an increase in intracellular reactive oxygen species (ROS) [93]. Decker and colleagues have demonstrated that blockade of GluN1 expression through the infection of primary cultured neurons with amplicon vectors encoding an anti-GluN1 antisense RNA, inhibited ADDLs binding to synapses [109]. In the same study, they showed that there was a great reduction in ADDL-instigated ROS formation in GluN1 knockdown neurons [109]. In addition, it has recently been reported that different regulatory GluN2 subunits would also be involved in the binding of ADDLs to synaptic sites [110, 111]. Liu and colleagues have suggested that increasing the activity of GluN2A and/or reducing that of GluN2B, may alter or reduce the expression of cytotoxic effects mediated by ADDLs in neuronal cultures [110]. On the other hand, Balducci and colleagues showed that there was an alteration in the trafficking of GluN2A and GluN2B subunits in mutant mice expressing an amyloidogenic human form of APP [111]. However, more precise studies on the possible specific interaction between different subunits of the NMDAR and the A β peptide, both in normal and pathological conditions, is necessary to further clarify the specific site for ADDLs binding. It should be taken into account that the decrease in GluN1, which is essential for assembly and for the membrane allocation of the receptor, would lead to a decrease of all the NMDAR subunits at the synaptic surface [112].

The genes for microtubule-associated protein tau (*MAPT*) and α -synuclein (*SNCA*) play central roles in neurodegenerative disorders. Peruzzi and colleagues have recently generated amplicon-based iBAC-vectors carrying either the 143-kb *MAPT* or the 135-kb *SNCA* loci [113]. They have used these vectors to study regulation of gene expression of both *MAPT* and *SNCA*, showing functional complementation between them in cultured neurons and in organotypic brain slices. Infected neurons were able to express functional genes under physiological regulation, including the generation of multiple splice variants. In particular, multiple *MAPT* transcripts were expressed under strict developmental and cell type-specific control. Primary cultures from *Mapt*($-/-$) embryos had been shown to be resistant to A β peptide-induced toxicity, suggesting that the tau protein might be mediating the neurotoxicity of A β . To test the functionality of the *MAPT* transgene, the authors examined whether the responsiveness to A β peptide could be restored in the *Mapt*($-/-$) neurons and in organotypic brain slices. In both preparations from *Mapt*($-/-$) mice infected with the vector bearing the *MAPT* transgene, the tau protein was expressed as detected by ELISA and immunocytochemistry, and the sensitivity of *Mapt*($-/-$) neurons to A β peptide was restored [113]. As stated by the authors, the faithful retention of gene expression and phenotypic complementation by this system provides a novel and powerful approach to analyze neurological disease genes.

Parkinson Disease

Parkinson's disease (PD) is the second most common neurodegenerative disorder. A typical feature of this disease is the progressive loss of dopaminergic neurons in the substantia nigra

and a decrease in the level of dopaminergic inputs to the striatum [114]. PD is clinically characterized by akinesia or bradykinesia, rigidity and tremor that are directly related to dopaminergic striatal loss [115]. Several studies have used amplicons in experimental settings of PD. During and co-workers were the first to report the use of amplicons to deliver human tyrosine hydroxylase (TH) into the partially denervated striatum of 6-hydrodopamine-lesioned rats, used as a model of PD [116]. Efficient behavioral and biochemical recovery has persisted for one year after gene transfer.

In an attempt to improve that vector, Sun and collaborators developed another vector able to co-express two enzymes involved in the synthesis of L-DOPA: TH and the aromatic amino acid decarboxylase (AADC), under a modified neurofilament gene promoter that supports long-term expression in forebrain neurons [117]. This improved vector was injected in the right striatum of adult rats where PD-like symptoms had been previously induced by injection of 6-OHDA in the substantia nigra. Histologic analyses demonstrated neuronal-specific coexpression of TH and AADC from 4 days to 7 months after gene transfer. In these rats, vector injection was able to correct in about 80% the apomorphine-induced rotational behaviour present in this 6-OHDA rat model of PD [117]. Later on, in a series of elegant studies, Sun and co-workers further compared the activities of tissue-specific promoters to drive gene expression, particularly the promoters of TH, the neurofilament and the vesicular glutamate transport 1 (VGLUT1) [22, 118-122]. Taking advantage of the large transgene capacity of HSV derived vectors, they developed two vectors: one that co-express the three dopamine biosynthetic enzymes (TH, GTP CH I, and AADC, a 3-genes-vector) and another carrying all the three dopamine biosynthetic enzymes and the vesicular monoamine transporter (TH, GTP CH I, AADC, and VMAT-2, a 4-genes-vector). The authors compared the effects of both vectors. The 4-genes-vector supported higher levels of correction of apomorphine-induced rotational behaviour than did the 3-genes-vector, and this correction was maintained for 6 months. Proximal to the injection sites, the 4-genes-vector, but not the 3-genes-vector, supported extracellular levels of dopamine and dihydroxyphenylacetic acid (DOPAC) similar to those observed in normal rats, and only the 4-genes-vector supported significant K⁺-dependent release of dopamine [118].

The effect of amplicon-mediated transduction of the dominant-negative fibroblast growth factor (FGF) receptor-1 mutant protein (FGFR1(TK-)) into the rat substantia nigra, was evaluated *in vivo* as a possible strategy to mimic the reduced FGF signaling already documented to occur in PD. Following intra-nigral delivery of the FGFR1(TK-) amplicon, the number of substantia nigra neurons expressing TH was significantly reduced, leading to the conclusion that reduced FGF signaling in the substantia nigra of Parkinsonian patients could play a role in the impaired dopaminergic transmission in PD [123]. In a further study, the same group analyzed the effect of *ex vivo* transduction of mesencephalic reagggregates with the anti-apoptotic protein bcl-2, on grafted dopamine neuron survival. Using an amplicon expressing bcl-2 under the control of the TH promoter (HSV-TH9bcl-2) to transduce mesencephalic reagggregates, it was shown that, in spite of the high efficiency of the infection (since many cells were effectively transduced), amplicon-mediated overexpression of bcl-2 did not lead to an increase in grafted TH-immune-reactive neuron number [124].

Mitochondrial alterations are detected in most neurodegenerative disorders and may contribute to the dysfunction and demise of neurons. Rotenone or 1-methyl-4-phenyl-1,2,3,6-tetrahydropyridina (MPTP) inhibit the mitochondrial complex I, causing death of dopaminergic neurons in the substantia nigra, thus providing an acute model of PD. It has been recently

demonstrated that mitochondrial hexokinase II promotes neuronal survival in rotenone treated cells, and that this enzyme acts downstream of glycogen synthase kinase-3 (GSK-3), which is considered to be a critical factor in regulating neuronal cell survival and death [125]. More recently, the same group generated amplicons expressing hexokinase II; overexpression of this protein in the substantia nigra of mice subsequently administered with rotenone or MPTP, prevented neuronal cell death induced by both drugs and reduced the associated motor deficits. These results provide the first proof that hexokinase II could protect against dopaminergic neurodegeneration *in vivo*, and suggest that increase of hexokinase II expression could represent a promising approach to treat PD [126].

Narcolepsy

Narcolepsy is a neurodegenerative sleep disorder that is linked to the loss of neurons containing the neuropeptide orexin (also known as hypocretin). Liu and collaborators inoculated an amplicon vector expressing pre-pro-orexin into the lateral hypothalamus of orexin knockout mice and showed that exogenous expression of orexin significantly improved sleeping in these animals [127].

AMPLICON VECTORS AND NERVOUS SYSTEM DAMAGE

Ischemia

Apoptosis plays a critical role in many neurological diseases, including stroke. To study the protective role of the antiapoptotic factor *bcl-2* in ischemia, an amplicon vector encoding *bcl-2* (HSV**bcl-2**) was infused in rat cerebral cortex 24 hours prior to ischemia induction. Expression of *bcl-2* was confirmed by immunohistochemistry in animals injected with the HSV**bcl2** expression vector. Tissue viability significantly increased at the injection site in HSV**bcl2**, but not in animals injected with a similar vector expressing *E. coli* LacZ (HSV**LacZ**) [128]. Later, HSV**bcl-2** was used to inject gerbils unilaterally into the CA1 region of the hippocampus, 24 hours prior to induce transient global ischemia [129]. Results showed that the local increase in *bcl-2* expression using the HSV**bcl-2** vector, may protect CA1 pyramidal cells from the delayed neuronal death caused by transient global ischemia, and that this happened only in the HSV**bcl-2** injected hemisphere. In the same line, other group generated another *bcl-2* expressing amplicon that was used to infect hippocampal cell cultures. *Bcl-2* expressing vectors enhanced neuron survival after exposure to adriamycin, glutamate and hypoglycemia. Furthermore, dichlorofluorescein measurements indicated that there was a significant reduction in the accumulation of oxygen radicals associated with these insults [130]. Improved neuron survival could be attributed to the fact that the *bcl-2* blocks nuclear Apoptosis-inducing factor (AIF) translocation [131]. In an experimental therapeutic approach, Lawrence and collaborators injected rats with this *Bcl-2* expressing vector after ischemia induction, then leading to a reduction in neuronal loss [132]. Furthermore, they showed that there is a time window (30 minutes to 4 hours after reperfusion) where the injection was effective [132]. In accordance with these results, amplicon vectors expressing the inducible heat shock protein (hsp) 72 can also attenuate cerebral ischemic injury when introduced into the rat striatum, even during post-

ischemia [133]. Furthermore, amplicons expressing hsp72 also protected neurons of CA1 hippocampal region from ischemia, and this protection would be mediated, at least in part, by increased expression of bcl-2 [134].

Neurotoxicity

Increases in cytoplasmic Ca^{2+} concentration can lead to neurotoxicity and neuronal death. The increase of Ca^{2+} could be induced by neurological trauma associated with aging and some neurological diseases. To prevent neurotoxic effects of cytoplasmic Ca^{2+} increases, an amplicon vector that expresses the calcium-binding protein calbindin D28K (calbindin D28KHSV) was used to infect neurons, both *in vitro* and *in vivo* [135, 136]. Cultured neurons infected with this vector responded to hypoglycemia and glutamatergic insults with decreased cytoplasmic $[\text{Ca}^{2+}]$ measured by microfluorometry and increased neuron survival relative to controls [135]. Furthermore, *in vivo* injection of calbindin D28KHSV vector in the hippocampal dentate gyrus increases neuronal survival after application of the antimetabolite 3-acetylpyridine, and increases transynaptic neuronal survival in area CA3 following kainic acid neurotoxicity [136].

Reactive oxygen species (ROS) and oxidative stress damage plays an important role in neuronal death. Amplicon vectors expressing different antioxidant enzymes were used to counteract oxidative damages. The cDNA of catalase and glutathione peroxidase, two enzymes involved in hydrogen peroxide degradation, were subcloned into amplicon vectors. These vectors were shown to decrease neurotoxicity induced by different agents in primary cultures of hippocampus or cerebral cortex cells [137]. A further study using amplicons to express the antioxidant enzyme Cu–Zn–SOD, showed that these vectors were able to protect hippocampal neurons through the induction of glutathione peroxidase expression, though only in the case of neurons treated with sodium cyanide. The authors pointed out that the effect of the amplicons actually worsens the toxic effects of kainic acid, another classical ROS inducer, raising a cautionary note concerning gene therapy against oxidative damage [138]. Amplicons expressing glutamic acid decarboxylase (GAD67) were shown to protect non-differentiated cortical neurons from glutamate toxicity mediated by oxidative stress [139].

It was reported that the HIV glycoprotein of 120kDa (HIV-gp120) could be neurotoxic at certain doses and that the hsp70, hsp25 or hsp90 overexpression would protect neurons from HIV-gp120 effect. Therefore, Lim and collaborators overexpressed hsp70 with a HSV-amplicon vector in cultured hippocampal neurons; in this way, they demonstrated that hsp70 overexpression protected cultured hippocampal neurons from HIV-gp120 neurotoxicity [140].

Hypoglycemia could result in an insult for neurons. With amplicons expressing the rat brain glucose transporter (GT), it was shown that these vectors: (i) can maintain neuronal metabolism and reduce the extent of neurons loss in cultures, after a period of hypoglycemia [141]; (ii) protected cultured hippocampal, spinal cord and septal neurons against various necrotic insults, including hypoglycemia, glutamate and 3-nitropropionic acid [142]; and (iii) can enhance glucose uptake in adult rat hippocampus and in hippocampal cultures [143].

Neurotrophins

Neurotrophins are a family of growth factors that play important roles in the development and maintenance of the nervous system. The human brain-derived neurotrophic factor (BDNF) is one of the most important neurotrophins. The knowledge about the complex function of BDNF in mammalian nervous system is continuously expanding; it plays a key role as mediator of activity-induced long term potentiation (LTP) in the hippocampus, as well as in behaviour and memory [144]; BDNF has been implicated in neurodegenerative diseases [145], motor diseases [146] and fragile X syndrome [147]. This neurotrophin participates in maturation and function of mammalian auditory neurons. For this reason, amplicon vectors expressing BDNF were used to evaluate the feasibility of gene therapy of deafness. First, Geschwind and collaborators used an amplicon vector bearing a BDNF cDNA and *E. coli* β -galactosidase (HSVbdnflac), to evaluate the expression and biological activity in established cell lines and explant cultures, prepared from spiral ganglia of the murine ear [148]. Using two BDNF-responsive cell lines, PC12trkB and MG87trkB, they demonstrated efficient secretion of biologically active BDNF [148]. Also, the transduction of explanted spiral ganglia with HSVbdnflac, elicited a robust neuritic process outgrowth, that is comparable to the effect of exogenously added BDNF [148]. Later on, the amplicon vector expressing BDNF was used in mice to infect damaged spiral ganglion. Four weeks post-infection, stable production of BDNF was observed; and it supported the survival of auditory neurons by preventing their loss due to trophic factor deprivation-induced apoptosis [149]. In addition, the use of amplicons expressing BDNF promoted neuronal survival up to the same maximal level seen by adding exogenous BDNF, in a model of avian cochlear neuron cultures [150]. Other neurotrophins were implicated in the proper function of the auditory system. Amplicons expressing neurotrophin-3 (NT-3) were also used in murine cochlear explants model. After infection, the cochlear explants were exposed to cisplatin to induce destruction of hair cells and neurons in the auditory system. This toxicity, defined as ototoxicity, is a major dose-limiting side effect of cisplatin chemotherapy for cancer patients. Amplicon-mediated NT-3 transduction was shown to attenuate the ototoxic action of cisplatin, demonstrating the potency of NT-3 in protecting spiral ganglion neurons from degeneration [151]. Moreover, aged mouse injected in the peripheral auditory system with amplicon vectors expressing NT-3, showed significantly more spiral ganglion surviving neurons (SGNs) and lower incidence of cisplatin-induced apoptosis or necrosis, than those injected with a control vector [152]. Therefore, this approach seemed to be a promising treatment for prevention of chemical-induced hearing disorders and potentially for hearing degeneration due to normal aging.

Amplicon vectors were used to compare BDNF ability with that of the glial-derived neurotrophic factor (GDNF) to protect nigrostriatal neurons in a rat model of PD. According to this study, GDNF was significantly more effective than BDNF for both correcting behavioral deficits and protecting nigrostriatal dopaminergic neurons. The expression of both neurotrophic factors from the same vector was not more effective than expressing GDNF only [153]. In a further study addressing the effect of this trophic factor, it was shown that intracerebral administration of amplicon vectors expressing GDNF, following prior occlusion of the middle cerebral artery, displayed neuroprotection in ischemic injury. Treated animals showed reduced motor deficits and, after 1 month, there was a reduction in tissue loss, in glial fibrillary acidic protein (GFAP) and in caspase-3 immunostaining [154].

In neonatal rats, the combined delivery of NT-3 and the GluN2D subunit of the NMDA receptor, that was possible thanks to the large capacity of amplicon vectors, was used to strengthen monosynaptic connections in contused cords and to induce the appearance of weak but functional multi-synaptic connections, in double hemi-sected cords. On the other hand, the expression of either NT-3 or GluN2D alone failed to induce appearance of synaptic responses through the hemi-sected region of newborn rats [155]. More recently, the same group has treated adult rats with the following agents: (i) anti-Nogo-A antibodies to neutralize the growth-inhibitor Nogo-A; (ii) NT-3 via engineered fibroblasts, to promote neuron survival and plasticity; and (iii) the GluN2D subunit via an HSV-1 amplicon vector, to elevate NMDA receptor function by reversing its Mg^{2+} blockade, thereby enhancing synaptic plasticity and promoting the effects of NT-3 [156]. The combined treatment resulted in slightly better motor function in the absence of adverse effects (i.e., pain), suggesting that this novel combined treatment may help to improve function of the damaged spinal cord [156].

Neuropathic Pain

Neuropathic pain is a chronic form of pain that results from an injury to the nervous system. It could be due to multiple causes, like alcoholism, chemotherapy, diabetes, facial nerve problems among others. Amplicon vectors expressing proenkephalin (pHSV-hPPE-lacZ, SHPZ), were used to investigate the antinociceptive effect of leu-enkephalin (a product of proenkephalin processing). The amplicons were injected into the ventral periaqueductal grey [157] or into the cortex of rats [21]. Both studies showed that the injection of SHPZ, but not of a control vector only expressing the reporter gene lacZ, attenuated neuropathic pain and reduced induced hypersensitivity in rats [21, 157].

ACKNOWLEDGMENTS

ALE and DAJ would like to thank the French *CNRS* and the Argentinean *CONICET* institutions for supporting the International Associated Laboratory (LIA) DEVENIR. Part of this work was supported by *Association Française contre les Myopathies* (AFM). MEM PhD studies were supported by AFM. CAB participation in this work was supported by the French Embassy in Argentina with a “*Bernardo Houssay*” postdoctoral position.

REFERENCES

- [1] Fatahzadeh, M. and R.A. Schwartz, Human herpes simplex labialis. *Clin Exp Dermatol*, 2007. 32(6): p. 625-30.
- [2] Kelly, B.J., et al., Functional roles of the tegument proteins of herpes simplex virus type 1. *Virus Res*, 2009. 145(2): p. 173-86.
- [3] Grunewald, K. and M. Cyrklaff, Structure of complex viruses and virus-infected cells by electron cryo tomography. *Curr Opin Microbiol*, 2006. 9(4): p. 437-42.

- [4] Zhou, Z.H., et al., Visualization of tegument-capsid interactions and DNA in intact herpes simplex virus type 1 virions. *J Virol*, 1999. 73(4): p. 3210-8.
- [5] Oh, J. and N.W. Fraser, Temporal association of the herpes simplex virus genome with histone proteins during a lytic infection. *J Virol*, 2008. 82(7): p. 3530-7.
- [6] Bataille, D. and A.L. Epstein, Herpes simplex virus type 1 replication and recombination. *Biochimie*, 1995. 77(10): p. 787-95.
- [7] McGeoch, D.J., et al., The complete DNA sequence of the long unique region in the genome of herpes simplex virus type 1. *J Gen Virol*, 1988. 69 (Pt 7): p. 1531-74.
- [8] McGeoch, D.J., et al., Complete DNA sequence of the short repeat region in the genome of herpes simplex virus type 1. *Nucleic Acids Res*, 1986. 14(4): p. 1727-45.
- [9] McGeoch, D.J., et al., Sequence determination and genetic content of the short unique region in the genome of herpes simplex virus type 1. *J Mol Biol*, 1985. 181(1): p. 1-13.
- [10] de Silva, S. and W.J. Bowers, Herpes Virus Amplicon Vectors. *Viruses*, 2009. 1(3): p. 594-629.
- [11] Vlazny, D.A. and N. Frenkel, Replication of herpes simplex virus DNA: localization of replication recognition signals within defective virus genomes. *Proc Natl Acad Sci U S A*, 1981. 78(2): p. 742-6.
- [12] Hardwicke, M.A. and P.A. Schaffer, Cloning and characterization of herpes simplex virus type 1 oriL: comparison of replication and protein-DNA complex formation by oriL and oriS. *J Virol*, 1995. 69(3): p. 1377-88.
- [13] Davison, A.J. and N.M. Wilkie, Nucleotide sequences of the joint between the L and S segments of herpes simplex virus types 1 and 2. *J Gen Virol*, 1981. 55(Pt 2): p. 315-31.
- [14] Deiss, L.P., J. Chou, and N. Frenkel, Functional domains within the a sequence involved in the cleavage-packaging of herpes simplex virus DNA. *J Virol*, 1986. 59(3): p. 605-18.
- [15] Nasser, M. and E.S. Mocarski, The cleavage recognition signal is contained within sequences surrounding an a-a junction in herpes simplex virus DNA. *Virology*, 1988. 167(1): p. 25-30.
- [16] McVoy, M.A., et al., Sequences within the herpesvirus-conserved pac1 and pac2 motifs are required for cleavage and packaging of the murine cytomegalovirus genome. *J Virol*, 1998. 72(1): p. 48-56.
- [17] DeLuca, N.A., A.M. McCarthy, and P.A. Schaffer, Isolation and characterization of deletion mutants of herpes simplex virus type 1 in the gene encoding immediate-early regulatory protein ICP4. *J Virol*, 1985. 56(2): p. 558-70.
- [18] Ozuer, A., et al., Evaluation of infection parameters in the production of replication-defective HSV-1 viral vectors. *Biotechnol Prog*, 2002. 18(3): p. 476-82.
- [19] Spaete, R.R. and N. Frenkel, The herpes simplex virus amplicon: a new eucaryotic defective-virus cloning-amplifying vector. *Cell*, 1982. 30(1): p. 295-304.
- [20] Jerusalinsky, D., M.V. Baez, and A.L. Epstein, Herpes simplex virus type 1-based amplicon vectors for fundamental research in neurosciences and gene therapy of neurological diseases. *J Physiol Paris*, 2012. 106(1-2): p. 2-11.
- [21] Zou, W., et al., Intrathecal herpes simplex virus type 1 amplicon vector-mediated human proenkephalin reduces chronic constriction injury-induced neuropathic pain in rats. *Mol Med Rep*, 2011. 4(3): p. 529-33.
- [22] Zhang, G.R. and A.I. Geller, A helper virus-free HSV-1 vector containing the vesicular glutamate transporter-1 promoter supports expression preferentially in VGLUT1-containing glutamatergic neurons. *Brain Res*, 2010. 1331: p. 12-9.

- [23] Marconi, P., R. Manservigi, and A.L. Epstein, HSV-1-derived helper-independent defective vectors, replicating vectors and amplicon vectors, for the treatment of brain diseases. *Curr Opin Drug Discov Devel*, 2010. 13(2): p. 169-83.
- [24] Blanc, A.M., et al., Induction of humoral responses to BHV-1 glycoprotein D expressed by HSV-1 amplicon vectors. *J Vet Sci*, 2012. 13(1): p. 59-65.
- [25] Laimbacher, A.S., et al., HSV-1 amplicon vectors launch the production of heterologous rotavirus-like particles and induce rotavirus-specific immune responses in mice. *Mol Ther*, 2012. 20(9): p. 1810-20.
- [26] Saydam, O., D.L. Glauser, and C. Fraefel, Construction and packaging of herpes simplex virus/adeno-associated virus (HSV/AAV) Hybrid amplicon vectors. *Cold Spring Harb Protoc*, 2012. 2012(3): p. 352-6.
- [27] de Oliveira, A.P. and C. Fraefel, Herpes simplex virus type 1/adeno-associated virus hybrid vectors. *Open Virol J*, 2010. 4: p. 109-22.
- [28] Mandegar, M.A., et al., Functional human artificial chromosomes are generated and stably maintained in human embryonic stem cells. *Hum Mol Genet*, 2011. 20(15): p. 2905-13.
- [29] Moralli, D., et al., A novel human artificial chromosome gene expression system using herpes simplex virus type 1 vectors. *EMBO Rep*, 2006. 7(9): p. 911-8.
- [30] Cuchet, D., et al., HSV-1 amplicon vectors: a promising and versatile tool for gene delivery. *Expert Opin Biol Ther*, 2007. 7(7): p. 975-95.
- [31] Epstein, A.L., HSV-1-derived amplicon vectors: recent technological improvements and remaining difficulties--a review. *Mem Inst Oswaldo Cruz*, 2009. 104(3): p. 399-410.
- [32] Fraefel, C., Gene delivery using helper virus-free HSV-1 amplicon vectors. *Curr Protoc Neurosci*, 2007. Chapter 4: p. Unit 4 14.
- [33] Manservigi, R., R. Argani, and P. Marconi, HSV Recombinant Vectors for Gene Therapy. *Open Virol J*. 4: p. 123-56.
- [34] Kwong, A.D. and N. Frenkel, Herpes simplex virus amplicon: effect of size on replication of constructed defective genomes containing eucaryotic DNA sequences. *J Virol*, 1984. 51(3): p. 595-603.
- [35] Boehmer, P.E. and I.R. Lehman, Herpes simplex virus DNA replication. *Annu Rev Biochem*, 1997. 66: p. 347-84.
- [36] Cunningham, C. and A.J. Davison, A cosmid-based system for constructing mutants of herpes simplex virus type 1. *Virology*, 1993. 197(1): p. 116-24.
- [37] Fraefel, C., et al., Helper virus-free transfer of herpes simplex virus type 1 plasmid vectors into neural cells. *J Virol*, 1996. 70(10): p. 7190-7.
- [38] Heister, T.G., et al., Construction of HSV-1 BACs and their use for packaging of HSV-1-based amplicon vectors. *Methods Mol Biol*, 2004. 256: p. 241-56.
- [39] Saeki, Y., X.O. Breakefield, and E.A. Chiocca, Improved HSV-1 amplicon packaging system using ICP27-deleted, oversized HSV-1 BAC DNA. *Methods Mol Med*, 2003. 76: p. 51-60.
- [40] Saeki, Y., et al., Improved helper virus-free packaging system for HSV amplicon vectors using an ICP27-deleted, oversized HSV-1 DNA in a bacterial artificial chromosome. *Mol Ther*, 2001. 3(4): p. 591-601.
- [41] Geller, A.I., A new method to propagate defective HSV-1 vectors. *Nucleic Acids Res*, 1988. 16(12): p. 5690.

- [42] Geller, A.I., et al., An efficient deletion mutant packaging system for defective herpes simplex virus vectors: potential applications to human gene therapy and neuronal physiology. *Proc Natl Acad Sci U S A*, 1990. 87(22): p. 8950-4.
- [43] Lim, F., et al., Generation of high-titer defective HSV-1 vectors using an IE 2 deletion mutant and quantitative study of expression in cultured cortical cells. *Biotechniques*, 1996. 20(3): p. 460-9.
- [44] Pechan, P.A., et al., A novel 'piggyback' packaging system for herpes simplex virus amplicon vectors. *Hum Gene Ther*, 1996. 7(16): p. 2003-13.
- [45] Zhang, X., et al., An efficient selection system for packaging herpes simplex virus amplicons. *J Gen Virol*, 1998. 79 (Pt 1): p. 125-31.
- [46] Logvinoff, C. and A.L. Epstein, Genetic engineering of herpes simplex virus and vector genomes carrying loxP sites in cells expressing Cre recombinase. *Virology*, 2000. 267(1): p. 102-10.
- [47] Logvinoff, C. and A.L. Epstein, Intracellular Cre-mediated deletion of the unique packaging signal carried by a herpes simplex virus type 1 recombinant and its relationship to the cleavage-packaging process. *J Virol*, 2000. 74(18): p. 8402-12.
- [48] Zaupa, C., V. Revol-Guyot, and A.L. Epstein, Improved packaging system for generation of high-level noncytotoxic HSV-1 amplicon vectors using Cre-loxP site-specific recombination to delete the packaging signals of defective helper genomes. *Hum Gene Ther*, 2003. 14(11): p. 1049-63.
- [49] Suzuki, M., K. Kasai, and Y. Saeki, Plasmid DNA sequences present in conventional herpes simplex virus amplicon vectors cause rapid transgene silencing by forming inactive chromatin. *J Virol*, 2006. 80(7): p. 3293-300.
- [50] *Content and Review of Chemistry, Manufacturing, and Control (CMC) Information for Human Gene Therapy Investigational New Drug Applications (INDs)*. U.S. Department of Health and Human Services / Food and Drug Administration, 2008.
- [51] Ho, I.A., et al., Targeting human glioma cells using HSV-1 amplicon peptide display vector. *Gene Ther*, 2010. 17(2): p. 250-60.
- [52] Kaestle, C., et al., Imaging herpes simplex virus type 1 amplicon vector-mediated gene expression in human glioma spheroids. *Mol Imaging*, 2011. 10(3): p. 197-205.
- [53] Boutell, C., et al., Reciprocal activities between herpes simplex virus type 1 regulatory protein ICP0, a ubiquitin E3 ligase, and ubiquitin-specific protease USP7. *J Virol*, 2005. 79(19): p. 12342-54.
- [54] Lomonte, P., K.F. Sullivan, and R.D. Everett, Degradation of nucleosome-associated centromeric histone H3-like protein CENP-A induced by herpes simplex virus type 1 protein ICP0. *J Biol Chem*, 2001. 276(8): p. 5829-35.
- [55] Cuchet, D., et al., Characterization of antiproliferative and cytotoxic properties of the HSV-1 immediate-early ICP0 protein. *J Gene Med*, 2005. 7(9): p. 1187-99.
- [56] Saydam, O., et al., Herpes simplex virus 1 amplicon vector-mediated siRNA targeting epidermal growth factor receptor inhibits growth of human glioma cells in vivo. *Mol Ther*, 2005. 12(5): p. 803-12.
- [57] Saydam, O., et al., HSV-1 amplicon-mediated post-transcriptional inhibition of Rad51 sensitizes human glioma cells to ionizing radiation. *Gene Ther*, 2007. 14(15): p. 1143-51.
- [58] Ho, I.A., K.M. Hui, and P.Y. Lam, Targeting proliferating tumor cells via the transcriptional control of therapeutic genes. *Cancer Gene Ther*, 2006. 13(1): p. 44-52.

- [59] Ho, I.A., K.M. Hui, and P.Y. Lam, Glioma-specific and cell cycle-regulated herpes simplex virus type 1 amplicon viral vector. *Hum Gene Ther*, 2004. 15(5): p. 495-508.
- [60] Ho, I.A., W.H. Ng, and P.Y. Lam, FasL and FADD delivery by a glioma-specific and cell cycle-dependent HSV-1 amplicon virus enhanced apoptosis in primary human brain tumors. *Mol Cancer*, 2010. 9: p. 270.
- [61] Prabhakar, S., et al., Imaging and therapy of experimental schwannomas using HSV amplicon vector-encoding apoptotic protein under Schwann cell promoter. *Cancer Gene Ther*, 2010. 17(4): p. 266-74.
- [62] Whitley, R.J. and B. Roizman, Herpes simplex virus infections. *Lancet*, 2001. 357(9267): p. 1513-8.
- [63] York, I.A., et al., A cytosolic herpes simplex virus protein inhibits antigen presentation to CD8+ T lymphocytes. *Cell*, 1994. 77(4): p. 525-35.
- [64] D'Antuono, A., et al., HSV-1 amplicon vectors that direct the in situ production of foot-and-mouth disease virus antigens in mammalian cells can be used for genetic immunization. *Vaccine*, 2010. 28(46): p. 7363-72.
- [65] Hocknell, P.K., et al., Expression of human immunodeficiency virus type 1 gp120 from herpes simplex virus type 1-derived amplicons results in potent, specific, and durable cellular and humoral immune responses. *J Virol*, 2002. 76(11): p. 5565-80.
- [66] Wang, X., et al., Cellular immune responses to helper-free HSV-1 amplicon particles encoding HIV-1 gp120 are enhanced by DNA priming. *Vaccine*, 2003. 21(19-20): p. 2288-97.
- [67] Gorantla, S., et al., Human dendritic cells transduced with herpes simplex virus amplicons encoding human immunodeficiency virus type 1 (HIV-1) gp120 elicit adaptive immune responses from human cells engrafted into NOD/SCID mice and confer partial protection against HIV-1 challenge. *J Virol*, 2005. 79(4): p. 2124-32.
- [68] Santos, K., et al., Effect of promoter strength on protein expression and immunogenicity of an HSV-1 amplicon vector encoding HIV-1 Gag. *Vaccine*, 2007. 25(9): p. 1634-46.
- [69] Tyler, C.M., et al., HSV amplicons: neuro applications. *Curr Gene Ther*, 2006. 6(3): p. 337-50.
- [70] Brockman, M.A. and D.M. Knipe, Herpes simplex virus vectors elicit durable immune responses in the presence of preexisting host immunity. *J Virol*, 2002. 76(8): p. 3678-87.
- [71] Lauterbach, H., et al., Reduced immune responses after vaccination with a recombinant herpes simplex virus type 1 vector in the presence of antiviral immunity. *J Gen Virol*, 2005. 86(Pt 9): p. 2401-10.
- [72] Adrover, M.F., et al., Hippocampal infection with HSV-1-derived vectors expressing an NMDAR1 antisense modifies behavior. *Genes Brain Behav*, 2003. 2(2): p. 103-13.
- [73] Cheli, V.T., et al., Gene transfer of NMDAR1 subunit sequences to the rat CNS using herpes simplex virus vectors interfered with habituation. *Cell Mol Neurobiol*, 2002. 22(3): p. 303-14.
- [74] Cheli, V., et al., Knocking-down the NMDAR1 subunit in a limited amount of neurons in the rat hippocampus impairs learning. *J Neurochem*, 2006. 97 Suppl 1: p. 68-73.
- [75] Neill, J.C., et al., Enhanced auditory reversal learning by genetic activation of protein kinase C in small groups of rat hippocampal neurons. *Brain Res Mol Brain Res*, 2001. 93(2): p. 127-36.
- [76] Malinow, R. and R.C. Malenka, AMPA receptor trafficking and synaptic plasticity. *Annu Rev Neurosci*, 2002. 25: p. 103-26.

- [77] Rumpel, S., et al., Postsynaptic receptor trafficking underlying a form of associative learning. *Science*, 2005. 308(5718): p. 83-8.
- [78] Kandel, E.R., The molecular biology of memory storage: a dialogue between genes and synapses. *Science*, 2001. 294(5544): p. 1030-8.
- [79] Han, J.H., et al., Neuronal competition and selection during memory formation. *Science*, 2007. 316(5823): p. 457-60.
- [80] Han, J.H., et al., Increasing CREB in the auditory thalamus enhances memory and generalization of auditory conditioned fear. *Learn Mem*, 2008. 15(6): p. 443-53.
- [81] Brightwell, J.J., et al., Hippocampal overexpression of mutant creb blocks long-term, but not short-term memory for a socially transmitted food preference. *Learn Mem*, 2005. 12(1): p. 12-7.
- [82] Brightwell, J.J., et al., Transfection of mutant CREB in the striatum, but not the hippocampus, impairs long-term memory for response learning. *Neurobiol Learn Mem*, 2008. 89(1): p. 27-35.
- [83] Barrot, M., et al., Regulation of anxiety and initiation of sexual behavior by CREB in the nucleus accumbens. *Proc Natl Acad Sci U S A*, 2005. 102(23): p. 8357-62.
- [84] Liu, J., et al., Binge alcohol drinking is associated with GABAA alpha2-regulated Toll-like receptor 4 (TLR4) expression in the central amygdala. *Proc Natl Acad Sci U S A*, 2011. 108(11): p. 4465-70.
- [85] Cossee, M., et al., Evolution of the Friedreich's ataxia trinucleotide repeat expansion: founder effect and premutations. *Proc Natl Acad Sci U S A*, 1997. 94(14): p. 7452-7.
- [86] Lim, F., et al., Functional recovery in a Friedreich's ataxia mouse model by frataxin gene transfer using an HSV-1 amplicon vector. *Mol Ther*, 2007. 15(6): p. 1072-8.
- [87] Gomez-Sebastian, S., et al., Infectious delivery and expression of a 135 kb human FRDA genomic DNA locus complements Friedreich's ataxia deficiency in human cells. *Mol Ther*, 2007. 15(2): p. 248-54.
- [88] Gimenez-Cassina, A., et al., Infectious delivery and long-term persistence of transgene expression in the brain by a 135-kb iBAC-FXN genomic DNA expression vector. *Gene Ther*, 2011. 18(10): p. 1015-9.
- [89] Verhagen, M.M., et al., Presence of ATM protein and residual kinase activity correlates with the phenotype in ataxia-telangiectasia: a genotype-phenotype study. *Hum Mutat*, 2012. 33(3): p. 561-71.
- [90] Cortes, M.L., et al., HSV-1 amplicon vector-mediated expression of ATM cDNA and correction of the ataxia-telangiectasia cellular phenotype. *Gene Ther*, 2003. 10(16): p. 1321-7.
- [91] Cortes, M.L., et al., Expression of human ATM cDNA in Atm-deficient mouse brain mediated by HSV-1 amplicon vector. *Neuroscience*, 2006. 141(3): p. 1247-56.
- [92] Cortes, M.L., et al., Targeted integration of functional human ATM cDNA into genome mediated by HSV/AAV hybrid amplicon vector. *Mol Ther*, 2008. 16(1): p. 81-8.
- [93] De Felice, F.G., et al., Abeta oligomers induce neuronal oxidative stress through an N-methyl-D-aspartate receptor-dependent mechanism that is blocked by the Alzheimer drug memantine. *J Biol Chem*, 2007. 282(15): p. 11590-601.
- [94] Shankar, G.M., et al., Natural oligomers of the Alzheimer amyloid-beta protein induce reversible synapse loss by modulating an NMDA-type glutamate receptor-dependent signaling pathway. *J Neurosci*, 2007. 27(11): p. 2866-75.

- [95] Thinakaran, G. and E.H. Koo, Amyloid precursor protein trafficking, processing, and function. *J Biol Chem*, 2008. 283(44): p. 29615-9.
- [96] Shankar, G.M., et al., Amyloid-beta protein dimers isolated directly from Alzheimer's brains impair synaptic plasticity and memory. *Nat Med*, 2008. 14(8): p. 837-42.
- [97] Haass, C. and D.J. Selkoe, Soluble protein oligomers in neurodegeneration: lessons from the Alzheimer's amyloid beta-peptide. *Nat Rev Mol Cell Biol*, 2007. 8(2): p. 101-12.
- [98] Klein, W.L., Synaptic targeting by A beta oligomers (ADDLS) as a basis for memory loss in early Alzheimer's disease. *Alzheimers Dement*, 2006. 2(1): p. 43-55.
- [99] Lesne, S., et al., A specific amyloid-beta protein assembly in the brain impairs memory. *Nature*, 2006. 440(7082): p. 352-7.
- [100] Roselli, F., et al., Soluble beta-amyloid1-40 induces NMDA-dependent degradation of postsynaptic density-95 at glutamatergic synapses. *J Neurosci*, 2005. 25(48): p. 11061-70.
- [101] Bowers, W.J., et al., HSV amplicon-mediated Abeta vaccination in Tg2576 mice: differential antigen-specific immune responses. *Neurobiol Aging*, 2005. 26(4): p. 393-407.
- [102] Barnes, C.A., Memory deficits associated with senescence: a neurophysiological and behavioral study in the rat. *J Comp Physiol Psychol*, 1979. 93(1): p. 74-104.
- [103] Frazer, M.E., et al., Reduced pathology and improved behavioral performance in Alzheimer's disease mice vaccinated with HSV amplicons expressing amyloid-beta and interleukin-4. *Mol Ther*, 2008. 16(5): p. 845-53.
- [104] Cousins, S.L., et al., Amyloid precursor protein 695 associates with assembled NR2A- and NR2B-containing NMDA receptors to result in the enhancement of their cell surface delivery. *J Neurochem*, 2009. 111(6): p. 1501-13.
- [105] Hoey, S.E., R.J. Williams, and M.S. Perkinson, Synaptic NMDA receptor activation stimulates alpha-secretase amyloid precursor protein processing and inhibits amyloid-beta production. *J Neurosci*, 2009. 29(14): p. 4442-60.
- [106] Mori, H. and M. Mishina, Structure and function of the NMDA receptor channel. *Neuropharmacology*, 1995. 34(10): p. 1219-37.
- [107] Monyer, H., et al., Developmental and regional expression in the rat brain and functional properties of four NMDA receptors. *Neuron*, 1994. 12(3): p. 529-40.
- [108] Moriyoshi, K., et al., Molecular cloning and characterization of the rat NMDA receptor. *Nature*, 1991. 354(6348): p. 31-7.
- [109] Decker, H., et al., N-methyl-D-aspartate receptors are required for synaptic targeting of Alzheimer's toxic amyloid-beta peptide oligomers. *J Neurochem*, 2010. 115(6): p. 1520-9.
- [110] Liu, J., et al., Amyloid-beta induces caspase-dependent loss of PSD-95 and synaptophysin through NMDA receptors. *J Alzheimers Dis*, 2010. 22(2): p. 541-56.
- [111] Balducci, C., et al., Cognitive deficits associated with alteration of synaptic metaplasticity precede plaque deposition in AbetaPP23 transgenic mice. *J Alzheimers Dis*, 2010. 21(4): p. 1367-81.
- [112] Paoletti, P., Molecular basis of NMDA receptor functional diversity. *Eur J Neurosci*, 2011. 33(8): p. 1351-65.
- [113] Peruzzi, P.P., et al., Physiological transgene regulation and functional complementation of a neurological disease gene deficiency in neurons. *Mol Ther*, 2009. 17(9): p. 1517-26.

- [114] Hornykiewicz, O., Basic research on dopamine in Parkinson's disease and the discovery of the nigrostriatal dopamine pathway: the view of an eyewitness. *Neurodegener Dis*, 2008. 5(3-4): p. 114-7.
- [115] Rodriguez-Oroz, M.C., et al., Initial clinical manifestations of Parkinson's disease: features and pathophysiological mechanisms. *Lancet Neurol*, 2009. 8(12): p. 1128-39.
- [116] During, M.J., et al., Long-term behavioral recovery in parkinsonian rats by an HSV vector expressing tyrosine hydroxylase. *Science*, 1994. 266(5189): p. 1399-403.
- [117] Sun, M., et al., Correction of a rat model of Parkinson's disease by coexpression of tyrosine hydroxylase and aromatic amino acid decarboxylase from a helper virus-free herpes simplex virus type 1 vector. *Hum Gene Ther*, 2003. 14(5): p. 415-24.
- [118] Sun, M., et al., Coexpression of tyrosine hydroxylase, GTP cyclohydrolase I, aromatic amino acid decarboxylase, and vesicular monoamine transporter 2 from a helper virus-free herpes simplex virus type 1 vector supports high-level, long-term biochemical and behavioral correction of a rat model of Parkinson's disease. *Hum Gene Ther*, 2004. 15(12): p. 1177-96.
- [119] Zhang, G.R., et al., The vesicular glutamate transporter-1 upstream promoter and first intron each support glutamatergic-specific expression in rat postrhinal cortex. *Brain Res*, 2011. 1377: p. 1-12.
- [120] Zhang, G.R., et al., A tyrosine hydroxylase-neurofilament chimeric promoter enhances long-term expression in rat forebrain neurons from helper virus-free HSV-1 vectors. *Brain Res Mol Brain Res*, 2000. 84(1-2): p. 17-31.
- [121] Cao, H., et al., Enhanced nigrostriatal neuron-specific, long-term expression by using neural-specific promoters in combination with targeted gene transfer by modified helper virus-free HSV-1 vector particles. *BMC Neurosci*, 2008. 9: p. 37.
- [122] Gao, Q., et al., Isolation of an enhancer from the rat tyrosine hydroxylase promoter that supports long-term, neuronal-specific expression from a neurofilament promoter, in a helper virus-free HSV-1 vector system. *Brain Res*, 2007. 1130(1): p. 1-16.
- [123] Corso, T.D., et al., Transfection of tyrosine kinase deleted FGF receptor-1 into rat brain substantia nigra reduces the number of tyrosine hydroxylase expressing neurons and decreases concentration levels of striatal dopamine. *Brain Res Mol Brain Res*, 2005. 139(2): p. 361-6.
- [124] Sortwell, C.E., et al., Effects of ex vivo transduction of mesencephalic reagggregates with bcl-2 on grafted dopamine neuron survival. *Brain Res*, 2007. 1134(1): p. 33-44.
- [125] Gimenez-Cassina, A., et al., Mitochondrial hexokinase II promotes neuronal survival and acts downstream of glycogen synthase kinase-3. *J Biol Chem*, 2009. 284(5): p. 3001-11.
- [126] Corona, J.C., et al., Hexokinase II gene transfer protects against neurodegeneration in the rotenone and MPTP mouse models of Parkinson's disease. *J Neurosci Res*, 2010. 88(9): p. 1943-50.
- [127] Liu, M., et al., Orexin (hypocretin) gene transfer diminishes narcoleptic sleep behavior in mice. *Eur J Neurosci*, 2008. 28(7): p. 1382-93.
- [128] Linnik, M.D., et al., Expression of bcl-2 from a defective herpes simplex virus-1 vector limits neuronal death in focal cerebral ischemia. *Stroke*, 1995. 26(9): p. 1670-4; discussion 1675.

- [129] Antonawich, F.J., H.J. Federoff, and J.N. Davis, BCL-2 transduction, using a herpes simplex virus amplicon, protects hippocampal neurons from transient global ischemia. *Exp Neurol*, 1999. 156(1): p. 130-7.
- [130] Lawrence, M.S., et al., Overexpression of Bcl-2 with herpes simplex virus vectors protects CNS neurons against neurological insults in vitro and in vivo. *J Neurosci*, 1996. 16(2): p. 486-96.
- [131] Zhao, H., et al., Bcl-2 transfection via herpes simplex virus blocks apoptosis-inducing factor translocation after focal ischemia in the rat. *J Cereb Blood Flow Metab*, 2004. 24(6): p. 681-92.
- [132] Lawrence, M.S., et al., Herpes simplex viral vectors expressing Bcl-2 are neuroprotective when delivered after a stroke. *J Cereb Blood Flow Metab*, 1997. 17(7): p. 740-4.
- [133] Hoehn, B., et al., Overexpression of HSP72 after induction of experimental stroke protects neurons from ischemic damage. *J Cereb Blood Flow Metab*, 2001. 21(11): p. 1303-9.
- [134] Kelly, S., et al., Gene transfer of HSP72 protects cornu ammonis 1 region of the hippocampus neurons from global ischemia: influence of Bcl-2. *Ann Neurol*, 2002. 52(2): p. 160-7.
- [135] Meier, T.J., et al., Gene transfer of calbindin D28k cDNA via herpes simplex virus amplicon vector decreases cytoplasmic calcium ion response and enhances neuronal survival following glutamatergic challenge but not following cyanide. *J Neurochem*, 1998. 71(3): p. 1013-23.
- [136] Phillips, R.G., et al., Calbindin D28K gene transfer via herpes simplex virus amplicon vector decreases hippocampal damage in vivo following neurotoxic insults. *J Neurochem*, 1999. 73(3): p. 1200-5.
- [137] Wang, H., et al., Over-expression of antioxidant enzymes protects cultured hippocampal and cortical neurons from necrotic insults. *J Neurochem*, 2003. 87(6): p. 1527-34.
- [138] Zemlyak, I., et al., Gene therapy in the nervous system with superoxide dismutase. *Brain Res*, 2006. 1088(1): p. 12-8.
- [139] Lamigeon, C., et al., Enhancement of neuronal protection from oxidative stress by glutamic acid decarboxylase delivery with a defective herpes simplex virus vector. *Exp Neurol*, 2003. 184(1): p. 381-92.
- [140] Lim, M.C., S.M. Brooke, and R.M. Sapolsky, gp120 neurotoxicity fails to induce heat shock defenses, while the over expression of hsp70 protects against gp120. *Brain Res Bull*, 2003. 61(2): p. 183-8.
- [141] Lawrence, M.S., et al., Herpes simplex virus vectors overexpressing the glucose transporter gene protect against seizure-induced neuron loss. *Proc Natl Acad Sci U S A*, 1995. 92(16): p. 7247-51.
- [142] Ho, D.Y., et al., Defective herpes simplex virus vectors expressing the rat brain glucose transporter protect cultured neurons from necrotic insults. *J Neurochem*, 1995. 65(2): p. 842-50.
- [143] Ho, D.Y., E.S. Mocarski, and R.M. Sapolsky, Altering central nervous system physiology with a defective herpes simplex virus vector expressing the glucose transporter gene. *Proc Natl Acad Sci U S A*, 1993. 90(8): p. 3655-9.
- [144] Cowansage, K.K., J.E. LeDoux, and M.H. Monfils, Brain-derived neurotrophic factor: a dynamic gatekeeper of neural plasticity. *Curr Mol Pharmacol*, 2010. 3(1): p. 12-29.

- [145] Gupta, V.K., et al., TrkB Receptor Signaling: Implications in Neurodegenerative, Psychiatric and Proliferative Disorders. *Int J Mol Sci*, 2013. 14(5): p. 10122-42.
- [146] He, Y.Y., et al., Role of BDNF in Central Motor Structures and Motor Diseases. *Mol Neurobiol*, 2013.
- [147] Castren, M. and E. Castren, BDNF in fragile X syndrome. *Neuropharmacology*, 2013.
- [148] Geschwind, M.D., et al., Defective HSV-1 vector expressing BDNF in auditory ganglia elicits neurite outgrowth: model for treatment of neuron loss following cochlear degeneration. *Hum Gene Ther*, 1996. 7(2): p. 173-82.
- [149] Staecker, H., et al., Brain-derived neurotrophic factor gene therapy prevents spiral ganglion degeneration after hair cell loss. *Otolaryngol Head Neck Surg*, 1998. 119(1): p. 7-13.
- [150] Garrido, J.J., et al., Defining responsiveness of avian cochlear neurons to brain-derived neurotrophic factor and nerve growth factor by HSV-1-mediated gene transfer. *J Neurochem*, 1998. 70(6): p. 2336-46.
- [151] Chen, X., et al., HSV amplicon-mediated neurotrophin-3 expression protects murine spiral ganglion neurons from cisplatin-induced damage. *Mol Ther*, 2001. 3(6): p. 958-63.
- [152] Bowers, W.J., et al., Neurotrophin-3 transduction attenuates cisplatin spiral ganglion neuron ototoxicity in the cochlea. *Mol Ther*, 2002. 6(1): p. 12-8.
- [153] Sun, M., et al., Comparison of the capability of GDNF, BDNF, or both, to protect nigrostriatal neurons in a rat model of Parkinson's disease. *Brain Res*, 2005. 1052(2): p. 119-29.
- [154] Harvey, B.K., et al., HSV amplicon delivery of glial cell line-derived neurotrophic factor is neuroprotective against ischemic injury. *Exp Neurol*, 2003. 183(1): p. 47-55.
- [155] Arvanian, V.L., et al., Combined delivery of neurotrophin-3 and NMDA receptors 2D subunit strengthens synaptic transmission in contused and staggered double hemisected spinal cord of neonatal rat. *Exp Neurol*, 2006. 197(2): p. 347-52.
- [156] Schnell, L., et al., Combined delivery of Nogo-A antibody, neurotrophin-3 and the NMDA-NR2d subunit establishes a functional 'detour' in the hemisected spinal cord. *Eur J Neurosci*, 2011. 34(8): p. 1256-67.
- [157] Zou, W., et al., Microinjection of HSV-1 amplicon vector-mediated human proenkephalin into the periaqueductal grey attenuates neuropathic pain in rats. *Int J Neurosci*, 2012. 122(4): p. 189-94.

Reviewed by

Dr. Roberto Manservigi, PhD
 E-mail: mns@unife.it
 Address: Università di Ferrara
 Dipartimento di Scienze della Vita e Biotecnologie
 Via Luigi Borsari 46
 44100 – Ferrara
 Italia

Chapter 98

ADVANCES IN VIRAL GENOME RESEARCH OF PAPILLOMAVIRUSES

***A. C. Freitas, A. L. S. Jesus, A. P. D. Gurgel
and M. A. R. Silva***

Department of Genetics,
Federal University of Pernambuco, Brazil

ABSTRACT

Papillomaviruses (PVs) infect the epithelium of amniotes, where they can cause tumours or persist asymptotically. PVs are classified in the *Papillomaviridae* family, that contains 29 genera of PVs isolated from humans (120 types), non-human mammals, birds and reptiles (69 types). PVs have circular double-stranded DNA genomes approximately 8 kb in size and typically contain eight genes. Studies aiming the identification of PVs genomes use techniques such as PCR with consensus primers, rolling circle amplification and metagenomic methods. Advances in papillomaviral genome research have allowed the knowledge of PVs diversity and evolution of the poorly known PVs genera types, revealing that there is still a limited understanding of PVs diversity. Particularly, recent studies in Bovine Papillomavirus (BPV) have shown the identification of novel BPV types and several putative new virus types in cattle. This chapter will show new contributions in PVs genome studies.

Keywords: Papillomavirus, molecular techniques, viral metagenomics

INTRODUCTION

Papillomavirus (PVs) is a group of viruses that induce warts (or papillomas) in a variety of animals. The PVs are widespread in nature and have been recognized primarily in higher vertebrates. Viruses have been characterized from humans, cattle, rabbits, horses, dogs, sheep, elk, deer, nonhuman primates, the harvest mouse, and the multimammate mouse (*Mastomys natalensis*) [1]. Some PVs also have malignant potential for animals and people. A

number of Human Papillomaviruses (HPVs) have been implicated as the etiologic agents for cervical cancer and other epithelial tumors.

PVs are small, nonenveloped, icosahedral DNA viruses that replicate in the nucleus of squamous epithelial cells. The virions consist of a single molecule of double-stranded circular DNA about 8,000 bp in size, contained within a spherical capsid coat, composed of 72 capsomers. The genome of the PVs can be divided, in general, into three major regions: early, late, and a long control region (LCR or noncoding region [NCR])[2]. The late region of the viral genome is expressed only in differentiated layers of warts and other productive lesions, while the early region express both in warts and transformed cells [2,3]. The products of the early region encode viral regulatory proteins, including those viral proteins that are necessary for initiating viral DNA replication and transformation of the host cell [4]. The late region of PVs genomes lies downstream of the early region and encodes L1 and L2 ORFs for translation of a major (L1) and a minor (L2) capsid protein. The LCR region has no protein-coding function, but bears the origin of replication as well as multiple transcription factor binding sites that are important in regulation of RNA polymerase II-initiated transcription from viral early as well as late promoters [5].

PVs have often been classified primarily accordingly to the host species they infect and the sites or diseases with which they are associated. DNA sequencing of many PVs genomes has led their phylogenetic organization, according to the comparative homology of the L1 ORF [6]. The L1 gene is useful for classification and construction of phylogenetic trees, as it is reasonably well conserved and can be aligned for all known PVs [7].

These similarities are consistent with the conclusion that PVs have accompanied their host species during evolution and have evolved with them [8]. Although all PVs share similar genetic organization, the L1 DNA sequence identity is just above 40% between the most divergent genomes. PVs were designated as a distinct family, *Papillomaviridae*, in the 7th Report of the ICTV [9]. De Villiers et al. [6] described the topology of phylogenetic trees based on the nucleotide sequence comparisons and biologically distinguishing features (host species, target tissues, pathogenicity, and genome organization) that determine the classification of PVs on the level of genera.

A nomenclature of these genera based on the Greek alphabet was introduced and has rapidly become accepted and widely used by the ICTV and community of PVs researchers.

Sixteen groups of PVs or individual PVs fulfilled the criterion of genera, and the Greek alphabet from the letters alpha to pi was employed to create their nomenclature. The last official classification of PVs genera ended with the genus Pi-PVs. The description of 13 new PVs genera however, exhausts the Greek alphabet. In order to create a system that continues with the Greek alphabet, it was proposed the use of the Greek alphabet a second time, employing the prefix “dyo”, (i.e., Greek “a second time”) [7].

Within a given genus, the L1 DNA of all members shares more than 60% identity. A viral type with a species has 71 to 89% identity with other types within the species. Within a type can exist subtypes, which share 90 to 98% identity, and variants, which have more than 98% identity [7].

Viruses can be identified by a wide range of techniques. Traditional methods include electron microscopy, cell culture, inoculation studies and serology [10].

Whereas many of the viruses known today were first identified by these techniques, the methods have limitations. For these viruses, the molecular-based techniques provide sensitive and rapid means of virus detection and identification [11].

One such approach uses sequence information from known types to identify related but undiscovered types through cross-hybridization. Another advance has involved PCR amplification of the viral genome. Such PCR-based methods comprise conventional PCR, degenerate PCR, sequence-independent PCR, and rolling circle amplification (RCA). Technological advances have also resulted in the development of metagenomics, the culture-independent study of the collective set of microbial populations in a sample by analyzing the sample's nucleotide sequence content [10].

In the next sections, it will be presented the main strategies for the identification of new types and variants of PVs and a particular broach about the recent advances in the Bovine Papillomavirus (BPV) research.

STRATEGIES USED TO IDENTIFY PAPILLOMAVIRUS GENOMES

Since PVs life cycle requires keratinocyte differentiation, there are no conventional cell lines that allow the growth these viruses [12]. Moreover, PVs could not be transmitted to laboratory animals [13]. Hence, molecular methods have been widely used to discover and characterize the genome of new PVs as well as for diagnostics of PVs that are clinically relevant [14–37].

Among these, molecular methods such as polymerase chain reaction (PCR), hybridization, and metagenomic methodologies have been frequently used in order to detect human [14–29,31–33,35–40] and animal PVs [41–43] (Figure 1). In this section, we will discuss about several methods to identify and characterize PVs genome.

Various molecular methods, such as PCR have been used to detect specific viruses or viral families. PCR can be used to detect a few copies of a particular nucleic acid, and the viral load in a sample can be quantified with real-time PCR [44].

Although many of the PCR assays investigate the presence of a specific virus, assays have been developed to detect several viruses simultaneously. The most used molecular method to study PVs through its genome is the PCR. This methodology has been widely used to detect a broad range of PVs types in human and animals. PCR-based method has been routinely used due to high sensitivity (PCR-based method can be used to detect a few copies of PVs DNA and viral load samples) and specificity [45–48].

In addition, multiple-primed rolling circle amplification (RCA) technique, hybridization, and metagenomic methods have became a convenient tool for molecular biological research on PVs [10]. Thus, the development of molecular techniques reflected advances in knowledge of PVs genomes. Particularly, viral metagenomics have provided a powerful technology to investigate the viral flora of healthy and sick human and animals [49]. By applying these methodologies to PVs studies, it is possible to gain a better understanding in the etiology of PVs, as well as deepen the knowledge into the PVs and others virus circulating in nature, such as PVs in flies, ticks, fomites and body fluids.

Also, it may provide a better understanding of the complex interaction between virus and host.

PCR-Based Methods

Nowadays, there are several molecular assays that use PCR methods for investigate PVs. These methodologies essentially differ in the PVs gene segment amplified and in the number of primers used. For clinical studies of PVs prevalence, the most frequently used primer sets are MY9/MY11, FAP59/64, and PGMY9/ PGMY11. These consensus primer sets were designed to amplify a partial sequence of highly conserved PVs L1 gene. Also, these primers are made with the use of degenerate nucleotides that increase the range of PVs detection, however, may lower its sensitivity [50]. The MY9/MY11 primer set is the most commonly method to detect HPV infection in cervical samples [51]. These primers amplify approximately 450 base pair (bp) fragments and they allow amplification of 47 HPV DNA types. The PGMY09/PGMY11 primer set is an extended version of MY09/11 primer set, which permits to increase the sensitivity to detect HPV types [52]. Alternatively, the association of fixed nucleotide primers, such as GP5+/GP6+, with degenerate nucleotide primers may be employed to increase the sensitivity to HPV.

An alternative PCR approach using primers FAP59/64 also amplify a broad spectrum of HPV types. As well as the MY09/11 primer set, the FAP59/64 primers have degenerate nucleotides and were designed to amplify a partial sequence of highly conserved PVs L1 gene [53]. The primers FAP59/64 allow the detection of 46 HPV types. Apart from HPV, FAP59/64 and MY09/11 can be used to identify a broad range of animal PVs DNA. For instance, Ogawa et al. [22] showed that MY09/11 primers allow to detect BPV1 and BPV3. Moreover, the primers FAP59/64 can detect 13 BPV types [27,28] as well as, possible putative new BPV types [22]. Furthermore, the use of FAP59/64 primers allowed to identify PVs in chimpanzees, gorillas, macaques, monkeys, lemurs, cows and elks [54].

More recently, real-time PCR has been used to detect PVs as well as viral load assessment. This method is based on the release of fluorescence at each amplification cycle, which is directly proportional to the amount of amplicon generated. With regard to Human PVs, studies have showed that real-time PCR can be used to quantify viral load, detection and genotyping in cervical lesion or cervical cancer [55–58]. Moreover, real-time PCR has been used to detect and quantify viral load of animal PV types, such as BPV1, BPV2, BPV12 [17,59–62].

Apart from the clinical importance, the use of PCR-based methods and direct sequencing allowed to demonstrate the genetic diversity within PVs. This genetic diversity of PVs DNA can occur due to synonymous and nonsynonymous mutations [35,36,63], insertions [64] and recombination [65]. These changes can alter the structure and the function of proteins and, consequently, their biological activity [66–69]. For instance, several studies have been demonstrated that variants of HPV16, HPV18, HPV31, HPV33, HPV58 and other HPV types are related to persistent infection, progression and oncogenicity [66,67,69–71]. Moreover, studies that used PCR-based method allowed demonstrating that the genetic diversity within HPV16 and HPV18 DNA co-evolved with the three major human phylogenetic branches: African, Caucasian and Asian. Therefore, the PCR-based method is an important approach used in clinical as well as evolutionary studies with regard PVs.

Although PCR-based methods have high sensitivity and specificity, this methodology presents some disadvantages, such as the necessity of specific primer to detect PVs and short amplified fragments. In contrast, multiple primed RCA method has been used to detect new PVs. The rolling circle sequence-independent amplification technique makes use of the property of circular DNA molecules such as plasmids or viral genomes replicating through a

rolling circle mechanism. Briefly, the RCA method is a DNA synthesis reaction that uses the phi29 DNA polymerase to amplify circular DNA molecules [72].

The uses of random hexamers primers, which bind at multiple locations on a circular DNA template, allow the amplification of circular DNA without requiring prior knowledge of the sequence. Recently, several studies showed the detection of new PVs using the RCA method. Although technically more demanding than other methods of sequence-independent amplification, the RCA approach has facilitated the identification of a novel PVs. For instance, the use of RCA allowed to identify 20 new HPV types in the *Betapapillomavirus* genus, 47 in *Gamma papillomavirus* genus, and 7 new HPV types in *Alphapapillomavirus* genus [13].

Similarly, RCA allowed to identify new animal PVs, such as the Canine Papillomavirus (CPV) [73,74], Feline Papillomavirus (FdPV) [75], Equine Papillomavirus (EcPV) [76], Mouse Papillomavirus (MusPV) [77], Bovine Papillomavirus (BPV) [78], *Bettongiapienicillata* Papillomavirus (BpPV) [79], *Cervuselaphus* Papillomavirus (CePV)[80] and *M. fascicularis* Papillomavirus type 1 (MfPV-1)[81].

Metagenomic Methods

As previously mentioned, several microorganisms cannot be cultivated, and therefore, some may go unnoticed and unstudied. Molecular methods widely employed such as PCR have the limitation of being used to detect specific viruses or viral families. However, metagenomics is a more general approach to study the genetic composition of uncultured samples [82]. Viral metagenomics investigates the complete genetic viral population in the sample studied [83]. These studies have also contributed to knowledge of undiscovered or low studied PVs due to its presence in no conventional sites.

In the first assays of metagenomics, the products of the amplification technologies have usually been cloned into bacteria to create libraries. Then, these cloned products were characterized by Sanger sequencing to identify any potential viruses. However, this process is quite laborious, and due to a combination of the high background of contaminating host nucleic acid and the occasionally low levels of virus, a vast number of clones may have to be sequenced before a viral sequence is identified [49]. One of these techniques, the DNase-SISPA assays includes sequence-independent amplification, cloning and sequencing of amplified viral nucleic fragments followed by in silico searches for sequence similarities to known viruses. This method requires a prior digestion of non-pathogen specific nucleic acid before viral DNA/RNA. DNase-SISPA technique have been widely used for identifying known and unknown viruses [84–86], including PVs [33].

Other independent platforms were introduced for sequencing with no need for traditional cloning after amplification and yield several hundred thousand sequence reads in one run. These techniques have different sequencing principles but all have high throughput (relative to Sanger sequencing) in common and are often referred to as next-generation sequencing or high throughput sequencing [87, 88]. For metagenomic studies, 454 sequencing is often used, mainly due to the longer read length, which makes de novo assembly easier.

Next-generation sequencing (NGS)-based metagenomic have allowed the identification of several PVs, including new PVs. Hence, the use of sequencing technologies such as the 454 sequencing platform and high throughput sequencing (metagenomic sequencing and analysis) has permitted to identify new PVs in several biological samples. With regard to Human PVs,

the use of metagenomic method identified the genomes of HPV116 in rectal swab [89] and nine putative HPV in cutaneous lesions skin [90]. Moreover, the use of this methodology seems to be important to identify PVs infection in HPV-negative samples [91] as well as sites of less common infections [92]. With regard to animal PVs, the metagenomic method allowed to identify BPV10 in teat wart of cattle [33], PVs in mosquitoes [42], a novel *Miniopterusschreibersii* Papillomavirus type 1 (MscPV1) and *Rousettusaegyptiacus* Papillomavirus type 1(RaP1) in bats [38, 93] and Tornovirus 1 (STTV1) in sea turtle [42].

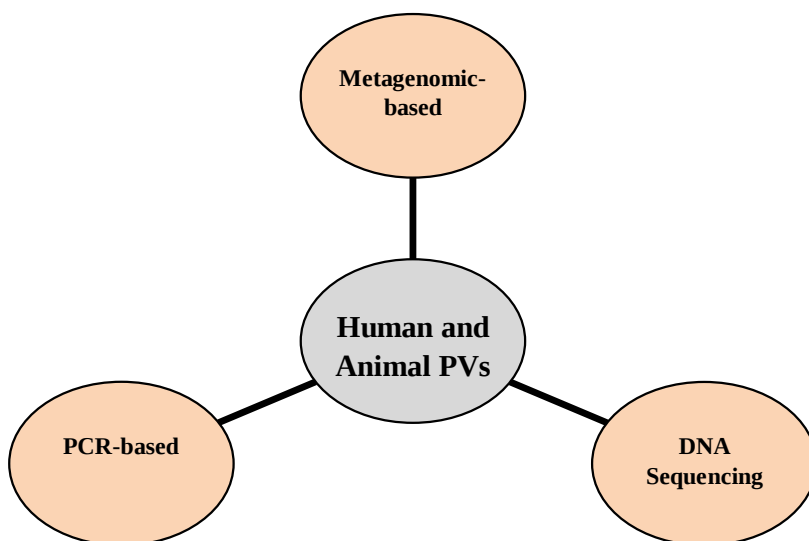


Figure 1. Current methodologies used to detect Human and Animal PVs.

A MODEL FOR RESEARCH IN PAPILLOMAVIRUS: ADVANCES IN BOVINE PAPILLOMAVIRUS KNOWLEDGE

Among PVs, BPV have a major role in veterinary medicine. They can cause benign and malignant lesions in cattle and are associated with horse, zebra, buffalo, yaks, giraffes, tapirs and bison lesions [17, 26, 40, 94, 95]. BPV has been widely distributed in cattle worldwide [96]. There are reports of the presence of BPV in the Americas [28], Europe [97], Asia [98], Oceania [99] and Africa [26]. BPV is considered the etiological agent of Bovine Papillomatosis and cancers of the bladder and upper gastrointestinal tract in its natural host. Although there is no global survey on the economic impact of BPV, the numbers about world cattle population that exceeded 1 billion head of cattle reflect the importance of studying BPV. Developing countries like India, Brazil and China have the largest cattle population in the world, whereas USA, Brazil and China are the largest beef producer, respectively [100]. Cutaneous papillomas are responsible for significant economic damages to farmers due to the retarded growth of the animals, loss of weight caused by the reduction in food consumption and decrease in milk production [101,102]. Since lesions are commonly found in young animals whose immune system is still developing, its appearance is an important observation for calves because of the morbidity and mortality and high costs of treatment [103].

The presence of mucosal papillomas or tumors may progress to cancer, and the upper gastrointestinal tract and the urinary bladder are the most common locations for the development of BPV-associated carcinomas [104]. The highest economic damages caused by BPV are found in cattle from tropical and subtropical regions due to the large presence of bracken fern and its consequent chronic ingestion by the animals [104].

To date, 13 BPV types have been identified and classified into three genera: *Deltapapillomavirus* (BPV1, 2 and 13), *Epsilonpapillomavirus* (BPV5 and 8) and *Xipapillomavirus* (BPV3, 4, 6, 9, 10, 11 and 12). One BPV type is still unclassified (BPV7) [78]. All BPV types characterized so far are associated with different histopathological lesions. *Xipapillomaviruses* are pure epitheliotropic (causing true papillomas), *Deltapapillomaviruses* induce fibropapillomas and *Epsilonpapillomaviruses* can cause both types of lesions [105]. Members of *Deltapapillomavirus* genus are also associated with infections in non-epithelial sites such as blood and semen [27,106].

Besides the importance of BPV in veterinary medicine, BPV has been also largely investigated as animal model to better understand the transforming activity of PVs. BPV attracted the interest of molecular biologists since it was the first PVs able to induce transformation in cultured non-epithelial cells; furthermore its genome was the first among PVs to be completely sequenced [107]. Despite its importance, the understanding of BPV diversity is limited, probably underestimated. In contrast with HPV that presents more than 150 described HPV genomes, less than 70 non-human PVs species have been isolated and sequenced so far. However, the numbers of studies have attempted to assess the diversity and characterization of BPV genome has increased in recent years. Beyond 13 BPVs described, about 30 new putative types were isolated from various geographical regions in countries such as Brazil, Japan and Sweden. For example, BPV7, 8, 9 and 10 were designated from putative BPV types BAPV6, 2, BPV type I and BPV type II, respectively [14, 22, 54, 98, 108–111]. These studies usually employ the PCR technique with consensus primers, and as commonly these studies describe new putative types and subtypes of BPV, it is believed that there may be an underestimation of the identification of BPV types [112].

BPV are also classified into subtypes and variants, which are useful to the knowledge of biology and diversification of PVs. The differences found in nucleotide sequences of variants of PVs could be responsible for changes in the oncogenic potential, cellular location, host immune response, function of PVs proteins (affecting pathogenesis), or binding capacity of a transcription factor [5, 35, 113, 114]. A few studies related to BPV genetic variability were done. The majority of these studies were associated with equine sarcoid, aiming at understanding differences between the disease in bovine and equine [109,115].

The improvement of molecular techniques for detection of HPV allowed the discovery of these various putative new BPV types, besides the discovery of frequent multiple infections in cattle caused by various BPV types in a single skin lesion [28, 97, 116]. BPV DNA is detected by a variety of PCR-based techniques. These PCRs are based frequently on the detection of one or two BPV types using degenerated or type-specific primers. Genotyping is performed either by real-time detection [33] or by sequence analysis [117] or restriction fragment length polymorphism (RFLP) analyses [18] of the generated PCR fragments. Usually the discovery of putative new types, subtypes and variant are made in works that employs consensus primers capable of identifying potentially more than two BPV types [22]. This methodology has suggested the existence of numerous yet uncharacterized BPV types in cattle herds from diverse geographical regions. Beyond the use of consensus primers designed for HPV, such as

FAP59/FAP64 and MY09/MY11 successful in identifying cutaneous and mucosal PVs types, respectively [118,119], other consensus primers designed exclusively for L1 of BPV have been used [98].

There are reports indicating that consensus primers sometimes fail to diagnose a Papillomavirus infection because of sequence differences between consensus primers and putative PV types [54]. Beyond the use of these primers, the multiple-primed RCA technique has been applied to amplify whole genomes of PVs [120,121] and also BPV [111] and became a convenient tool for molecular biological research on PVs. Also, in cattle other sequence-independent amplification techniques like DNase-SISPA have been use for characterizing genome of BPV. Due to this plasticity, BPV genomes and BPV-like genome has been detected in several sites and hosts [27, 27, 96, 122], and metagenomic methods are very suitable for samples like plasma, serum, soft tissues, respiratory secretions, cerebrospinal fluid, urine, faeces and filterable environmental [83, 86]. Therefore, these new methods seem to be very suitable for BPV studies.

The advances in knowledge of BPV genomes and putative new genomes have been reflected by the development of molecular techniques and enabled new strategies of studies with the purpose of prevention and treatment of Papillomaviruses in cattle and other animals.

REFERENCES

- [1] Howley PM: Papillomaviruses and Their Replication. In: Knipe, D.M., Howley, P.M. (Eds.), *Fields Virology*, vol. 2. Lippincott, Williams & Wilkins, pp. 3634–3687.
- [2] Zheng Z-M, Baker CC: Papillomavirus genome structure, expression, and post-transcriptional regulation. *Front Biosci. J. Virtual Libr.* 2006;11:2286–2302.
- [3] Jia R, Zheng Z-M: Regulation of Bovine Papillomavirus type 1 gene expression by RNA processing. *Front Biosci. J. Virtual Libr.* 2009
- [4] Campo MS. Bovine Papillomavirus: old system, new lessons? In: Campo MS (ed) *Papillomavirus research: from natural history to vaccine and beyond. Caister Academic Press*, (2006) Wymondham.
- [5] Bernard H-U: Gene expression of genital Human Papillomaviruses and considerations on potential antiviral approaches. *Antivir. Ther.* 2002;7:219–237.
- [6] De Villiers E-M, Fauquet C, Broker TR, Bernard H-U, zur Hausen H: Classification of Papillomaviruses. *Virology* 2004 20;324:17–27.
- [7] Bernard H-U, Burk RD, Chen Z, van Doorslaer K, Hausen H zur, de Villiers E-M: Classification of Papillomaviruses (PVs) based on 189 PV types and proposal of taxonomic amendments. *Virology* 2010 25;401:70–79.
- [8] Bernard H-U, Calleja-Macias IE, Dunn ST: Genome variation of Human Papillomavirus types: phylogenetic and medical implications. *Int. J. Cancer J. Int. Cancer* 2006 1;118:1071–1076.
- [9] Van Regenmortel M, International, Fauquet C: *Virus Taxonomy: Classification and Nomenclature of Viruses: Seventh Report of the International Committee on Taxonomy of Viruses*. Academic Press, 2000.
- [10] Bexfield N, Kellam P: Metagenomics and the molecular identification of novel viruses. *Vet. J. Lond. Engl.* 2011;190:191–198.

- [11] Belak S, Thoren P, LeBlanc N, Viljoen G: Advances in viral disease diagnostic and molecular epidemiological technologies. *Expert Rev. Mol. Diagn.* 2009;9:367–381.
- [12] Geimanen J, Isok-Paas H, Pipitch R, Salk K, Laos T, Orav M, et al.: Development of a cellular assay system to study the genome replication of high- and low-risk mucosal and cutaneous Human Papillomaviruses. *J. Virol.* 2011 1;85:3315–3329.
- [13] De Villiers E-M: Cross-roads in the classification of Papillomaviruses. *Virology* DOI: 10.1016/j.virol.2013.04.023
- [14] Maeda Y, Shibahara T, Wada Y, Kadota K, Kanno T, Uchida I, et al.: An outbreak of teat papillomatosis in cattle caused by Bovine Papillomavirus (BPV) type 6 and unclassified BPVs. *Vet. Microbiol.* 2007;121:242–248.
- [15] Xi LF, Demers GW, Koutsky LA, Kiviat NB, Kuypers J, Watts DH, et al.: Analysis of Human Papillomavirus Type 16 Variants Indicates Establishment of Persistent Infection. *J. Infect. Dis.* 1995;172:747–755.
- [16] Wu Y, Chen Y, Li L, Yu G, He Y, Zhang Y: Analysis of mutations in the E6/E7 oncogenes and L1 gene of Human Papillomavirus 16 cervical cancer isolates from China. *J. Gen. Virol.* 2006;87:1181–1188.
- [17] Bogaert L, Martens A, Kast WM, Van Marck E, De Cock H: Bovine Papillomavirus DNA can be detected in keratinocytes of equine sarcoid tumors. *Vet. Microbiol.* 2010 15;146:269–275.
- [18] Carr EA, Théon AP, Madewell BR, Griffey SM, Hitchcock ME: Bovine Papillomavirus DNA in neoplastic and nonneoplastic tissues obtained from horses with and without sarcoids in the western United States. *Am. J. Vet. Res.* 2001;62:741–744.
- [19] Stocco dos Santos RC, Lindsey CJ, Ferraz OP, Pinto JR, Mirandola RS, Benesi FJ, et al.: Bovine Papillomavirus transmission and chromosomal aberrations: an experimental model. *J. Gen. Virol.* 1998;79:2127–2135.
- [20] Wosiacki SR, Claus MP, Alfieri AF, Alfieri AA: Bovine Papillomavirus type 2 detection in the urinary bladder of cattle with chronic enzootic haematuria. *Memórias Inst. Oswaldo Cruz* 2006;101:635–638.
- [21] Borzacchiello G, Ambrosio V, Roperto S, Poggiali F, Tsirimonakis E, Venuti A, et al.: Bovine Papillomavirus type 4 in oesophageal papillomas of cattle from the south of Italy. *J. Comp. Pathol.* 2003;128:203–206.
- [22] Ogawa T, Tomita Y, Okada M, Shinozaki K, Kubonoya H, Kaiho I, et al.: Broad-spectrum detection of Papillomaviruses in bovine teat papillomas and healthy teat skin. *J. Gen. Virol.* 2004;85:2191–2197.
- [23] Van Dyk E, Bosman AM, van Wilpe E, Williams JH, Bengis RG, van Heerden J, et al.: Detection and characterisation of Papillomavirus in skin lesions of giraffe and sable antelope in South Africa. *J. S. Afr. Vet. Assoc.* 2011;82:80–85.
- [24] Gnanamony M, Peedicayil A, Subhashini J, Ram TS, Rajasekar A, Gravitt P, et al.: Detection and quantitation of HPV 16 and 18 in plasma of Indian women with cervical cancer. *Gynecol. Oncol.* 2010;116:447–451.
- [25] Ho C-M, Yang S-S, Chien T-Y, Huang S-H, Jeng C-J, Chang S-F: Detection and quantitation of Human Papillomavirus type 16, 18 and 52 DNA in the peripheral blood of cervical cancer patients. *Gynecol. Oncol.* 2005;99:615–621.
- [26] Van Dyk E, Oosthuizen MC, Bosman A-M, Nel PJ, Zimmerman D, Venter EH: Detection of Bovine Papillomavirus DNA in sarcoid-affected and healthy free-roaming zebra (*Equus zebra*) populations in South Africa. *J. Virol. Methods* 2009;158:141–151.

- [27] Silva MAR, Pontes NE, Da Silva KMG, Guerra MMP, Freitas AC: Detection of Bovine Papillomavirus type 2 DNA in commercial frozen semen of bulls (*Bos taurus*). *Anim. Reprod. Sci.* 2011;129:146–151.
- [28] Carvalho CCR, Batista MVA, Silva MAR, Balbino VQ, Freitas AC: Detection of Bovine Papillomavirus types, co-Infection and a putative new BPV11 subtype in cattle. *Transbound. Emerg. Dis.* 2012 Jan 9; DOI: 10.1111/j.1865-1682.2011.01296.x
- [29] Leishangthem GD: Detection of Bovine Papillomaviruses in cutaneous warts/papillomas in cattle. *Indian J. Vet. Pathol.* 32:15–20.
- [30] Shen Z-Y, Hu S-P, Lu L-C, Tang C-Z, Kuang Z-S, Zhong S-P, et al.: Detection of Human Papillomavirus in esophageal carcinoma. *J. Med. Virol.* 2002;68:412–416.
- [31] Baldez da Silva MFPT, Chagas BS, Guimaraes V, Katz LMC, Felix PM, Miranda PM, et al.: HPV31 and HPV33 incidence in cervical samples from women in Recife, *Brazil*. *Genet. Mol. Res.* 2009;8:1437–1443.
- [32] Badaracco G, Venuti A, Morello R, Muller A, Marcante ML: Human Papillomavirus in head and neck carcinomas: prevalence, physical status and relationship with clinical/pathological parameters. *Anticancer Res.* 2000;20:1301–1305.
- [33] Rai GK, Saxena M, Singh V, Somvanshi R, Sharma B: Identification of Bovine Papillomavirus 10 in teat warts of cattle by DNase-SISPA. *Vet. Microbiol.* 2011;147:416–419.
- [34] Finan RR, Tamim H, Almawi WY: Identification of Chlamydia trachomatis DNA in Human Papillomavirus (HPV) positive women with normal and abnormal cytology. *Arch. Gynecol. Obstet.* 2002;266:168–171.
- [35] Chagas BS, Batista MVA, Guimaraes V, Balbino VQ, Crovella S, Freitas AC: New variants of E6 and E7 oncogenes of Human Papillomavirus type 31 identified in Northeastern Brazil. *Gynecol. Oncol.* 2011;123:284–288.
- [36] Chagas BS, Batista MV de A, Crovella S, Gurgel APAD, Silva Neto J da C, Serra IGSS, et al.: Novel E6 and E7 oncogenes variants of Human Papillomavirus type 31 in Brazilian women with abnormal cervical cytology. *Infect. Genet. Evol.* 2013;16C:13–18.
- [37] Watts KJ, Thompson CH, Cossart YE, Rose BR: Sequence variation and physical state of Human Papillomavirus type 16 cervical cancer isolates from Australia and New Caledonia. *Int. J. Cancer J. Int. Cancer* 2002;97:868–874.
- [38] Tse H, Tsang AKL, Tsoi H-W, Leung ASP, Ho C-C, Lau SKP, et al.: Identification of a novel bat Papillomavirus by metagenomics. *Plos One* 2012;7:e43986.
- [39] Ge X, Li Y, Yang X, Zhang H, Zhou P, Zhang Y, et al.: Metagenomic analysis of Viruses from the bat fecal samples reveals many novel viruses in insectivorous bats in China. *J. Virol.* 2012; DOI: 10.1128/JVI.06671-11.
- [40] Pangty K, Singh S, Goswami R, Saikumar G, Somvanshi R: Detection of BPV-1 and -2 and Quantification of BPV-1 by Real-Time PCR in Cutaneous Warts in Cattle and Buffaloes. *Transbound. Emerg. Dis.* 2010;57:185–196.
- [41] Christian L, Claude F, Mathias A, Jessica G, Elisabeth V, Kurt T: Novel snake Papillomavirus does not cluster with other non-mammalian papillomaviruses. *Virol. J.* 2011;8:4362011.
- [42] Ng TFF, Manire C, Borrowman K, Langer T, Ehrhart L, Breitbart M: Discovery of a novel single-stranded DNA virus from a sea turtle fibropapilloma by using viral metagenomics. *J. Virol.* 2009;83:2500–2509.

- [43] Herbst LH, Lenz J, Van Doorslaer K, Chen Z, Stacy BA, Wellehan Jr. JFX, et al.: Genomic characterization of two novel reptilian Papillomaviruses, *Chelonia mydas* Papillomavirus 1 and *Caretta caretta* Papillomavirus 1. *Virology* 2009;383:131–135.
- [44] Mackay IM, Arden KE, Nitsche A: Real-time PCR in virology. *Nucleic. Acids Res.* 2002;30:1292–1305.
- [45] Gravitt PE, Peyton C, Wheeler C, Apple R, Higuchi R, Shah KV: Reproducibility of HPV 16 and HPV 18 viral load quantitation using TaqMan real-time PCR assays. *J. Virol. Methods* 2003;112:23–33.
- [46] Hesselink AT, Meijer CJLM, Poljak M, Berkhof J, Kemenade FJ van, Salm ML van der, et al.: Clinical validation of the Abbott RealTime High Risk (HR) HPV assay according to the guidelines for HPV DNA test requirements for cervical screening. *J. Clin. Microbiol.* 2013; DOI: 10.1128/JCM.00633-13
- [47] Nagao S, Yoshinouchi M, Miyagi Y, Hongo A, Kodama J, Itoh S, et al.: Rapid and sensitive detection of physical status of Human Papillomavirus type 16 DNA by quantitative real-time PCR. *J. Clin. Microbiol.* 2002;40:863–867.
- [48] Peitsaro P, Johansson B, Syrjänen S: Integrated Human Papillomavirus type 16 is frequently found in cervical cancer precursors as demonstrated by a novel quantitative real-time PCR technique. *J. Clin. Microbiol.* 2002;40:886–891.
- [49] Blomström A-L: *Applications of viral metagenomics in the veterinary field.* 2010.
- [50] Qu W, Jiang G, Cruz Y, Chang CJ, Ho GY, Klein RS, et al.: PCR detection of Human Papillomavirus: comparison between MY09/MY11 and GP5+/GP6+ primer systems. *J. Clin. Microbiol.* 1997;35:1304–1310.
- [51] Resnick RM, Cornelissen MTE, Wright DK, Eichinger GH, Fox HS, Schegget J ter, et al.: Detection and typing of Human Papillomavirus in archival cervical cancer specimens by DNA amplification with consensus primers. *J. Natl. Cancer Inst.* 1990;82:1477–1484.
- [52] Gravitt PE, Peyton CL, Alessi TQ, Wheeler CM, Coutlée F, Hildesheim A, et al.: Improved Amplification of Genital Human Papillomaviruses. *J. Clin. Microbiol.* 2000;38:357–361.
- [53] Forslund O: Genetic diversity of cutaneous human papillomaviruses. *J. Gen. Virol.* 2007;88:2662–2669.
- [54] Antonsson A, Hansson BG: Healthy skin of many animal species harbors Papillomaviruses which are closely related to their human counterparts. *J. Virol.* 2002;76:12537–12542.
- [55] Liao Y, Zhou Y, Guo Q, Xie X, Luo E, Li J, et al.: Simultaneous detection, genotyping, and quantification of Human Papillomaviruses by multicolor real-Time PCR and melting curve analysis. *J. Clin. Microbiol.* 2013;51:429–435.
- [56] Seaman WT, Andrews E, Couch M, Kojic EM, Cu-Uvin S, Palefsky J, et al.: Detection and quantitation of HPV in genital and oral tissues and fluids by real time PCR. *Virol. J.* 2010;7:194.
- [57] Iftner T, Germ L, Swoyer R, Kjaer SK, Breugelmans JG, Munk C, et al.: Study comparing Human Papillomavirus (HPV) real-time multiplex PCR and hybrid capture II INNO-LiPA v2 HPV genotyping PCR Assays. *J. Clin. Microbiol.* 2009;47:2106–2113.
- [58] Micalessi MI, Boulet GA, Vorsters A, De Wit K, Jannes G, Mijs W, et al.: A real-time PCR approach based on SPF10 primers and the INNO-LiPA HPV genotyping extra assay

- for the detection and typing of Human Papillomavirus. *J. Virol. Methods* 2013;187:166–171.
- [59] Pathania S, Dhama K, Saikumar G, Shahi S, Somvanshi R: Detection and quantification of Bovine Papillomavirus type 2 (BPV-2) by real-time PCR in urine and urinary bladder lesions in enzootic bovine haematuria (EBH)-affected cows. *Transbound. Emerg. Dis.* 2012;59:79–84.
- [60] Dong J, Zhu W, Goto Y, Haga T: Initial detection of a circular genome deletion in a naturally Bovine Papillomavirus-infected sample. *J. Vet. Med. Sci.* 2013;75:179–182.
- [61] Haralambus R, Burgstaller J, Klukowska-Rötzler J, Steinborn R, Buchinger S, Gerber V, et al.: Intralesional Bovine Papillomavirus DNA loads reflect severity of equine sarcoid disease. *Equine Vet. J.* 2010;42:327–331.
- [62] Roperto S, Russo V, Ozkul A, Sepici-Dincel A, Maiolino P, Borzacchiello G, et al.: Bovine Papillomavirus type 2 infects the urinary bladder of water buffalo (*Bubalus bubalis*) and plays a crucial role in bubaline urothelial carcinogenesis. *J. Gen. Virol.* 2012;94:403–408.
- [63] De Freitas AC, Gurgel APAD, Chagas BS, Coimbra EC, do Amaral CMM: Susceptibility to cervical cancer: An overview. *Gynecol. Oncol.* 2012;126:304–311.
- [64] Campitelli M, Jeannot E, Peter M, Lappartient E, Saada S, de la Rochefordiere A, et al.: Human Papillomavirus mutational insertion: specific marker of circulating tumor DNA in cervical cancer patients. *Plos One* 2012;7:e43393.
- [65] Robles-Sikisaka R, Rivera R, Nollens HH, St Leger J, Durden WN, Stolen M, et al.: Evidence of recombination and positive selection in cetacean Papillomaviruses. *Virology* 2012;427:189–197.
- [66] Yamada T, Manos MM, Peto J, Greer CE, Munoz N, Bosch FX, et al.: Human Papillomavirus type 16 sequence variation in cervical cancers: a worldwide perspective. *J. Virol.* 1997;71:2463–2472.
- [67] Londesborough P, Ho L, Terry G, Cuzick J, Wheeler C, Singer A: Human Papillomavirus genotype as a predictor of persistence and development of high-grade lesions in women with minor cervical abnormalities. *Int. J. Cancer J. Int. Cancer* 1996;69:364–368.
- [68] Zehbe I, Lichtig H, Westerback A, Lambert PF, Tommasino M, Sherman L: Rare Human Papillomavirus 16 E6 variants reveal significant oncogenic potential. *Mol. Cancer* 2011;10:77.
- [69] Schiffman M, Rodriguez AC, Chen Z, Wacholder S, Herrero R, Hildesheim A, et al.: A Population-Based prospective study of carcinogenic Human Papillomavirus variant lineages, viral persistence, and cervical neoplasia. *Cancer Res.* 2010;70:3159–3169.
- [70] Cento V, Rahmatalla N, Ciccozzi M, Perno CF, Ciotti M: Intratype variations of HPV 31 and 58 in Italian women with abnormal cervical cytology. *J. Med. Virol.* 2011;83:1752–1761.
- [71] Raiol T, Wyant PS, de Amorim RMS, Cerqueira DM, Milanezi N von G, Brigido M de M, et al.: Genetic variability and phylogeny of the high-risk HPV-31, -33, -35, -52, and -58 in central Brazil. *J. Med. Virol.* 2009;81:685–692.
- [72] Johne R, Müller H, Rector A, van Ranst M, Stevens H: Rolling-circle amplification of viral DNA genomes using phi29 polymerase. *Trends Microbiol.* 2009;17:205–211.
- [73] Lange CE, Tobler K, Lehner A, Vetsch E, Favrot C: A case of a canine pigmented plaque associated with the presence of a Xipapillomavirus. *Vet. Dermatol.* 2012;23:76–e19.

- [74] Lange CE, Ackermann M, Favrot C, Tobler K: Entire Genomic Sequence of Novel Canine Papillomavirus Type 13. *J. Virol.* 2012;86:10226–10227.
- [75] Lange CE, Tobler K, Markau T, Alhaidari Z, Bornand V, Stöckli R, et al.: Sequence and classification of FdPV2, a Papillomavirus isolated from feline Bowenoid in situ carcinomas. *Vet. Microbiol.* 2009;137:60–65.
- [76] Scase T, Brandt S, Kainzbauer C, Sykora S, Bijmolt S, Hughes K, et al.: *Equus caballus* Papillomavirus-2 (EcPV-2): An infectious cause for equine genital cancer? *Equine Vet. J.* 2010;42:738–745.
- [77] Joh J, Jenson AB, King W, Proctor M, Ingle A, Sundberg JP, et al.: Genomic analysis of the first laboratory-mouse Papillomavirus. *J. Gen. Virol.* 2011;92:692–698.
- [78] Lunardi M, Alfieri AA, Otonel RAA, de Alcântara BK, Rodrigues WB, de Miranda AB, et al.: Genetic characterization of a novel Bovine Papillomavirus member of the *Deltapapillomavirus* genus. *Vet. Microbiol.* 2013;162:207–213.
- [79] Bennett MD, Reiss A, Stevens H, Heylen E, Ranst MV, Wayne A, et al.: The first complete Papillomavirus genome characterized from a marsupial host: a novel isolate from *Bettongia penicillata*. *J. Virol.* 2010;84:5448–5453.
- [80] Scagliarini A, Gallina L, Battilani M, Turrini F, Savini F, Lavazza A, et al.: *Cervus elaphus* Papillomavirus (CePV1): New insights on viral evolution in deer. *Vet. Microbiol.* DOI: 10.1016/j.vetmic.2013.03.012
- [81] Joh J, Hopper K, Doorslaer KV, Sundberg JP, Jenson AB, Ghim S-J: *Macaca fascicularis* Papillomavirus type 1: a non-human primate *Betapapillomavirus* causing rapidly progressive hand and foot papillomatosis. *J. Gen. Virol.* 2009;90:987–994.
- [82] Handelsman J: Metagenomics: Application of Genomics to Uncultured Microorganisms. *Microbiol. Mol. Biol. Rev.* 2004;68:669–685.
- [83] Delwart EL: Viral metagenomics. *Rev. Med. Virol.* 2007;17:115–131.
- [84] Allander T, Emerson SU, Engle RE, Purcell RH, Bukh J: A virus discovery method incorporating DNase treatment and its application to the identification of two Bovine Parvovirus species. *Proc. Natl. Acad. Sci. USA* 2001;98:11609–11614.
- [85] Breitbart M, Rohwer F: Method for discovering novel DNA viruses in blood using viral particle selection and shotgun sequencing. *BioTechniques* 2005;39:729–736.
- [86] Ambrose HE, Clewley JP: Virus discovery by sequence-independent genome amplification. *Rev. Med. Virol.* 2006;16:365–383.
- [87] Kircher M, Kelso J: High-throughput DNA sequencing-concepts and limitations. *Bioessays* 2010;32:524–536.
- [88] Shendure J, Ji H: Next-generation DNA sequencing. *Nat. Biotechnol.* 2008;26:1135–1145.
- [89] Li L, Barry P, Yeh E, Glaser C, Schnurr D, Delwart E: Identification of a novel Human *Gammapapillomavirus* species. *J. Gen. Virol.* 2009;90:2413–2417.
- [90] Ekström J, Bzhalava D, Svenback D, Forslund O, Dillner J: High throughput sequencing reveals diversity of Human Papillomaviruses in cutaneous lesions. *Int. J. Cancer J. Int. Cancer* 2011;129:2643–2650.
- [91] Johansson H, Bzhalava D, Ekström J, Hultin E, Dillner J, Forslund O: Metagenomic sequencing of “HPV-negative” condylomas detects novel putative HPV types. *Virology* 2013;440:1–7.

- [92] Mokili JL, Dutilh BE, Lim YW, Schneider BS, Taylor T, Haynes MR, et al.: Identification of a novel Human Papillomavirus by metagenomic analysis of samples from patients with febrile respiratory illness. *Plos One* 2013;8:e58404.
- [93] Baker KS, Leggett RM, Bexfield NH, Alston M, Daly G, Todd S, et al.: Metagenomic study of the viruses of African straw-coloured fruit bats: Detection of a chiropteran poxvirus and isolation of a novel adenovirus. *Virology* 2013;441:95–106.
- [94] Löhr CV, Juan-Sallés C, Rosas-Rosas A, García AP, Garner MM, Teifke JP: Sarcoids in Captive zebras (*Equus burchelli* II): Association with Bovine Papillomavirus type I infection. *J. Zoo Wildl. Med.*;36:74–81.
- [95] Silvestre O, Borzacchiello G, Nava D, Iovane G, Russo V, Vecchio D, et al.: Bovine Papillomavirus Type 1 DNA and E5 Oncoprotein Expression in Water Buffalo Fibropapillomas. *Vet. Pathol. Online* 2009;46:636–641.
- [96] Freitas AC: Recent insights into Bovine Papillomavirus. *Afr. J. Microbiol. Res.* 2011;5. DOI: 10.5897/AJMRX11.020
- [97] Schmitt M, Fiedler V, Müller M: Prevalence of BPV genotypes in a German cowshed determined by a novel multiplex BPV genotyping assay. *J. Virol. Methods* 2010;170:67–72.
- [98] Hatama S, Ishihara R, Ueda Y, Kanno T, Uchida I: Detection of a novel Bovine Papillomavirus type 11 (BPV-11) using *Xipapillomavirus* consensus polymerase chain reaction primers. *Arch. Virol.* 2011;156:1281–1285.
- [99] Munday JS, Knight CG: Amplification of feline sarcoid-associated Papillomavirus DNA sequences from bovine skin. *Vet. Dermatol.* 2010;21:341–344.
- [100] Lowe M, Gereffi G. A value chain analysis of the U.S. beef and dairy industries. *Report prepared for Environmental Defense Fund.* 2009.
- [101] Meyers SA, Read WK: Squamous cell carcinoma of the vulva in a cow. *J. Am. Vet. Med. Assoc.* 1990;196:1644–1646.
- [102] Scott DW, Anderson WI. Bovine cutaneous neoplasms: literature review and retrospective analysis of 62 cases (1968 to 1990). *Comp. Cont. Educ. Pract. Vet.* 1992; 14:10.
- [103] Wellenberg GJ, van der Poel WHM, Van Oirschot JT: Viral infections and bovine mastitis: a review. *Vet. Microbiol.* 2002;88:27–45.
- [104] Hatama S. Cutaneous papillomatosis in cattle. *J. Disaster Res.* 2012.
- [105] Freitas ACF, Gurgel APAD, Chagas BS, Balbino VQ, Batista MVA. Genetic diversity of human and animal Papillomaviruses. In: Julian A, Cervantes Amaya, Miguel MFJ (Org.). *Genetic Diversity: New Research.* Hauppauge, NY: Nova Science Publishers, Inc., 2012.
- [106] Silva MAR, Silva KMG, Jesus ALS, Barros LO, Corteggio A, Altamura G, et al.: The presence and gene expression of Bovine Papillomavirus in the peripheral blood and semen of healthy horses. *Transbound Emerg. Dis* 2012; DOI: 10.1111/tbed.12036
- [107] Chen EY, Howley PM, Levinson AD, Seeburg PH: The primary structure and genetic organization of the Bovine Papillomavirus type 1 genome. *Nature* 1982;299:529–534.
- [108] Tomita Y, Literak I, Ogawa T, Jin Z, Shirasawa H: Complete genomes and phylogenetic positions of Bovine Papillomavirus type 8 and a variant type from a European bison. *Virus Genes.* 2007;35:243–249.

- [109] Claus MP, Lunardi M, Alfieri AF, Ferracin LM, Fungaro MHP, Alfieri AA: Identification of unreported putative new Bovine Papillomavirus types in Brazilian cattle herds. *Vet. Microbiol.* 2008;132:396–401.
- [110] Hatama S, Nobumoto K, Kanno T: Genomic and phylogenetic analysis of two novel Bovine Papillomaviruses, BPV-9 and BPV-10. *J. Gen. Virol.* 2008;89:158–163.
- [111] Zhu W, Dong J, Shimizu E, Hatama S, Kadota K, Goto Y, et al.: Characterization of novel Bovine Papillomavirus type 12 (BPV-12) causing epithelial papilloma. *Arch. Virol.* 2012;157:85–91.
- [112] Silva MAR, Batista MVA, Pontes NE, Santos EUD, Coutinho LCA, Castro RS, et al.: Comparison of two PCR strategies for the detection of Bovine Papillomavirus. *J. Virol. Methods* 2013;192:55–58.
- [113] Sichero L, Villa LL: Epidemiological and functional implications of molecular variants of Human Papillomavirus. *Braz. J. Med. Biol. Res.* 2006;39:707–717.
- [114] Chambers G, Ellsmore VA, O'Brien PM, Reid SWJ, Love S, Campo MS, et al.: Sequence variants of Bovine Papillomavirus E5 detected in equine sarcoids. *Virus Res.* 2003;96:141–145.
- [115] Jarrett WFH, McNeil PE, Grimshaw WTR, Selman IE, McIntyre WIM: High incidence area of cattle cancer with a possible interaction between an environmental carcinogen and a papilloma virus. *Nature* 1978;274:215–217.
- [116] Batista MVA, Silva MAR, Pontes NE, Reis MC, Corteggio A, Castro RS, et al.: Molecular epidemiology of bovine papillomatosis and the identification of a putative new virus type in Brazilian cattle. *Vet. J.* 2013; DOI: 10.1016/j.tvjl.2013.01.019
- [117] Brandt S, Haralambus R, Schoster A, Kirnbauer R, Stanek C: Peripheral blood mononuclear cells represent a reservoir of Bovine Papillomavirus DNA in sarcoid-affected equines. *J. Gen. Virol.* 2008;89:1390–1395.
- [118] Forslund O, Antonsson A, Nordin P, Stenquist B, Hansson BG: A broad range of Human Papillomavirus types detected with a general PCR method suitable for analysis of cutaneous tumours and normal skin. *J. Gen. Virol.* 1999;80 (Pt 9):2437–2443.
- [119] Manos MM, Ting Y, Wright DK, Lewis AJ, Broker TR, Wolinsky SM. The use of polymerase chain reaction amplification for the detection of genital Human Papillomaviruses. *Mol. Diagn. Human* 1989;209–214.
- [120] Rector A, Bossart GD, Ghim S-J, Sundberg JP, Jenson AB, Van Ranst M: Characterization of a novel close-to-root Papillomavirus from a Florida manatee by using multiply primed rolling-circle amplification: *Trichechus manatus latirostris* Papillomavirus Type 1. *J. Virol.* 2004;78:12698–12702.
- [121] Rector A, Tachezy R, Van Doorslaer K, MacNamara T, Burk RD, Sundberg JP, et al.: Isolation and cloning of a Papillomavirus from a North American porcupine by using multiply primed rolling-circle amplification: the *Erethizon dorsatum* Papillomavirus type 1. *Virology* 2005;331:449–456.
- [122] Lindsey CL, Almeida ME, Vicari CF, Carvalho C, Yagui A, Freitas AC, et al.: Bovine Papillomavirus DNA in milk, blood, urine, semen, and spermatozoa of bovine papillomavirus-infected animals. *Genet. Mol. Res.* 2009;8:310–318.

Chapter 99

SYNTHETIC SYNTHESIS OF VIRAL GENOMES

***Monique R. Eller¹, Roberto S. Dias², Rafael L. Salgado³,
Cynthia C. da Silva⁴ and Sérgio O. De Paula^{2,*}***

¹Department of Food Technology,

²Laboratory of Molecular Immunovirology,
Department of General Biology,

³Department of Biochemistry and Molecular Biology,

⁴Department of Microbiology, Federal University of Viçosa, Av. PH Rolfs, s/n,
Campus da UFV, Viçosa, Minas Gerais, Brazil

ABSTRACT

Viral particles are important tools in Molecular Biology, acting as carriers of genetic material, immunogenic antigens, adjuvants, or even directly combating antibiotic-resistant microorganisms and their biofilms in hospital and industrial environments. However, the efficient use of these particles requires extensive knowledge about their characteristics and components, including those involved in their regulatory mechanisms of genome transcription and protein synthesis. The exploration of this knowledge becomes a challenge especially for scientists analyzing the virions in their natural environment, due to their interactions with the complex and diverse types of biological systems, which directly influence the regulation of the infective cycles. Thus, knowledge about viral genes, their function, organization, and modulation, beyond the comprehension of the viral components as parts of complex systems, consist of the main hurdles for the controlled and predictable handling and use of these particles. For this, the technology of Synthetic Synthesis of viral genomes is distinguished from traditional genetic engineering through the use of modularity and standardization to construct proof-of-principle systems and allow generalized circuits designs to be applied to different scenarios. This new technology is made possible thanks to advances in many areas of science, from the use of restriction enzymes until the development of techniques of genomic synthesis and sequencing, like the 454 Roche, Illumina, and SOLiD systems. This technology is becoming increasingly a multidisciplinary tool used in the investigations about these complex systems, as well in the engineering of new particles for the optimization of diverse viral functions and

* Corresponding Author's Email: depaula@ufv.br (Tel.: +55 31 3899-2589; Fax: +55 31 3899-2549).

alterations in their infectivity and affinity, or even in the development of completely new organisms and features without the need for a template. Nevertheless, advances in this technology are still limited by the lack of dynamic techniques for monitoring biological systems and efficient and standardized circuits. Here, we summarize the major characteristics of viral genomes, their organization and gene modulation, and highlight the main aspects of Synthetic Biology applied to viral genomes, as its main techniques and applications.

INTRODUCTION

Viruses are infective particles essential for the balance of ecosystems and are directly related to the evolutionary processes throughout the history of life on Earth. Therefore, the study of viral genomes attracts the interest from various areas of science, since the elucidation of the genomic sequences, its function, organization and modulation are sources of important discoveries about the evolution of species and may help to explain the biological behavior of the more complex living organisms. In addition, knowledge and handling of these genomes along with techniques of molecular biology have allowed these particles be used as vectors or gene recombination agents, as antimicrobials agents helping to control microbial contamination in industries and hospitals, or still used in the production of vaccines for the control of various pathogens, for the detection of micro-organisms, among others. For example, Lu and Collins demonstrate that an engineered bacteriophage was able to enhance the killing of antibiotic-resistant and biofilm-former bacteria, and act as a strong adjuvant for other bactericidal antibiotics (T K Lu & J J Collins 2007).

Viral genomes are extremely variable in size and structure of genes, but patterns can be observed along the evolutionary studies, especially among closely related viruses. The techniques for sequencing these genomes are already a practice and economic reality, allowing the study of genes, their function and modulation for handling these genomes can be accomplished in a rational and predictable way. Accordingly, the chemical synthesis of genes has been gradually made possible, but yet is not a reality in terms of genomes, and biological techniques are still necessary for the union of synthetically synthesized chains of nucleotides. The scientific community's interest in this issue is so meaningful that in 2012 was created a journal exclusive for this theme in which researchers could report the successful cases in the study and manipulation of viral genomes.

In this chapter we discuss about the main elements of the viral genomes, as well as modern techniques for sequencing and manipulating these genomes, concluding with an approach about the concepts of synthetic biology and examples of success in the synthesis of synthetic viral genomes reported throughout the world.

ESSENTIAL ELEMENTS OF VIRAL GENOMES

Essential Genes

In the cells and viral particles, the constancy of genetic networks form biological networks, which act synergistically to perform various metabolic functions (Figure 1) (Riccione et al.

2012). Thus, a major goal of Cellular Biology is to understand how these complex networks culminate in the observed cellular behaviors. One of the most used method to assess a gene function or genetic network is based on the isolation of mutants unable to complete a particular pathway. The mutations in the viral genomes can interfere with a range of properties of these particles such as infectivity, the type of lysis plaque formed by them, their host range, or the physicochemical properties of the particles. However, the production of lethal mutations generates unviable virions, thus making them undetectable. Strategies based on synthetic biology to study genetic circuits or viral motifs gene can be a solution to this problem. Thus, to obtain more complex circuitry is necessary to know which genes confer essential characteristics to each organism, which has led to a large number of studies devoted to this subject (Little 2010).

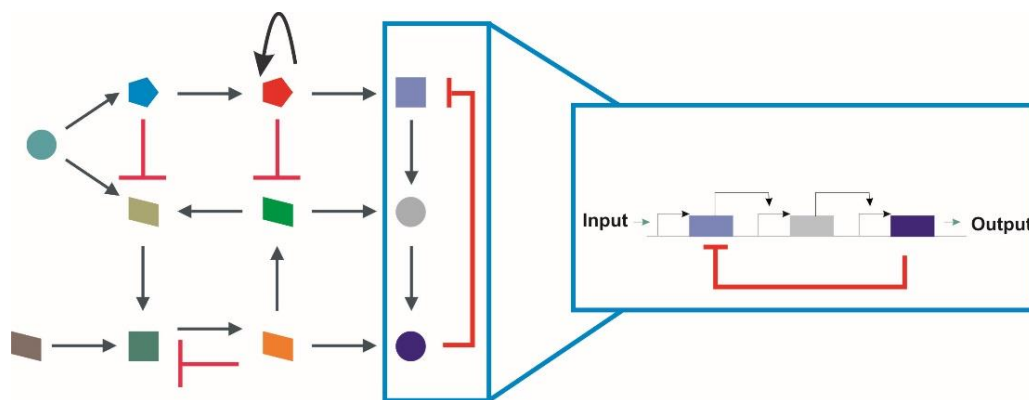


Figure 1. The more complex circuits have various integrated routes. Synthetic biology allows reconstructing parts of these circuits with different purposes, such as the study of these genes, the production of metabolites and control of cell populations (adapted from Riccione et al. 2012).

Viruses have high genetic diversity, ranging from extremely simple genomes, such as that of porcine circovirus, which has a circular genome of only 1.7 nucleotides and three Open Reading Frames (ORFs) (Figure 2), till extremely large genomes, such that of mimivirus, which genome codes for more than 900 ORFs (Abrescia et al. 2012). There is not a universally conserved gene among all viruses. However, some genes that encode proteins required for replication and particle formation are present in all members in a group (Dolja & Koonin 2011). Although there is similarity between sequences of the viral genomes analyzed until now, surprisingly the vast majority of sequences of viruses present in ocean has low similarity to known viral sequences, finding greater similarity to bacterial genes, some involved in the main processes of the cellular metabolism. Two not mutually exclusive hypotheses could explain this issue, one based in the contamination of samples with bacterial DNA, raising the possibility of fails in that protocols of metagenomic analysis, and the other in the fact that our current knowledge on viral genomes is not representative of the current virome which makes oceans the main viromes on the planet (Kristensen et al. 2010).

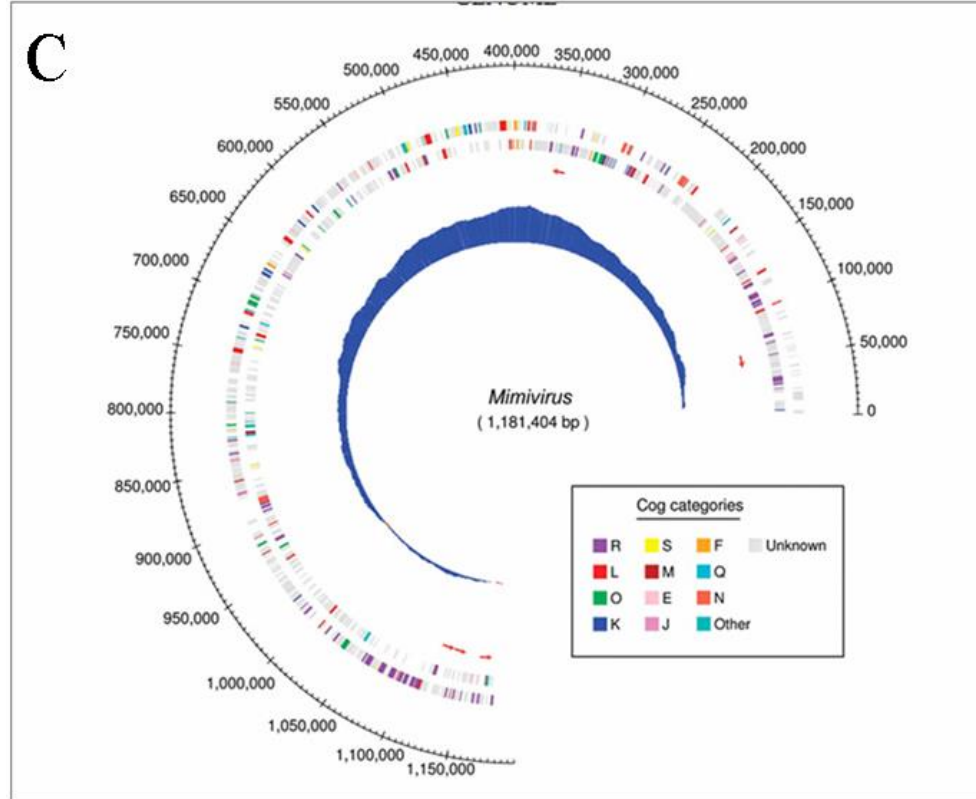


Figure 2. Genomes with different degrees of complexity. A) PCV genome, the simplest viral genome known, composed of essential genes for replication (ORF1), virion assembly (ORF2) and host apoptosis (ORF3). B) UFV-P2 genome, with intermediate complexity and at least two transcription modules of replication and virion assembly and release. C) mimivirus genome presenting high complexity, possessing 1.2 Mb and more than 900 genes (Raoult et al. 2004).

Due to the great diversity of viral genomes and the distinct genetic needs between different viruses, there are no genes that are essential to all of them. However, two proteins are considered the minimum requirements for the existence of a viral particle - the viral capsid protein and - the viral polymerase. Although the genes encoding these proteins are highly variable among different groups, its existence is almost unanimous among known viral genomes. The other genes present in the viral genomes are components of transcription modules that vary between different virus groups and depend on the particle structure and their mechanism of infection/replication. Next are cited groups of some viruses and their basic replication machinery.

1. Bacteriophages as well as plasmids, can be considered genetic elements, motile or not, present in prokaryotic cells. Once internalized, the genomes of these viruses are called prophages or replicons, and replicate intracellularly by the "replicon model" using key regulatory elements that interact with an origin of replication (Ori). The replication of linear viral genomes composed of dsDNA seems to be simple, despite its unidirectional mechanism of initiation, which leads to the loss of a portion of the genome in each replication cycle. Evolutionary studies showed that phages found at least four different mechanisms to address this issue: 1 - The *Bacillus subtilis* phage Φ 29 proteins uses specific proteins as portable primers, which remain covalently attached to the ends of the viral genome; 2 - The *Escherichia coli* phage T7 possess direct repeats at the terminal ends of its genome, which are regenerated by the processing of its concatamers by enzymes called terminases during the packing of monomeric genomes, 3 - The *Escherichia coli* phage N15 regenerates the ends of its linear double-stranded DNA by the action of the specialized enzyme protelomerase, similarly to what happens for eukaryotic genomes by the enzyme telomerase, 4 - linear genomes of some phages are circularized after its injection into the host, while others are integrated into the host chromosome in the form of prophage.
2. The porcine circovirus (PCV) are small, icosahedral viruses with approximately 17 nm in diameter and composed of a circular ssDNA genome. Replication of the genome of PCV-1 is made by rolling circle (RCR) mechanism, which is widely used by phages, plasmids, animal viruses and phytoviruses (Cheung 2012). The PCV have two opposing ORFs from the Ori region. The ORF1 is transcribed and processed, giving rise to the Rep proteins (Rep and Rep[']) related to viral replication. ORF2 is present in the complementary strand and is responsible for the synthesis of the viral capsid protein (Cheung 2012). The ORF3 was recently characterized as a non-structural protein, not essential for replication in cell cultures, and with a major role in the induction of apoptosis by activating the initiation of the via of caspase 8 and caspase 3 in the host (Liu et al. 2005).
3. Viruses of the *Flaviviridae* family have linear genomes consisting of single stranded RNA (ssRNA) with positive polarity, what means that it can be transcribed immediately after its entry into the host cell. These genomes are transcribed as a single polyprotein, which is cleaved generating the viral structural and non-structural proteins. Among these is the NS5 protein, a RNA-dependent RNA polymerase (RdRp), which is essential for the phage genome replication, since natural animal cells do not possess this enzyme. A second element essential to the replication of these viruses is the sequences for cyclization, present in the terminal ends of their genomes.

Recent studies show that the RdRp binds to the clamp formed after genome circularization and initiates its replication by synthesizing the negative strand RNA, which is the replicative intermediate of flaviviruses (Malet et al. 2008).

4. Until the discovery of mimiviruses, the concept of viruses was based on the filtration of solutions in membranes with 500 nm pores. However, mimivirus have a diameter of about 750 nm, beyond fibrils, which made this definition improper. Although little is known about the different stages of mimiviruses infection, it is known that their 981 genes characterize them as the first known viruses with complex genomes similar to other intracellular bacterial parasites. In addition, these viruses were the first in which components of the system of protein synthesis were observed, such as amino-acyl tRNA synthetases (Claverie & Abergel 2010). In addition to size and complexity of its genome, the mimivirus are characterized by a high gene density (90.5%), with something like 1262 ORFs, and of these 298 have function assigned to several central functions of cellular metabolism, ranging from the amino acid metabolism and transport, translation and post-translational modifications.

Organization and Modulation of Viral Genomes

The evolution of the viral particles led to the mobile interchangeable elements, known as modules, which carry specific biological functions. Thereby, viruses found ways to adapt their genomes to different niches through the combination of different modules. A detailed comparison of two genomes from phages belonging to different niches may shows similarities in genetic organization, although containing alternating regions of high and low sequence similarity (Botstein 1980). The high degree of conservation in the ordering of functional genes into transcriptional modules is evident, allowing the inference of functional proteins by comparing the gene sequences using bioinformatics (Veesler & Cambillau 2011). Furthermore, the homology in the organization of modules in phages from different families enables the establishment of hybrids between them.

The genetic modules have two basic characteristics: the efficient conducting of their biological functions and their ability to permute among genomes. Both features define the frequency at which each module is inserted (or lost) in a genome, and this decision is mainly related to the functional compatibility of a module with a variety of combinations with the other essential modules present in the genome of an organism. Thus, the presence/absence of specific viral modules in the genomes may be also indicative of the phylogenetic distance of viral hosts. Lawrence and coworkers proposed that the prophages and phage genomes could be described as an assembly of several genetic modules, which tend to remain associated independently of the recombinant nature of their genomes (Lawrence et al. 2002). Therefore, these modules can be classified according to the functions of their inner genes, as occurs in the modules of adsorption, replication, regulation, recombination, capsid, basal plate, proteins forming contractile tails, proteins forming long non-contractile tails, tail fibers, lysis, integration, excision and pathogenesis (Toussaint et al. 2007; Lima-Mendez et al. 2011).

The classification of viruses can be performed by the analysis of evolutionarily conserved modules (ECM). Some ECMs appear to be specific to certain types of phages, host groups, or type of infective cycle. For example, the ECM17 is composed of two sub-modules, the first of integration/regulation, which are real markers for integrative phages, and the replicative module

found in Gram-positive-infecting phages (Pellegrini et al. 1999; Lima-Mendez et al. 2011). Recently, a software was developed based on this type of classification and was named Phage Classification Tool Set - PHACTS, which uses genomic analysis to predict the infective cycle of different phages (McNair et al. 2012).

METHODOLOGIES FOR GENOMIC MANIPULATION

The term Synthetic Biology was first described in 1910, by Leduc, S. (Stéphane Leduc 1910). This technology has enabled the development of cells and biological devices with features molded according to a number of different interests. This has led to a demand for more dynamic, simple, and cost effective new technologies of DNA synthesis and sequencing (T K Lu & J J Collins 2007).

Genomic Sequencing

The first regulatory circuits were discovered over 40 years and consisted of a feedback inhibitors of amino acidic biosynthetic pathways (Westerhoff & Palsson 2004). The discovery, production and commercial viability of restriction enzymes, and the standardization of techniques as molecular cloning constituted major advances in the 70s, allowing the emergence of technologies such as Genetic Engineering and Biotechnology. Already in the 1980s some of the fundamental experimental approaches of Molecular Biology were developed and improved, and in the mid 90's the first automated DNA sequencers were glimpsed, leading the genetic sequencing to the genomic scale. The automation, miniaturization and multiplexing of multiple trials led to the generation of additional types of "omic" data, such as metabolomics, proteomics, and genomics (Westerhoff & Palsson 2004; Hunkapiller et al. 1991; Rowen et al. 1997; Uetz et al. 2000).

The first methodology for sequencing DNA was the chemical method of *Sanger*, in which labeled modified nucleotides are used for reading the template DNA during amplification. The Human Genome Project was performed by a variation of this method, the *shotgun sequencing*, in which the DNA is mechanically or enzymatically broken and the fragments generated are cloned and individually sequenced. After obtaining individual sequences, these are then overlapped for the assembly of the entire genome. The new sequencing methodologies known as next generation sequencing (NGS) are based on this principle, performing random reads throughout the genome and assembling the fragments generated by overlapping (J. Zhang et al. 2011). However, the number of bases that can be continuously read by NGS methods is very small when compared to Sanger, what is the major limitation of these methods since they generate a very large amount of data, making their processing difficult and time-consuming. Some of these new platforms are described below:

1. Roche GS-FLX 454 – It was as the first commercial sequencing platform, introduced in 2004, and has as main advantage the generation of the larger fragments among the NGS platforms. The key process is based on mixing the reactants for amplification in an emulsion. Inside the micelles, the amplification is performed in DNA-binding

beads. In this platform, the DNA is read through pyrosequencing, in which light is emitted and recorded when each nucleotide is incorporated by the polymerase during the fragment amplification.

2. Illumina - It was the second platform to be commercialized and is currently the most used due to its lower costs. This platform is also based on the sequencing by synthesis, in which DNA fragments are attached to fixed supports via adapters (oligo-primed). These oligo-adapters are also used to amplify the free end of the DNA fragment, thus generating clusters of identical fragments. The amplifications using terminators nucleotides labeled generate signals that are interpreted and give rise to the sequence of each fragment.
3. SOLiD - This platform is based on sequencing by ligation, in which the templates are attached to microspheres in emulsions and combined with universal sequencing primer, the enzyme ligase, and fluorescent probes. Sequencing occurs by hybridization of such fluorescent probes with the target fragment through several steps using different combinations of universal primers.

In the last two years new genetic sequencing platforms were presented and called third generation sequencing. One of them is based on changes in electric current generated by the passage of DNA in nanopores, with an ability to read million bases per hour. A second method is based on the variation of electrical conductance of a solution, which occurs due to the polymerase addition of nucleotides to the chain of nucleic acids, presenting processivity about 22 nucleotides per second (Y. S. Chen et al. 2013; Radford et al. 2012; Eisenstein 2012).

Genetic Manipulations

The mutations in the phage genomes may be obtained from the use of physical agents such as ultraviolet light, chemical mutagens, or by recombinant DNA technology. These techniques allow changing specifically or randomly the sequences of nucleotides that compose the viral genomes, thus generating particles with different characteristics in relation to the wild ones. For example, the technique of *DNA shuffling* allows the random recombination of fragments between homologous genes, thereby generating a series of particles with different versions of a particular gene. On the other hand, *error prone PCR* is a technique of inducing errors during amplification of a specific gene, generating directed punctual mutations. Homologous recombination has also been used for deleting genes or obtaining recombinant phages. The technology known as *recombineering* allows the introduction of *in vivo* changes, such as knock outs, deletions, insertions, or point mutations, in a genome. Yu and coworkers reported the development of a transformation system of *Escherichia coli* genome by inserting the recombination system of the phage λ (Exo and Beta) in the recombination system *RecBCD* (Marinelli et al. 2012; Yu et al. 2000).

The chemical synthesis of nucleotide sequences led to a race of biotechnology companies for more efficient ways to synthesize genetic material with increasing amounts of nucleotides, making it possible to sell affordable gene sequences. Currently, it is possible to a researcher to purchase gene sequences as large as 10,000 base pairs with a delivery time of about a month! The high-throughput DNA synthesis technology provides to researchers a new and important tool. However, the big question still lies in correctly performing the biological question to be

answered by this tool, which can be facilitated by detailed study and mathematical modeling of the known genetic circuits. The chemical modification of nucleotides and the generation and study of isolated circuits offer a way to study independently the influence of circuits in natural pathways without disturbing the other components of the metabolism.

Some of these circuits have been used in useful tasks, such as control of cell populations, decision making for biosensors, genetic timer for fermentation processes, and image processing. More recently these circuits have been used to solve medical and industrial solutions such as for the development of bacteria capable of invading cancer cells, for the dispersion of bacterial biofilms by engineered phages, as well as for the production of precursors of anti-malarial drugs for the generation of synthetic microbial routes. However, in such cases, engineered organisms have a single modified gene circuit, which does not include the potential offered by synthetic biology, so showing a discrepancy between the simplicity of these genetic circuits and the promise of assembly of these circuits in complex genetic networks (T K Lu et al. 2009). Although there are no limitations on the number of new circuits to be built, the number of interoperable and well-characterized parts hampers the development of more complex biological systems, thus creating a constant need to expand the set of available tools. This limits our ability to build truly modular circuits and highlights the need for an accurate characterization of the interactions between the different components of these systems so that negative interactions are minimized (T K Lu et al. 2009).

SYNTHESIS OF NEW VIRAL GENOMES

The beginning of the study and synthesis of genetic circuits evolved the manipulation of the genetic material of most organisms by a series of restriction enzymes, recombination and amplification reactions, and the expression of heterologous proteins in foreign systems. While these techniques are now used routinely in laboratories around the world, the need for templates for genetic manipulation and the high complexity of the regulatory system of the host led to the need to synthesize *de novo* genomes, from which it would be possible to predict the interactions between different genetic components. Thus, the chemical synthesis of long polymer chains of nucleotides and the construction and maintenance of DNA libraries have favored the emergence of this new way of creating and manipulating entirely new biological systems.

The synthesis and manipulation of genes for their insertion or deletion in the genomes is already a strategy routinely used in different fields of science, for example, in elucidating metabolic pathways, in discovering gene function, or to genetic improvement. However, the synthesis of new genomes without templates has been done only since the last decade, and the increasing knowledge on the functions of regulatory sequences have contributed to the synthesis of larger and more complex genomes, with increasingly predictable and rational designs. An example was the synthesis *de novo* of a bacterial genome in 2010 (Gibson et al. 2010). While the synthesis of genes considered natural and proximal regulatory aspects, the genomic synthesis should consider the different regulatory elements contained in the entire genome, the interactions between them, and the products of the different genes. This is only possible from a consolidated knowledge about the viral genomes, its function and modulation.

Genetic Circuits

Synthetic biology integrates the techniques of molecular biology to the principles of engineering, computer sciences, and mathematical models. From this integration, it is possible to design and construct genetic circuits that enable living cells to perform new functions. The term genetic circuitry has been widely used to refer to well-defined sequences of genetic material that, together, perform specific functions when transcribed. This concept is directly related to the rational construction of genomes with predefined characteristics. A large number of genetic circuits have been developed, most of them switches, oscillators, digital logic gates, filters, modular and interoperable memory devices, counters, sensors and protein scaffolds (Timothy K Lu 2010). By using these circuits is possible to predict, at least partially, the features and biological modulations of synthetic viral particles, which can be manipulated to present characteristics desirable in relation to the functions for which they are used. In the future, researchers envision that, by analogy to electronic circuits, it will be possible to generate viral particles programmable from the generic combination of genetic circuits which, when combined, generate an predictable phenotype (Ferber 2004). For this, it is necessary to develop well-defined stages of construction, monitoring and debugging and setting of these circuits.

Monitoring the Expression of Genetic Elements

Practices methodologies of large-scale monitoring of the expression of circuits are essential for the development of efficient networks and for obtaining reliable and reproducible results. These techniques are mainly based on the detection of reporter elements - molecules whose concentration is proportional to the expression of the target gene, as mRNAs and proteins. there are many methods based on this concept, but the optimal methodology for monitoring gene circuits, although it is dependent on the circuit involved, may generate strong signals with low noise, present specificity, low cost, non-invasiveness, multiplexability, and, if possible, generate dynamic real-time signals (Timothy K Lu 2010). It is possible to monitor a circuit for analyzing the total protein content of an organism without the need for a reporter element. However, this method is nonspecific and often non- accepted for the quantitative determination of expression, and thus does not significantly contribute to the understanding of these circuits. Thus, some macromolecules are frequently used as reference detectors agents. Initial studies relied on the determination of the promoter activity by the simple fusion of the promoter plus the gene of interest with a particular reporter gene. These systems are very effective and used in a number of systems, although the scientists still search for alternatives that are less invasive, cheaper and applicable to a larger number of systems.

Colorimetric Assays

The colorimetric assays for monitoring protein expression consist of simple but inaccurate techniques. These assays are based mainly on the co-expression of the gene of interest with an auxiliary gene encoding a colored product, or an enzyme capable of generating products detectable in visible light and therefore measurable. For example, one of the first systems used for reporter gene expression was based on colorimetric assays involving the enzyme β -galactosidase (β -gal) encoded by the gene *lacZ* in the lac operon in *Escherichia coli*. Cleavage

of galactosidic substrates by this enzyme generates colored products whose concentration can be measured in a spectrophotometer and is proportional to the amount of enzyme. Thus, the expression of the genes of interest under the promoter of the lac operon can be tracked by monitoring the expression of the β -gal gene. Advantages of this method include its simplicity and low cost, in addition to the vast knowledge about the parameters that could influence the results. Among the main disadvantages are sharing of a promoter, low precision and low reproducibility.

Green Fluorescent Protein (GFP)

The production and detection of fluorescent proteins is already a practical and well characterized technology for sensing and monitoring protein expression of genetic circuits. Furthermore, this technology is not specific and can be applied to a variety of circuits. The cloning of the GFP gene from jellyfish (*Aequorea victoria*) in heterologous systems allows monitoring of its expression via the emission of green fluorescence by the protein after excitation by blue or UV light. For this, scientists couple the gene coding for this protein to the gene of interest using an expression systems and, thus, are able to quantify it and even to locate it in various systems. However, because it is a protein, intrinsic and extrinsic factors involved in its stability, conformation and interactions with various components of the cell may interfere with the accuracy of the methodology. The main advantage of using this protein is the possibility of in vivo and in real time monitoring, without the need for extraction steps or destruction of the particle.

Luminescence

The systems for monitoring protein expression via luminescence emission are based on exothermic reactions that emit luminescence. The main example is undoubtedly those that uses the enzyme luciferase expressed by the *luc* gene, cloned from the genome of firefly (*Photinus pyralis*). This enzyme catalyzes the oxidation of luciferin in the presence of ATP, Mg^{2+} and O_2 , generating a light signal that is proportional to the amount of the enzyme. Systems based on luminescence present high sensitivity, speed and are relatively inexpensive. However, in the same way using GFPs, the results depend on the efficiency of the maintenance of the enzymatic activity by stabilizing the system conditions. Specifically, these system also present a great sensitivity to variations in ATP concentrations, which can make their use unfeasible in some cases.

Aptamers

Aptamers are fragments of macromolecules, specially nucleic acids, that are capable of specifically recognizing and detecting proteins. However, their in vivo use has not been widely applied to the monitoring of genetic circuits because the algorithms that could safely predict the binding affinity between aptamers and the proteins of interest would be extremely specific and therefore not viable, since they are not applied to different cases. However, alternatives that use these molecules also can be developed for detection of proteins in several different systems.

Detection of Cellular mRNAs

The amplification of the mRNA for quantitation of the expression of certain genes constitutes a simple and well-characterized technique for monitoring the expression of genes contained in genetic circuits and governed by different modulators. However, besides not being an accurate picture of the protein content of the host, thus excluding the translational stages, techniques for detecting nucleic acids require the extraction of this material and consequently the destruction of the cells analyzed. This is therefore an invasive technique and impractical for real time analyses. The detection of RNAs by fluorescent proteic probes can be an alternative to these issues, but requires that the RNA of interest are modified so that it can specifically bind to the probe.

Debugging

The number of genetic circuits that have presented unexpected results is still very high. This is due to our still limited knowledge about the different parameters involving biological circuits for the synthesis of viral genomes, and the interactions between the components of these circuits and the machinery of the host. In addition, living organisms do not respond linearly according to its components due to the co-operability between molecules, which makes even more challenging to develop predictive mathematical models for these systems. The debugging of genetic circuits involves the detailed characterization of these systems, all available nodes, modifications by trial and error to characterize the interactions between the molecules involved, and specific modeling.

EXAMPLES OF PROJECTS ON SYNTHETIC GENOMES

1. *Bacterial gene circuits*: In 2004, the group of researchers leaded by Hideki Kobayashi and colleagues engineered genetic circuits coupled to the cellular machinery of an *E. coli* strain. For this, they inserted plasmids containing a gene essential for biofilm formation under the control of a repressor regulated by RecA protein, which is responsible for mediating the SOS response. The circuit designed was capable of responding to biological signals (damage to genomic DNA) in a predictable and programmable way, so that cells formed biofilm when stimulated. With this, the researchers showed that the genetic circuits can be used to program bacterial cells, through coupling these circuits to the interactions networks of the cell itself (Kobayashi et al. 2004).
2. *Synthetic viral genomes*: Aiming the degradation of bacterial biofilms, American researchers engineered the genome of bacteriophage T7, an *E. coli*-specific phage, in which they inserted the gene for the enzyme dispersin B, which is able to reduce biofilms of different species of bacteria. The gene was inserted under the control of the T7phi10 promoter, so that the enzyme was expressed during infection of the phage and thus released after bacterial lysis. The engineered phage was about two orders of magnitude more efficient in degrading the biofilm matrix as compared to the wild phage (T K Lu & J J Collins 2007).

3. *Synthetic viral genomes*: using tools of synthetic biology, Paul R. Jaschke and colleagues recently published a study aiming establish methods enabling the synthesis, assembly, and recovery of a bacteriophage genome via yeast. They used the bacteriophage øX174 given its small circular genome with 5386 nucleotides that codes 11 gene products with overlapping open reading frames. The researchers questioned whether this overlap was essential or might be eliminated. For this, they synthesized a 6302 nucleotide synthetic genome separating all ORFs and their regions of translational control. The synthesis of the synthetic genome was performed by a combination of two different strategies: chemical synthesis and PCR amplification. For the synthetic genome was reduced to the exact size of the wild genome, the F gene, which encodes the coat protein, was truncated. The researchers were able to recover infective virions containing the truncated and decompressed genome from replication cycles in *E. coli* expressing the F gene under an inducible promoter. These results indicate that the overlap of genes in the genome of this phage is probably just a tool for genome compaction to facilitate its packaging into the viral particles, but is not necessary for the virions viability (Jaschke et al. 2012).

PROSPECTS

Although gene manipulation and processing of viral genomes are already routinely used in laboratories for the modification of viral characteristics, the development of entirely new synthetic genomes without the use of templates and considering the genetic circuits as an unit for assembly of genomes, comprises a new theme and is still relatively obscure. The hurdles to predict and forecast the synergistic behavior of these circuits lead to security issues that must be answered before this method becomes popular. This only will be possible from the knowledge on the interactions between the different components of this circuit, and the monitoring and debugging of the already created circuits in order to increase the existing mathematical models and algorithms to optimize the predictability of these systems. Even so, the creation of entirely new viral genomes is considered one of the most promising technologies for the development of new methods of genetic transfer and recombination, for the generation of new more specific and efficient antimicrobials, for the planning of new vaccines and specific antiviral drugs with lower side effects, and also for the detection of microorganisms in various industrial and hospital environments, where strict control of the microbial population is essential.

CONFLICT OF INTEREST

None of the authors of this paper has a financial or personal relationship with other people or organizations that could inappropriately influence or bias the content of this chapter.

ACKNOWLEDGMENT

This work was supported by Brazilian funding agencies CNPq and FAPEMIG.

REFERENCES

- Abrescia, N. G. et al. (2012). Structure unifies the viral universe. *Annual review of biochemistry*, 81, 795–822. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/22482909>.
- Botstein, D. (1980). A theory of modular evolution for bacteriophages. *Annals of the New York Academy of Sciences*, 354, 484–490. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/6452848>.
- Chen, Y. S. et al. (2013). DNA sequencing using electrical conductance measurements of a DNA polymerase. *Nature nanotechnology*. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/23644569>.
- Cheung, A. K. (2012). Porcine circovirus: transcription and DNA replication. *Virus research*, 164(1-2), 46–53. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/22036834>.
- Claverie, J. M. & Abergel, C. (2010). Mimivirus: the emerging paradox of quasi-autonomous viruses. *Trends in genetics : TIG*, 26(10), 431–437. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/20696492>.
- Dolja, V. V & Koonin, E. V. (2011). Common origins and host-dependent diversity of plant and animal viromes. *Current opinion in virology*, 1(5), 322–331. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/22408703>.
- Eisenstein, M. (2012). Oxford Nanopore announcement sets sequencing sector abuzz. *Nature biotechnology*, 30(4), 295–296. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/22491260>.
- Ferber, D. (2004). Synthetic biology. Microbes made to order. *Science* (New York, N.Y.), 303(5655), 158–61. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/14715986> [Accessed March 11, 2013].
- Gibson, D. G. et al. (2010). Creation of a bacterial cell controlled by a chemically synthesized genome. *Science* (New York, N.Y.), 329(5987), 52–6. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/20488990> [Accessed February 27, 2013].
- Hunkapiller, T. et al. (1991). Large-scale and automated DNA sequence determination. *Science*, 254(5028), 59–67. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/1925562>.
- Jaschke, P. R. et al. (2012). A fully decompressed synthetic bacteriophage øX174 genome assembled and archived in yeast. *Virology*, 434(2), 278–84. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/23079106> [Accessed May 15, 2013].
- Kobayashi, H. et al. (2004). Programmable cells: interfacing natural and engineered gene networks. *Proceedings of the National Academy of Sciences of the United States of America*, 101(22), 8414–9. Available at: <http://www.pubmedcentral.nih.gov/article-render.fcgi?artid=420408&tool=pmcentrez&rendertype=abstract> [Accessed March 18, 2013].

- Kristensen, D. M. et al. (2010). New dimensions of the virus world discovered through metagenomics. *Trends in microbiology*, 18(1), 11–19. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/19942437>.
- Lawrence, J. G., Hatfull, G. F. & Hendrix, R. W. (2002). Imbroglis of viral taxonomy: genetic exchange and failings of phenetic approaches. *Journal of bacteriology*, 184(17), 4891–4905. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/12169615>.
- Lima-Mendez, G., Toussaint, A. & Leplae, R. (2011). A modular view of the bacteriophage genomic space: identification of host and lifestyle marker modules. *Research in microbiology*, 162(8), 737–746. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/21767638>.
- Little, J. W. (2010). Evolution of complex gene regulatory circuits by addition of refinements. *Current biology: CB*, 20(17), R724–34. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/20833317>.
- Liu, J., Chen, I. & Kwang, J. (2005). Characterization of a previously unidentified viral protein in porcine circovirus type 2-infected cells and its role in virus-induced apoptosis. *Journal of virology*, 79(13), 8262–74. Available at: <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=1143760&tool=pmcentrez&rendertype=abstract> [Accessed May 17, 2013].
- Lu, T. K. & Collins, J. J. (2007). Dispersing biofilms with engineered enzymatic bacteriophage. *Proceedings of the National Academy of Sciences of the United States of America*, 104(27), 11197–11202. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/17592147>.
- Lu, T. K., Khalil, A. S. & Collins, J. J. (2009). Next-generation synthetic gene networks. *Nature biotechnology*, 27(12), 1139–1150. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/20010597>.
- Lu, Timothy K. (2010). Engineering scalable biological systems. *Bioengineered bugs*, 1(6), 378–84. Available at: <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3056087&tool=pmcentrez&rendertype=abstract> [Accessed April 3, 2013].
- Malet, H. et al. (2008). The flavivirus polymerase as a target for drug discovery. *Antiviral research*, 80(1), 23–35. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/18611413>.
- Marinelli, L. J., Hatfull, G. F. & Piuri, M. (2012). Recombineering: A powerful tool for modification of bacteriophage genomes. *Bacteriophage*, 2(1), 5–14. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/22666652>.
- McNair, K., Bailey, B. A. & Edwards, R. A. (2012). PHACTS, a computational approach to classifying the lifestyle of phages. *Bioinformatics*, 28(5), 614–618. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/22238260>.
- Pellegrini, M. et al. (1999). Assigning protein functions by comparative genome analysis: protein phylogenetic profiles. *Proceedings of the National Academy of Sciences of the United States of America*, 96(8), 4285–4288. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/10200254>.
- Raoult, D. et al. (2004). The 1.2-Mb Genome Sequence of Mimivirus. *Science*, 306, 1344–1350. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/15486256> [Accessed May 24, 2013].
- Radford, A. D. et al. (2012). Application of next-generation sequencing technologies in virology. *The Journal of general virology*, 93(Pt 9), 1853–1868. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/22647373>.

- Riccione, K. A. et al. (2012). A synthetic biology approach to understanding cellular information processing. *ACS synthetic biology*, 1(9), 389–402. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/23411668> [Accessed March 7, 2013].
- Rowen, L., Mahairas, G. & Hood, L. (1997). Sequencing the human genome. *Science*, 278(5338), 605–607. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/9381170>.
- Stéphane Leduc (1910). *Théorie Physico-Chimique de la Vie et Générations Spontanées*, [Physico-Chemical Theory of Life and Spontaneous Generations]. 1st ed. A. Poinat Editer, ed., Paris.
- Toussaint, A., Lima-Mendez, G. & Leplae, R. (2007). PhiGO, a phage ontology associated with the ACLAME database. *Research in microbiology*, 158(7), 567–571. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/17614261>.
- Uetz, P. et al. (2000). A comprehensive analysis of protein-protein interactions in *Saccharomyces cerevisiae*. *Nature*, 403(6770), 623–627. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/10688190>.
- Veesler, D. & Cambillau, C. (2011). A common evolutionary origin for tailed-bacteriophage functional modules and bacterial machineries. *Microbiology and molecular biology reviews: MMBR*, 75(3), 423–33, first page of table of contents. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/21885679>.
- Westerhoff, H. V. & Palsson, B. O. (2004). The evolution of molecular biology into systems biology. *Nature biotechnology*, 22(10), 1249–1252. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/15470464>.
- Yu, D. et al. (2000). An efficient recombination system for chromosome engineering in *Escherichia coli*. *Proceedings of the National Academy of Sciences of the United States of America*, 97(11), 5978–5983. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/10811905>.
- Zhang, J. et al. (2011). The impact of next-generation sequencing on genomics. *Journal of genetics and genomics = Yi chuan xue bao*, 38(3), 95–109. Available at: <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3076108&tool=pmcentrez&rendertype=abstract> [Accessed February 28, 2013].

Chapter 100

NOVEL BIOINFORMATICS METHOD TO ANALYZE MORE THAN 10,000 INFLUENZA VIRUS STRAINS EASILY AT ONCE: BATCH-LEARNING SELF ORGANIZING MAP (BLSOM)

***Yuki Iwasaki^{1,2}, Toshimichi Ikemura¹, Kennosuke Wada¹,
Yoshiko Wada^{1,3} and Takashi Abe^{1,4,*}***

¹Nagahama Institute of Bio-Science and Technology,

²Research Fellow of the Japan Society for the Promotion of Science,

³Shiga University of Medical Science, Shiga, Japan

⁴Niigata University, Niigata, Japan

ABSTRACT

With increase of microbial and viral genome sequence data obtained from high-throughput DNA sequencers, novel tools are needed for comprehensive analyses of the big sequences data. An unsupervised neural network algorithm, Self-Organizing Map (SOM), is an effective tool for clustering and visualizing high-dimensional complex data on a single map. We previously modified the conventional SOM for genome informatics on the basis of batch-learning SOM (BLSOM), by making the learning process and resulting map independent of the order of data input. Influenza virus is one of zoonotic viruses and shows clear host tropism. Important issues for bioinformatics studies of influenza viruses are prediction of genomic sequence changes in the near future and surveillance of potentially hazardous strains.

To characterize sequence changes of influenza virus genomes after invasion into humans from other animal hosts and to study molecular evolutionary processes of their host adaptation, we have constructed BLSOMs for oligonucleotide, codon, amino-acid, and peptide compositions in all genome sequences of influenza A and B viruses and found clear host-dependent clustering (self-organization) of the sequences.

Viruses isolated from humans and birds differ in mononucleotide composition from each other. In addition, host-dependent oligonucleotide and peptide compositions that

* Corresponding Author's Email: takaabe@ie.niigata-u.ac.jp.

cannot be explained with the host-dependent mononucleotide composition are revealed by these BLSOMs. Retrospective time-dependent directional changes of oligonucleotide compositions, which are visualized for human strains on BLSOMs, can provide predictive information about sequence changes of the newly invaded viruses from other animal sources.

Basing on this host-dependent oligonucleotide composition, we have proposed a strategy for prediction of directional changes of virus sequences and for surveillance of potentially hazardous strains when introduced into human populations from nonhuman sources. Millions of genomic sequences from infectious microbes and viruses will become available in the near future because of their medical importance, and BLSOM can characterize such big data easily and support efficient knowledge discovery.

INTRODUCTION

Due to the revolutionized advancement of decoding the genome sequences, their data have been growing into supermassive “big data” in the International DNA Data Banks, and therefore, large-scale data mining have become vital. The more the available information becomes larger, the more the rightfulness of the proposed model will become verifiable as long as the model is appropriate. At the same time, because of massive amounts of data available, analyzing only a limited part of the data should lead to an image of “the authors may have made convenient stories by extracting useful data for them”, especially when a novel discovery is publicized. It should be important to have a stance “all data belonging to a certain category have to be analyzed, even if the amount of the data is very large”. Therefore, in the post-genome era, the research that gets the picture of the whole will become increasingly important and the present chapter introduces a bioinformatics method suitable for this type of a large-scale analysis.

In more detail, we introduce the studies to unveil the species-specific characteristics in the genome sequences by using “BLSOM” developed by our group [1, 2]. This BLSOM for oligonucleotide composition has been developed for genome informatics, on the basis of SOM (Self-Organization Map) originally established by Kohonen and his colleagues [3, 4]. Because BLSOM can analyze millions of sequences simultaneously, almost all known genome sequences have been clustered and analyzable on a single map [5].

GC% has been used for a long period as a fundamental parameter for phylogenetic classification of microbial genomes, including viral genomes, but the GC% is apparently too simple a parameter to differentiate a wide variety of microbial genomes. Oligonucleotide composition, on the other hand, can be used to distinguish the species even with the same GC%, because the oligonucleotide composition varies significantly among the genomes and is called ‘genome signature’ [6]. Previously, we have constructed BLSOMs for di-, tri-, or tetranucleotide composition in all genomic fragments (e.g., 10, 50 and 100 kb) derived from a wide range of prokaryotic and eukaryotic species and found that, without giving any information of species, most of the fragment sequence have been separated (self-organized) according to species [1, 2, 5].

Epidemics of influenza viruses have been frequently repeated among various animal species including birds and swine besides human. Given that the strains derived from nonhuman sources are also capable of infecting humans, it should be impossible to eradicate influenza viruses from the planet. Influenza virus has developed into a pandemic among humans at least four times (in 1918, 1957, 1968 and 2009), causing considerable damages.

Virus requires many host factors (e.g., nucleotide, amino acid and tRNA pools and host proteins) when it grows; therefore, it inescapably depends on new host factors when causing a new infection to humans from other animal species. In such cases, the genome sequence of the virus will change to fit the new cellular environments including various host factors. If this view is correct, influenza viruses are likely to modify some part of the characteristics of their genome sequences depending on host, during epidemics in a new host. In other words, it enables us to predict, at least in part, the direction of sequence changes in the newly invaded virus, if we can properly find out the host-dependent characteristics of viral genome sequences. In this chapter we introduce examples of analysis on the host-dependent characteristics of influenza virus genome sequences, which is a key to conduct forecasting their sequence changes after changing hosts. In the previous study, we have attempted BLSOM analyses on viral genome sequences all available in that time [7]. Here, we introduce the BLSOM analysis, which includes the newly accumulated sequences and examine whether the directional changes proposed in the previous paper is actually observed in the sequences obtained after our previous paper. This type of confirmatory studies should be important to verify a novel bioinformatics method.

MATERIALS AND METHODS

Viral Genome Sequences

A total of 856,730 virus sequences analyzed in Figure 1 were obtained from the NCBI and a total of 12,370 influenza A and B virus strains analyzed in Figure 2 were obtained from the NCBI Influenza Virus Resource (<http://www.ncbi.nlm.nih.gov/genomes/FLU/>) [8].

Batch-Learning Self-Organizing Map (BLSOM)

SOM is an unsupervised neural network algorithm that implements a characteristic non-linear projection from the high-dimensional space of input data onto a two-dimensional plane [3, 4]. We modified the conventional SOM for genome informatics to make the learning process and resulting map independent of the order of data input on a basis of Batch-Learning SOM: BLSOM [1, 2, 9]. The initial weight vectors were defined by principal component analysis instead of random values. BLSOM learning for oligonucleotide composition was conducted as described previously [1,2]. BLSOM learning for synonymous codon usage and visualization of diagnostic codons for the category separation were conducted as described by Kanaya et al. [9]. BLSOM program was obtained from UNTROD, Inc. (y_wada@nagahama-i-bio.ac.jp).

Oligonucleotide composition is normalized for the sequence length and the normalized composition is used for BLSOM calculation. Mono-, di- or tri-amino acid composition in eight genes of influenza virus strains is calculated, and those of eight genes are summed up for each strain. The composition normalized with a total length is used for BLSOM calculation.

RESULT

Oligonucleotide BLSOM for All Virus Genome Sequences

To explain the clustering powers of BLSOM for a large number of genome sequences from a wide variety of viruses, we have initially constructed BLSOM for di- and trinucleotide composition in all viral genome sequences currently available (ca. 860,000 sequences), which have been classified into 67 phylogenetic families (Figure 1). Lattice points that contain sequences from one phylogenetic family are indicated in color, and those that include sequences from more than one family are indicated in black. A major portion is colored, which shows that a major portion of sequences is classified (self-organized) according to phylogenetic family; 95 and 98% of viral sequences are located in colored lattice points on di- and tri-BLSOMs, respectively. Notably, no information in regard to phylotype has been given during the BLSOM calculation. It should also be mentioned that approximately one million sequences are analyzed simultaneously on one plane.

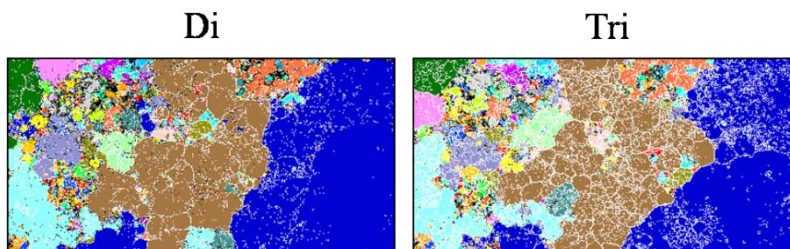


Figure 1. Oligonucleotide-BLSOMs for virus genome sequences. BLSOM has been constructed for di- or trinucleotide composition (Di or Tri) in 856,730 viral genome sequences. Lattice points that include sequences from more than one phylogenetic family are indicated in black, and those containing sequences from a single phylotype are indicated in a color representing the phylotype: Arteriviridae (■), Bunyaviridae (■), Coronaviridae (■), Flaviviridae (■), Geminiviridae (■), Hepadnaviridae (■), Herpesviridae (■), Orthomyxoviridae (■), Paramyxoviridae (■), Picornaviridae (■), Potyviridae (■), Reoviridae (■), Retroviridae (■), Rhabdoviridae (■) and others (various colors not specified here).

Oligonucleotide BLSOM for Influenza Virus Genomes

Over ten thousand genome sequences of influenza A and B virus strains are registered in the International Nucleotide Sequence Databases, even when we concern strains for which all eight segments have been sequenced. We have next constructed a tetranucleotide BLSOM for all influenza A or B strains registered in Influenza Virus Resource [8] in NCBI (Figure 2); for di- and trinucleotide BLSOMs, refer to our previous paper [7]. Because the direct target of natural selection is a virion containing a full set of eight genome segments, tetranucleotide frequencies in the eight segments are summed up for each strain and BLSOM is constructed with the summed frequency after normalization of a total nucleotide length of individual strains; this is because the length of genome sequences compiled by the Influenza Virus Resource differs slightly between strains (Figure 2). The present genome-level analysis should provide valuable information for characterizing individual strains, which may not be obtained by a

gene-level analysis, primarily conducted with sequence homology searches used in phylogenetic tree analyses.

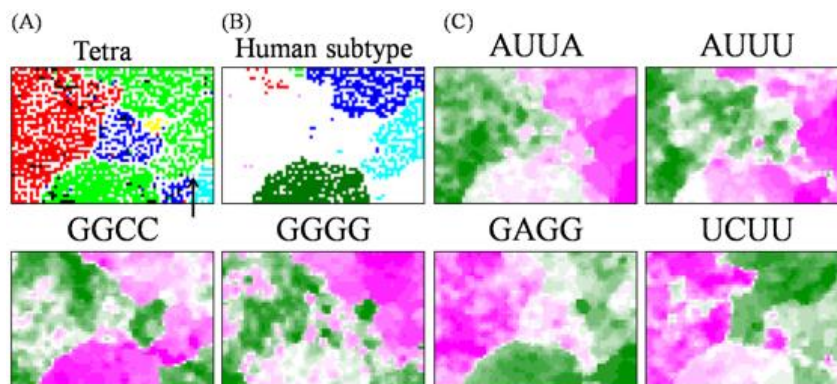


Figure 2. Tetranucleotide-BLSOM for influenza A and B virus genome sequences. (A) Tetranucleotide BLSOM (Tetra) has been constructed for 12,370 strains of influenza A and B virus strains. Lattice points containing sequences from strains isolated from more than one host are indicated in black, and those containing sequences from one host are indicated in a color: avian, red; human, green; swine, blue; equine, yellow; bat, grey; and influenza B strain, light blue. (B) Human subtype. On the tetra-BLSOM presented in (A), each human virus subtype is specified in a color: H1N1, light blue; H1N1/09, dark green; H3N2, blue; H5N1, red; H7 and H9, pink; and other human subtype, green. (C) Occurrence levels of tetranucleotides, which are diagnostic for host-dependent clustering, are indicated with different levels of two colors: pink (high), green (low), and achromatic (intermediate) as described by Abe et al. [1].

Lattice points containing virus strains isolated from one host species are indicated in a color representing the host and those containing strains isolated from more than one host are in black. Though only oligonucleotide frequencies have been given during the BLSOM calculation, viral sequences are clustered (self-organized) according to host. Viruses are inevitably dependent on many host factors for their growth (e.g., pools of nucleotides, amino acids and tRNAs) and at the same time have to escape from antiviral host mechanisms such as antibodies, cytotoxic T cells, interferons, and RNA interferences [10-13]. Host-dependent clustering of viral sequences visualized on BLSOM should reflect this host dependency in their growth.

In Figure 2B (Human Subtype), lattice points that contain human influenza A viruses of one subtype on the tetra-BLSOM (Figure 2A) are specified with one color representing the subtype. Among the 12,000 viral strains analyzed, approximately 3,000 strains correspond to the new pandemic H1N1 strain (H1N1/09) (dark green in Figure 2B), which has started its pandemics among humans around since April 2009. Although its origin is derived from avian strains, it has been infected to human via swine after genome segment reassortments. Interestingly, on BLSOM, they are located apart from seasonal human H1N1 or H3N2 strains (light blue or dark blue in Figure 2B) and surrounded by avian and swine territories (red and blue in Figure 2A and blank in Figure 2B). It can be possible that these H1N1/09 strains have not yet been best suited to growth under human cellular environments.

We next investigate human virus subtypes other than the H1N1/09 (Figure 2B). Human seasonal H1N1 and H3N2 are clearly separated from each other; light and dark blue in Figure 2B. In addition, in contrast to H1N1/09 (dark green), human H5N1 strains (red in Figure 2B and mainly black in Figure 2A) are rather scattered within the avian territory (red in Figure 2A

and blank in Figure 2B). This should show that the human H5N1 strains have jumped to humans but are not able to spread from human to human [12], and therefore, they have characteristics of avian viruses. These human H5N1 strains are more separated from the swine and human territories than H1N1/09 strains, and this difference may relate to their infection power in the human and swine populations.

In the case of influenza B strains (light blue in Figure 2A and blank in Figure 2B), which have repeatedly caused epidemics only among humans, they form a territory more distant from the avian territory than that of human seasonal strains. This characteristic of influenza B strains will be discussed later.

Diagnostic Tetranucleotides Responsible for Host-Dependent Separation

BLSOM provides a powerful ability to visualize diagnostic oligonucleotides responsible for the category-dependent clustering (the host-dependent clustering in this case). In order to study the oligonucleotides that may relate to host adaptation, we next identify diagnostic oligonucleotides for host-dependent separation on the tetranucleotide BLSOM. Six examples of diagnostic tetranucleotides for the host-dependent separation are presented in Figure 2C. A clear tendency is apparent; A- and U-rich oligonucleotides are more favored in humans than in avians (e.g., AUUA and AUUU); G- and C-rich oligonucleotides were more favored in avians than in humans (e.g., GAGG). This GC% effect was previously reported by Rabadan et al. [14]. However, GGCC and GGGG, which are composed only of G and/or C, are more favored in humans than in avians, and UCUU, a tetranucleotide rich in U, is preferred mainly in the avian territory. When we concern about interaction of viral RNAs (vRNAs and mRNAs) with various host factors (e.g., host proteins), oligonucleotide compositions, rather than the mononucleotide composition, will become important, because their interactions with host proteins primarily depend on oligonucleotide sequences in their sequences. This should also be true in considering escape processes from host antiviral mechanisms. To wrap up, oligonucleotide-BLSOM analyses of influenza viruses should provide valuable information of their host adaptation mechanisms (e.g., interactions of viral RNAs with host factors), which will become important for medical and pharmaceutical studies.

Chronological Changes Visualized for Human Viruses

In Figure 3, we next explain chronological changes of human seasonal strains using the same tetranucleotide BLSOM; only the lattice points containing human strains are shown as done in Figure 2B. In addition, the territory of human seasonal H1N1 (except for H1N1/09) and H3N2 strains in the following period is separately displayed in brown and blue; H1N1 and/or H3N2 strains isolated in 1930-1957, 1968-1974, 1975-1989, 1990-1999, 2000-2005, and 2006-2012 are displayed in brown and blue, respectively, in the respective panel. The seasonal strains isolated in 1930-1957 and 1968-1974 (i.e., the beginning of the epidemic of H1N1 and H3N2, respectively) are seen near avian and/or swine strains (red and blue in Figure 2A and blank in Figure 3), which have been estimated as their origin. Human seasonal strains are seen moving away from the avian and/or swine territory since 1930 (for H1N1) and 1968 (for H3N2), indicating the directional change of their oligonucleotide composition after

invasion into the human population. Therefore, at least from a specific point of view, genome sequence changes in the invader virus appear to be predictable, especially in an early stage of epidemics cycles. This has been found for H1N1/09 strains in our previous paper [7] even during the first one year, because of the high mutation rate and short generation time of influenza A viruses [12, 15].

To wrap up, BLSOM can visualize any categories of strains, in which experimental and medical groups will be interested, and therefore, has a power to visualize evolutionary histories of influenza viruses after invasion into new hosts.

Influenza B has caused repetitive epidemics only among humans and can be considered as it has already been well adapted to human cellular environments. In order to examine numerically whether the tetranucleotide composition of human seasonal A strains are approaching that of Influenza B with time, we have calculated Euclidean distance between the average composition in influenza B strains and that in H1N1 or H3N2 strains isolated in each year (Figure 4A). Euclidean distance shows the distance in the multidimensional scale (256 dimension for tetranucleotide composition); the distance between the strains that own an identical oligonucleotide composition is 0 and, as the difference in the oligonucleotide composition grows, the distance becomes larger.

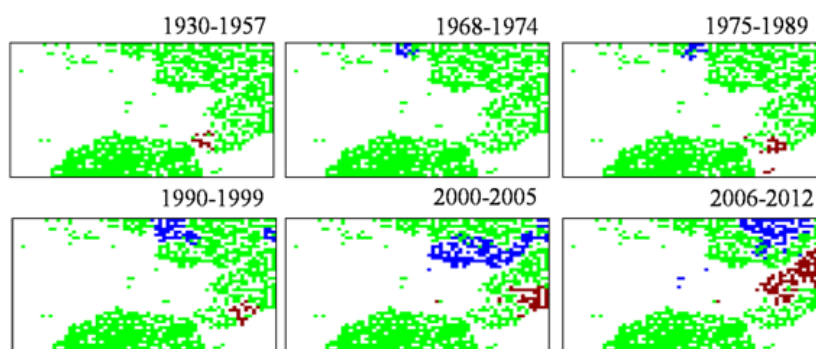


Figure 3. Chronological changes for seasonal human subtype strains. Seasonal human H1N1 and H3N2 strains that are isolated in the different period are separately marked in brown and blue, respectively.

The distance between the influenza B strain and either H1N1 (red + marks) or H3N2 (blue x marks) A strain has become smaller over time (Figure 4A), showing that the tetranucleotide composition in A strains has become analogous to that in B strain during the repeating pandemics among humans and directional changes have been accumulating in genome sequences in the seasonal A strains.

H17N10 Strains Isolated from Bat

One of the most feared influenza viruses is the strain possessing antigen types, to which most humans do not possess antibodies, and therefore, the strains having caused epidemics in nonhuman sources (e.g., birds and pigs) attract attention.

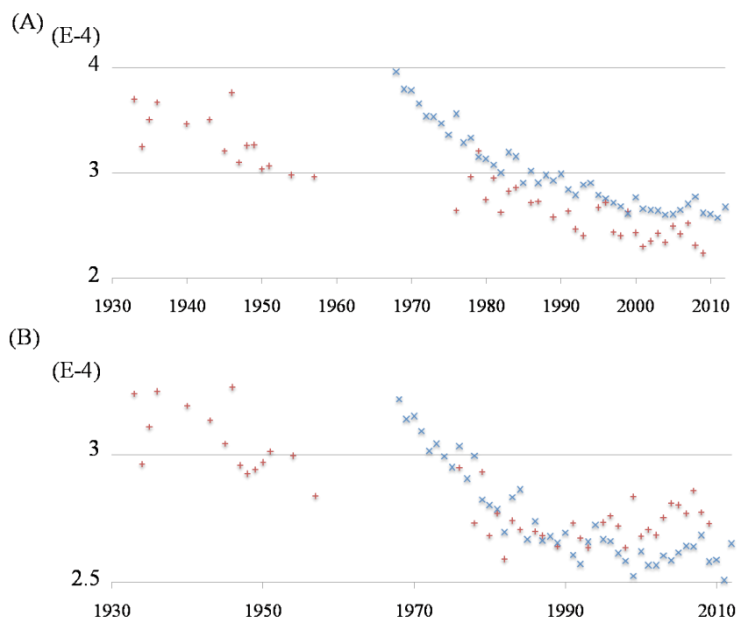


Figure 4. Euclidean distance between seasonal human A and B strains. (A) Euclidean distance of tetranucleotide composition between influenza B strain and human seasonal strain (influenza A strain). Euclidean distance between the average composition in influenza B strains and that in H1N1 (+) or H3N2 (x) strains isolated in each year is plotted. (B) Euclidean distance of tetranucleotide composition between the bat and human seasonal strains. Euclidean distance between the average composition in bat strains and that in H1N1 (+) or H3N2 (x) strains isolated in each year is plotted.

Influenza A viruses are categorized by the combinations of antigen types: 16 and 9 types of HA and NA, respectively. As a natural host, birds are capable of transmitting all of the $16 \times 9 = 144$ types of influenza A strains. During recent years, new H17N10 strains, which greatly vary from known influenza A strain, have been isolated from the bats in Republic of Guatemala and estimated as being separated from birds far into the past [16]. As previously mentioned, BLSOM have a power to focus on and visualize a specific category of strains with no difficulty, even massive amounts of strains are analyzed. When paying attention to the three bat strains on BLSOM (arrowed in Figure 2A), these are located near influenza B strains and thus far away from the territory of birds which should be their original sources. This indicates that these bat H17N10 strains have been well adapted to mammalian cellular environments during repetitive epidemics in the mammalian host. So far judged from locations on BLSOM, cellular environments of human and bat appear to resemble each other, because human influenza B and bat A strains are located in a close vicinity.

Next, this prediction has been numerically checked, as done for the human influenza B strain in Figure 4A. In Figure 4B, we have plotted the Euclidean distance between the average composition in three bat H17N10 strains and that in H1N1 or H3N2 strains isolated in each year. As observed in Figure 4A, both H1N1 and H3N2 strains have shortened their Euclidean distances from the bat strains over time, indicating that cellular environments of human and bat appreciably resemble each other for the influenza virus growth. To sum up the findings in Figure 4A and B, the tetranucleotide composition most suitable to the growth in mammalian cells differs significantly from that for the growth in avian cells, and the avian strains invaded into a new host have to change their sequences for obtaining the composition suitable to the

new cellular environments. This directional change can offer information on predictions of viral sequence changes that will occur in the near future.

BLSOM for Codon Usage

Synonymous codon choice sensitively reflects constraints imposed on genome sequences and thus provides a sensitive probe for searching for molecular mechanisms responsible for the constraints, e.g., genome GC% and tRNA composition in the cases of micro-organisms [17-20]. We have previously found that BLSOM efficiently detects species-specific codon-choice patterns of micro-organisms, resulting in self-organization of genes according to microbial species [9]. Furthermore, in the case of genes horizontally transferred relatively recently, synonymous codon choice reflects primarily that of the donor, but not the recipient, genome. In the case of influenza viruses, we have previously found host-dependent clustering on codon-BLSOM [7], and in this chapter we introduce BLSOM for synonymous codon usage in influenza A and B virus genes (Figure 5A), by analyzing sequences including those published after our previous paper. In order to know codon biases for each strain, codon usages in eight genes are summed up for each strain.

Human and avian territories are again clearly separated from each other, and human H1N1/09 strains (dark green in Figure 5B) are again separated from the major human territory and surrounded by avian, equine and swine territories.

Synonymous codon-choice patterns of newly invading viruses, such as H1N1/09, should be close to those of the original host viruses, at least for a period immediately after the invasion. Codon choice will most likely shift towards the pattern of seasonal human viruses during many infection cycles among humans, because viruses depend on many cellular factors for their growth. Actually, the directional sequence changes in H1N1/09 towards the codon usage pattern of the seasonal human strains have been observed in our previous paper [7].

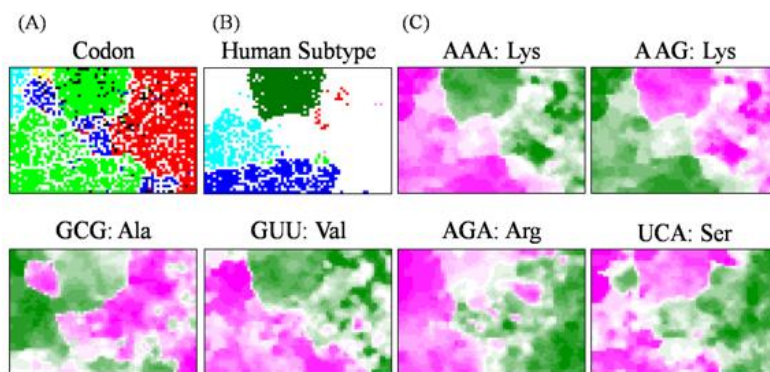


Figure 5. BLSOMs for codon usage. (A) Codon. BLSOM has been constructed for synonymous codon usage in 12,288 genes from influenza A and B strains, and lattice points are indicated in a color representing the host as described in Figure 2A. (B) Human subtype. On the Codon-BLSOM presented in (A), lattice points are indicated in a color representing the host as described in Figure 2B. (C) Occurrence levels of six codons, which are diagnostic for host-dependent separation, are indicated with different levels of two colors, as described in Figure 2C.

When we focus on diagnostic codons (Figure 5C), one simple tendency is observed. Codons ending with G or C are more favored for the avian strains than the human strains (AAG for Lys in Figure 5C), and vice versa. While the GC% effect is most apparent in two-codon boxes, this is also observed for many codons in four- or six-codon boxes (four examples in Figure 5C).

BLSOM for Amino-Acid and Peptide Composition

When we consider prediction of changes in virus sequences from a view of medical and pharmaceutical application, changes in amino acid sequences appear to be more informative than in nucleotide sequences, because amino acid sequences are more directly related to viral functions than genomic sequences; e.g., prediction of amino-acid sequence changes is very useful for designing vaccine and antiviral drug. Therefore, we next examine whether directional change of amino-acid and peptide composition can be detected with BLSOMs. We have constructed BLSOMs for mono amino-acid, dipeptide and tripeptide composition for a sum of major eight proteins (PB2, PB1, PA, HA, NP, NA, M1, and NS1) (Figures 6A and 7A, C).

Strains isolated from avian (red) and human (green) are again clustered (self-organized) according to host, forming their own large continuous territories. One simple tendency is immediately apparent. Human strains tend to prefer the amino acids coded by A- and/or U-rich codons (Asn, Ile, Lys, Phe, and Tyr; designated as A&U-AA), while avian strains appear to prefer the amino acids coded by C- and/or G-rich codons (Ala, Gly, Pro, and Arg; designated as C&G-AA) (Figure 6C). This shows that viral protein sequences are strongly affected by the host-dependent base compositions.

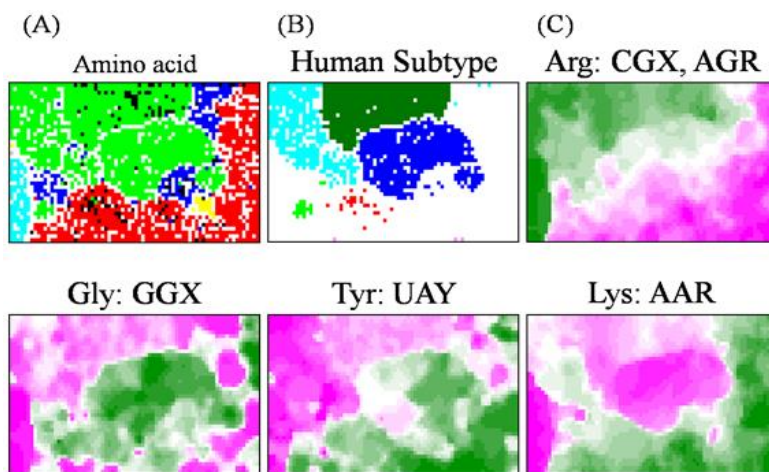


Figure 6. BLSOMs for amino-acid composition. (A) Amino acid. BLSOM has been constructed for amino-acid composition in 12,288 genes from influenza A and B strains, and lattice points are indicated in a color representing the host as described in Figure 2A. (B) Human subtype. On the Amino-BLSOM presented in (A), lattice points are indicated in a color representing the host as described in Figure 2B. (C) Occurrence levels of six amino acids, which are diagnostic for host-dependent separation, are indicated with different levels of two colors, as described in Figure 2C.

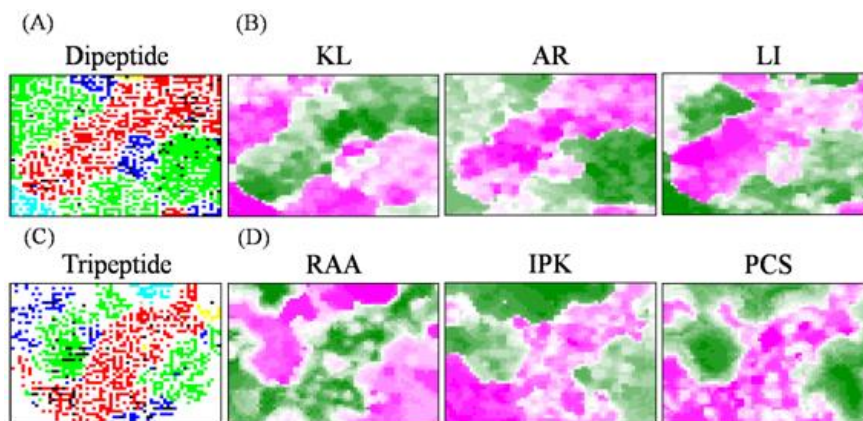


Figure 7. Oligopeptide-BLSOM for all influenza viruses. (A) Dipeptide-BLSOM. The color coding is the same as Figure 2A. (B) Occurrence levels of dipeptide, which are diagnostic for host-dependent separation, are indicated with different levels of two colors, as described in Figure 2C. (C) Tripeptide-BLSOM. The color coding is the same as Figure 2A. (D) Occurrence levels of tripeptide, which are diagnostic for host-dependent separation, are indicated with different levels of two colors, as described in Figure 2C.

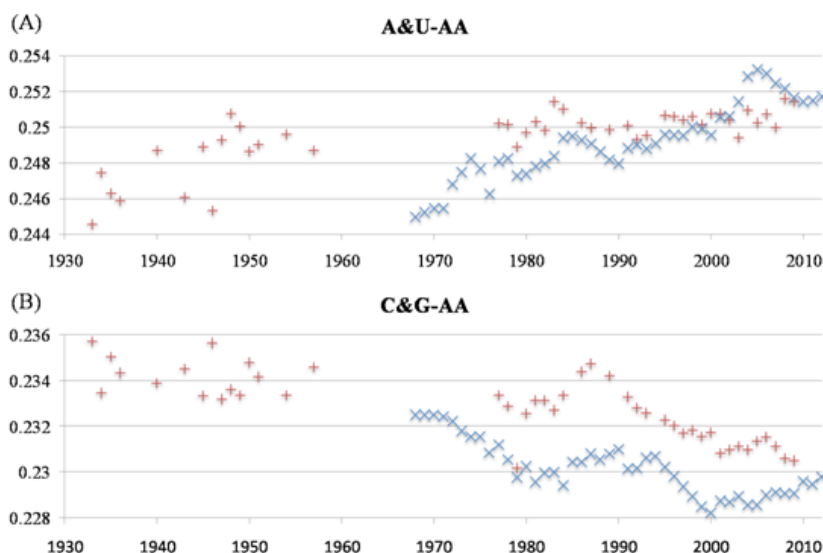


Figure 8. Directional changes of amino-acid frequency in human seasonal strains for A&U-AA and C&G-AA. (A) The A&U-AA frequencies in human seasonal H1N1(+) and H3N2(×) strains are plotted according to the chronological order. (B) The C&G-AA frequencies in human seasonal H1N1(+) and H3N2(×) strains are plotted according to the chronological order.

In order to examine the directional changes of amino-acid composition during the human virus evolution, we calculate usage frequencies of A&U-AA and C&G-AA separately, and plotted the average of their frequencies in the seasonal human H1N1 or H3N2 strains isolated in each year using the red + or blue x mark (Figure 8). A&U-AAs for both H1N1 and H3N2 strains have been gradually growing since the beginning of their pandemic while C&G-AA appear to be reducing, showing the directional change at the amino-acid level, which is consistent with the direction observed for viral genome GC%.

When we consider functions of individual viral proteins, a combination of amino acids such as di- and tripeptide may become more important than the simple amino-acid composition; i.e., prediction of directional changes of peptide compositions may provide information more directly related to change in functions and/or antigenicities of viral proteins. On both di- and tripeptide BLSOMs, strains isolated from avian (red) and human (green) are again clustered (self-organized) according to host, forming their own large continuous territories (Figure 7A, 7C). Six examples of diagnostic di- or tri-peptide for the host-dependent separation are presented in Figure 7B and 7D; most human strains prefer KL composed only of A&U-AAs and most avian strains prefer AR composed only of C&G-AAs. This is consistent to the prediction from their genome GC%. However, the trend of not relying on viral genome GC% is also seen for various di- and tri-peptides (e.g., RAA in Figure 7B). Host-dependent composition of the latter type of peptides appears to be interesting from a view of their biological significance, and therefore, detailed analyses of these peptides are in progress by our group.

CONCLUSION

Influenza viruses isolated from humans and birds differed in mono- and oligonucleotide composition and in mono amino-acid and peptide composition. Time-dependent directional changes of these compositions, which are observed for the newly invaded viruses into the human population from other animal hosts, can provide predictive information about sequence changes of the invaded viruses, which should be useful for medical and pharmaceutical purposes. Basing on the findings obtained by BLSOM analyses, we have previously proposed a strategy for efficient surveillance of potentially hazardous nonhuman strains that may cause new pandemics with a high probability [7]. Millions of influenza and other virus sequences will become available in the near future because of their medical and social importance, and BLSOM can characterize such big data without difficulty and support efficient knowledge discovery.

ACKNOWLEDGMENTS

This work was supported by Research Fellow of the Japan Society for the Promotion of Science, the Grant-in-Aid for JSPS Fellows (JSPS KAKENHI Number 24•9979), and Grant-in-Aid for Scientific Research (C) and for Young Scientists (B) from the Ministry of Education, Culture, Sports, Science and Technology of Japan. We wish to thank Drs. Masae Itoh and Kouki Yonezawa (Nagahama Institute of Bio-Science and Technology) and Dr Kimihito Ito (the Research Center for Zoonosis Control, Hokkaido University) for valuable suggestions and discussions and to Ms. Yukari Yonemori for manuscript preparation.

REFERENCES

- [1] Abe, T., Kanaya, S., Kinouchi, M., et al. (2003) Informatics for unveiling hidden genome signatures, *Genome Res.*, 13, 693–702.
- [2] Abe, T., Sugawara, H., Kinouchi, M., Kanaya, S. and Ikemura, T. (2005) Novel phylogenetic studies of genomic sequence fragments derived from uncultured microbe mixtures in environmental and clinical samples, *DNA Res.*, 12, 281–90.
- [3] Kohonen, T. (1982) Self-organized formation of topologically correct feature maps, *Biol. Cybern.*, 43, 59–69.
- [4] Kohonen, T., Oja, E., Simula, O., Visa, A. and Kangas, J. (1996) Engineering applications of the self-organizing map, *Proc. IEEE*, 84, 1358–84.
- [5] Abe, T., Sugawara, H., Kanaya, S. and Ikemura, T., Sequences from almost all prokaryotic, eukaryotic, and viral genomes available could be classified according to genomes on a large-scale Self-Organizing Map constructed with the Earth Simulator, *Journal of the Earth Simulator*, 6, 17-23, 2006.
- [6] Karlin, S., Campbell, A.M. and Mrazek, J. (1998) Comparative DNA analysis across diverse genomes, *Annu. Rev. Genet.*, 32, 185–225.
- [7] Iwasaki, Y., Abe, T., Wada, K., Itoh, M., and Ikemura, T. (2011), Prediction of Directional Changes of Influenza A Virus Genome Sequences with Emphasis on Pandemic H1N1/09 as a Model Case, *DNA res.*, 18, 125-136.
- [8] Bao, Y. (2008) The influenza virus resource at the National Center for Biotechnology Information, *J. Virol.*, 82, 596 – 601.
- [9] Kanaya, S., Kinouchi, M., Abe, T., et al. (2001) Analysis of codon usage diversity of bacterial genes with a self-organizing map (SOM): characterization of horizontally transferred genes with emphasis on the *E. coli* O157 genome. *Gene*, 276, 89-99.
- [10] García-Sastre, A. (2001) Inhibition of interferon-mediated antiviral responses by influenza A viruses and other negative-strand RNA viruses. *Virology*, 279, 375-384.
- [11] Voinnet, O. (2005) Induction and suppression of RNA silencing: insights from viral infections. *Nat. Rev. Genet.*, 6, 206-220.
- [12] Nelson, M.I., and Holmes, E.C. (2007) The evolution of epidemic influenza. *Nat. Rev. Genet.*, 8, 196-205.
- [13] Alexey, A., and Moelling, K. (2007) Dicer is involved in protection against influenza A virus infection. *J. Gen. Virol.*, 88, 2627-2635.
- [14] Rabadan, R., Levine, A.J., and Robins, H. (2006) Comparison of avian and human influenza A viruses reveals a mutational bias on the viral genomes. *J. Virol.*, 80, 11887-11891.
- [15] Domingo, E., and Holland, J.J. (1997) RNA virus mutations and fitness for survival. *Annu. Rev. Microbiol.*, 51, 151-178.
- [16] Tong, S., Li, Y., Rivallier, P., Conrardy, C., Castillo, DA., Chen, LM., et al. (2012) A distinct lineage of influenza A virus from bats. *PNAS*, 109, 4269-4274.
- [17] Ikemura, T. (1981) Correlation between the abundance of *Escherichia coli* transfer RNAs and the occurrence of the respective codons in its protein genes. *J. Mol. Biol.*, 146, 1-21.
- [18] Ikemura, T. (1985) Codon usage and transfer RNA content in unicellular and multicellular organisms. *Mol. Biol. Evol.*, 2, 13-34.

- [19] Sharp, P.M., and Matassi, G. 1994, Codon usage and genome evolution. *Curr. Opin. Gen. Dev.*, 4, 851-860.
- [20] Sueoka, N. (1995) Intrastrand parity rules of DNA base composition and usage biases of synonymous codons. *J. Mol. Evol.*, 40, 318-325.

Chapter 101

**ONCOGENES: CLASSIFICATION, MECHANISMS
OF ACTIVATION, AND ROLES IN CANCER
DEVELOPMENT ROLE OF ONCOGENES IN
GYNECOLOGICAL PATHOLOGY**

Ciro Comparetto^{1,*} and Franco Borruto²

¹Division of Obstetrics and Gynecology, City Hospital, Prato, Italy

²Obstetrics and Gynecology, Princess Grace Hospital, Monaco

ABSTRACT

An oncogene is a modified gene, or a series of nucleotides that encode a protein, and direct the cell to the development of a neoplastic phenotype. Usually, oncogenes are involved in tumor development and increase the possibility that the development (proliferation and differentiation) of a cell directs towards cancer. New researches indicate that small ribonucleic acids (RNA) of 21-25 nucleotides, called micro RNA (miRNA), can control these genes through down-regulation. The first oncogene was discovered in 1970 and named Src. Src was first discovered in a retrovirus of chickens. In 1976, J. Michael Bishop and Harold E. Varmus of the University of California showed that this oncogene was a defective proto-oncogene present in many organisms including humans. For their studies Bishop and Varmus were awarded of the Nobel Prize in 1989. A proto-oncogene is a normal gene that can become oncogenic due to mutations or to increased expression. Proto-oncogenes encode proteins that regulate the cell cycle and differentiation. They may also be involved in the signal transduction of the start of mitosis. A proto-oncogene becomes an oncogene even with minimum modifications of its original functions. There are two basic types of activation: 1) a mutation of nucleotides which produce a different protein causing: a) increased enzymatic activity of the protein; b) the loss of regulatory sites; and c) the creation of hybrid proteins; and 2) an increase of the concentration of proteins caused by: a) an increase of gene expression (through misregulation); b) an increase of stability (half-life) of the protein; and c) a duplication or amplification of the gene coding for the protein. Growth factors (GF), or fitogens, are usually secreted in a few

* Corresponding Author's Email: cicomp@tin.it.

cells specialized to induce proliferation in a paracrine, autocrine, or endocrine manner. If a cell that normally does not produce GF suddenly begins to produce them (because it has developed an oncogene), this induces the proliferation, without control, to adjacent cells (paracrine action) and to its cell type (autocrine action) increasing secretion. There are six known classes of protein kinases (PK) and related proteins that are potential oncogenes: 1) tyrosine kinase receptor (TKR), which becomes constitutively (permanently) activated, such as the epidermal growth factor receptor (EGFR), the platelet-derived growth factor receptor (PDGFR), and the vascular endothelial growth factor receptor (VEGFR); 2) cytoplasmic TK, such as TK enzymes of the Src family, Syk-ZAP-70 and Bruton's TK (BTK); 3) regulatory guanosine triphosphate (GTP)ase, like Ras; mutations that activate permanently Ras are found in 20-25% of all human tumors and up to 90% in some types of cancer, such as pancreatic ones; 4) cytoplasmic serine/threonine kinase and its regulatory units, such as Raf kinase and cyclin-dependent kinase (CDK); 5) adapter proteins in signal transduction (for example in the apoptotic pathway); and 6) transcription factors, for example the Myc gene. A large number of genes have been identified as proto-oncogenes. Most of them are responsible for the production of a positive signal that induces cell division. Other proto-oncogenes play an important role in the regulation of cell death. As mentioned before, the "altered" versions of these genes (oncogenes) may induce the cell to replicate unruly. Such a development can take place even in the absence of normal pro-growth signals, as for example the one provided by GF. A key feature in the activity of oncogenes is that a single altered copy is sufficient to induce an unregulated growth. Such behavior is in contrast with the one typical of tumor-suppressor genes (TSG) for which it is necessary that both copies of the gene are defective for triggering a process of abnormal cell division. Proto-oncogenes that have been identified until now have very different functions within the cell. Despite the differences in their normal functions, these genes all have the characteristic to contribute to unregulated cell division when present as mutated (oncogenic). The mutated proteins sometimes retain some features of their own, but are no longer sensitive to the regulation systems that determine and control the normal form of the protein itself. In this chapter we overview the role of oncogenes in gynecological pathology.

INTRODUCTION

During the past decades, the expansion of molecular biology has had a pivotal role in understanding the basis of cancer development and progression. In addition, real advances have been made in the application of deoxy-ribonucleic acid (DNA) recombinant technology to cancer therapy and patient management [1]. Early studies designed to investigate the molecular basis of oncogenesis indicated the existence of discrete genes which could cause neoplastic transformation of normal cells in vitro. These genes (which became known as oncogenes) were originally thought to be derived from oncogenic retroviruses and neoplastic transformation was believed to be the result of infection of normal cells by an oncogenic retrovirus. However, recent studies have demonstrated that normal eukaryotic cells contain gene sequences which are highly homologous to oncogenes but which do not cause neoplastic transformation. These genes (termed proto-oncogenes) have been found to code for proteins which are intimately involved in the regulation of mitosis. These include growth factors, receptors for growth factors, proteins (such as protein kinases [PK] and guanosine triphosphate [GTP] binding proteins [BP]) which transduce exogenous signals through the plasma membrane, and nuclear BP. Thus, if the function or control of a proto-oncogene (or its product) is altered by mutation,

gene rearrangement, or translocation, its effects on the cell will also change. This, in turn, can lead to the loss of normal mitotic control. The tumorigenic process is, of course, much more complicated than just unchecked cell growth. However, the insights into the molecular basis of cellular regulation have suggested new approaches to the early diagnosis of cancer and may ultimately lead to the development of more effective treatment protocols [2].

Gynecologic malignancies, representing 13% of all cancers affecting women, have a major impact on women's health. Cervical, endometrial, and ovarian cancers comprise the majority of these tumors and contribute significant morbidity and mortality to the female population. While cervical and endometrial cancers can be detected early in their development, sadly, many patients present with advanced disease, as do the majority of patients with ovarian cancer. Unfortunately, advanced cases of these malignancies are usually lethal despite modern therapeutic modalities. In order to impact upon these grim statistics, gynecologic researchers have turned to molecular biology in an attempt to elucidate the etiology of these cancers. Recent research describing dominant oncogene and tumor-suppressor gene (TSG) mutations common to these malignancies is providing a basis for the molecular genesis of these cancers. This information should offer new avenues for the development of early detection and chemoprevention, as well as novel treatment strategies [3].

ONCOGENES IN GYNECOLOGICAL CANCERS

The multifactorial process of carcinogenesis involves mutations in oncogenes, or TSG, as well as the influence of environmental etiological factors. Common DNA polymorphisms in low penetrance genes have emerged as genetic factors that seem to modulate an individual's susceptibility to malignancy. Genetic studies, which lead to a true association, are expected to increase understanding of the pathogenesis of each malignancy and to be a powerful tool for prevention and prognosis in the future [4]. Cancer is a genetic disease, and inherited or acquired genetic defects contribute to the initiation and progression of cancer [5]. The genes responsible for hereditary cancers such as retinoblastoma, colorectal cancer, and breast cancer have been identified through application of positional cloning from human molecular genetics. Linkage analysis is a powerful method to locate a disease gene, however a precise genetic model, detailing the mode of inheritance, gene frequencies and penetrance, is required for parametric methods, but not for nonparametric methods. The nonparametric methods ignore unaffected people, and look for alleles that are shared by affected individuals within nuclear families as well as extended families. Hence, the methods usually have been performed to identify disease genes in many hereditary diseases [6].

Comparative genomic hybridization (CGH) is a powerful new molecular cytogenetic method which allows genome-wide mapping of regions with DNA sequence copy number changes (both increases and decreases) in a single experiment without previous knowledge of the locations of the regions of abnormality. CGH is based on *in situ* hybridization (ISH) of differentially labeled total genomic tumor DNA and normal DNA to normal human metaphase chromosomes. After hybridization, copy number variations among the sequences in the tumor DNA are detected by measuring the tumor/normal fluorescence intensity ratio for each locus in the target chromosomes. Many previously unknown chromosomal regions with relative copy number changes have been detected in various tumors by CGH. Some changes have been

identified as genetic markers associated with biological and clinico-pathological characteristics (i.e., histopathological grade and clinical outcome) [7, 8].

Advances in molecular biology have facilitated the recent investigation of gynecological malignancies. The presence of certain oncogenes within gynecological tumors indicates that transformation may be associated with genetic alteration of normal regulatory processes. Several oncogenes have been implicated in the transformation of gynecological tissues [9]. Among them, the human epidermal growth factor receptor (EGFR) 2 (HER-2)/neu and micro-ribonucleic acids (miRNA) are two examples of the most often studied.

The HER-2/neu proto-oncogene (also known as c-erbB2, neu, and HER2) encodes a 185-kDa trans-membrane glycoprotein with intrinsic tyrosine kinase (TK) activity that resembles the receptor for EGF. Aberrant HER-2/neu protein overexpression occurs in human gynecologic adenocarcinomas, including those of the ovary, endometrium, breast, fallopian tube, and cervix, and is secondary to gene amplification and/or overexpression of the p185HER2 protein. Overexpression of HER-2/neu was found to be a poor prognostic factor for survival from advanced-stage ovarian cancer, node-positive breast cancer, and endometrial cancer. Although a specific ligand has not been definitively identified, HER-2/neu may have unusually complex activation pathways because it can form both homodimeric and heterodimeric associations with other related receptor proteins. Preliminary findings suggest that serum HER-2/neu levels may be used as a tumor marker in a subset of patients with tumors that overexpress the HER-2/neu receptor. Receptor-targeted therapeutics currently being studied include the use of receptor antibodies, liposomally delivered antisense DNA, antigen-activated cytotoxic lymphocytes (CTL), and adenovirus-mediated E1A delivery to overexpressing tumor cells. HER-2/neu appears to play an important role in the biologic behavior of ovarian, endometrial, and breast cancers and holds potential as a target for oncogene-directed therapies [10].

miRNA are 21-22 nucleotide non-coding RNA that regulate gene expression and play fundamental roles in biological processes. These small molecules bind to target messenger-RNA (mRNA), leading to translational repression and/or mRNA degradation. Aberrant miRNA expression is associated with several human diseases such as cancer, cardiovascular disorders (CVD), inflammatory diseases, and gynecological pathology. miRNA have a role in four gynecological disorders that affect the ovary or the uterus, one benign and frequent disease (endometriosis) that is classified as a tumor-like lesion and three malignant gynecological diseases (endometrial, cervical, and ovarian cancers). In cancer, miRNA have an important role as regulatory molecules, acting as oncogenes (oncomiR) or tumor-suppressors. Endometrial cancer is one of the most frequent gynecological malignancies in the developed countries. Cervical cancer, also one of the most common cancers in women, is associated with high risk Human Papillomaviruses (HPV), although this infection alone may not be enough to induce the malignant transformation. Ovarian cancer is the fifth leading cause of all cancer-related deaths among women. Over 80% of cases are diagnosed at an advanced stage, with a reduced five-year survival rate. Recent studies have shown that miRNA are aberrantly expressed in different human cancer types, including endometrial, cervical, and ovarian cancer, and that specific dysregulated miRNA may act as biomarkers of patients' outcome. Recently, miRNA have been detected in serum and plasma, and circulating miRNA expression profiles have now been associated with a range of different tumor types. Their accessibility in peripheral blood and stability, given the fact that miRNA circulate confined within exosomes, make researchers foster hope in their role as emerging biomarkers of cancer and other disorders. The development

of therapies that might block the expression or mimic the functions of miRNA could represent new therapeutic strategies for any of the aforementioned gynecological disorders [11, 12].

Breast Cancer

Breast cancer is the most common female malignancy in many industrialized countries [13]. Colony stimulating factor (CSF-1) and its receptor (CSF-1R, product of *c-fms* proto-oncogene) were initially implicated as essential for normal monocyte development as well as for trophoblastic implantation. However, studies have demonstrated that CSF-1 and CSF-1R have additional roles in mammary gland development during pregnancy and lactation. This apparent role for CSF-1/CSF-1R in normal mammary gland development is very intriguing because this receptor/ligand pair has also been found to be important in the biology of breast cancer, in which abnormal expression of CSF-1 and its receptor correlates with tumor cell invasiveness and adverse clinical prognosis. Recent findings also implicate tumor-produced CSF-1 in promotion of bone metastasis in breast cancer, and a certain membrane-associated form of CSF-1 appears to induce immunity against tumors. Recent findings on the role of CSF-1 and its receptor in normal and neoplastic mammary development may elucidate potential relationships of GF-induced biological changes in the breast during pregnancy and tumor progression [14].

The mammary gland seems to be the only organ that is not fully developed at birth. Estrogens stimulate breast tissue via estrogen receptors (ER). In the mammary gland, ER-mediated mechanisms have been shown to regulate:

1. various GF, such as transforming GF (TGF)-alpha and TGF-beta;
2. enzymes, such as cathepsin D and plasminogen-activator;
3. proto-oncogenes, such as *c-fos*, *c-myc*, and *HER-2/neu*;
4. cyclines and other regulatory substances that provide signaling systems for cell division and differentiation; and
5. other steroid receptors and EGFR.

Estrogen target genes contain estrogen-responsive elements. In these genes, transcription will be activated through interaction with the estrogen/ER protein complex. Subsequent activation of proto-oncogenes provides an explanation for the stimulating effect of estrogens on the glandular breast. Progesterone may be the key in influencing the risk of breast cancer with the peak of mitotic activity in the breast during the luteal phase of the menstrual cycle. On the other hand, in human breast cancer cell lines, both proliferation and inhibition have been observed with various progestational agents.

Relevant biological and clinical issues are pregnancy and exposure to exogenous hormones. The intense hormonal stimulation of pregnancy (both estrogen and progesterone) has no adverse impact on the course of breast cancer. Pregnancy, with its mammatogenic differentiation, results in the protection of this organ from carcinogenesis. Characterization of specific lobular morphology serves as an indicator of the level of differentiation achieved by the organ, and thus provides means to assess the risk of the gland undergoing neoplastic transformation when exposed to given agents. Sufficient evidence exists to indicate the possibility of a slightly increased risk of breast cancer after approximately one decade of

postmenopausal estrogen use. A review of the epidemiologic studies of postmenopausal hormone replacement therapy (HRT) and the risk of breast cancer fails to provide definitive evidence. Recent information derives from observations of cellular proliferation, plasma and tissue estradiol and progesterone receptor levels, and the percentage of apoptotic epithelial cells in human breast tissue. Several studies suggest that short-term, continuous combined HRT does not increase breast cancer recurrence or mortality. The participation of sexual hormones in the mammatogenic process during pregnancy might serve as an intermediate end point in assessing the effectiveness of hormones as chemopreventive agents. Investigations based on history, and breast morphology, should enable us to select estrogens and progestogens for HRT, and adopt optimal therapeutic regimens [15].

Breast cancer emerges by a multistep process which can be broadly equated to transformation of normal cells via the steps of hyperplasia, premalignant change, and carcinoma in situ (CIS). The elucidation of molecular interdependencies, which lead to development of primary breast cancer, its progression, and its formation of metastases is the main focus for new strategies targeted at prevention and treatment. Cytogenetic and molecular genetic analysis of breast cancer samples demonstrates that tumor development involves the accumulation of various genetic alterations including amplification of oncogenes and mutation or loss of TSG. Amplification of certain oncogenes with concomitant overexpression of the oncoprotein seems to be specific for certain histological types. Loss of normal tumor-suppressor protein function can occur through sequential gene mutation events (somatic alteration) or through a single mutational event of a remaining normal copy, when a germ-line mutation is present. The second event is usually chromosome loss, mitotic recombination, or partial chromosome deletion. Chromosome loci 16q and 17p harbor TSG, which seem to be pathognomonic for the development or progression of a specific histological subtype. There are an overwhelming number of abnormalities that have been identified at the molecular level which fit the model of multistep carcinogenesis of breast cancer. When the functions of all of these genes are known and how they participate in malignant progression, we will have the tools for a more rational approach to diagnosis, prevention, and treatment. A key critical long-term step in the molecular analysis of breast cancer will be to link the specific molecular damage with the effects of environmental carcinogens [16].

The question addressed is how far a multi-step progression model for sporadic breast cancer would differ from that for hereditary breast cancer. Hereditary breast cancer is characterized by an inherited susceptibility to breast cancer on basis of an identified germ-line mutation in one allele of a high penetrance susceptibility gene (such as breast cancer [BRCA]1, BRCA2, checkpoint kinase [CHEK]2, tumor protein [TP]53, or phosphatase and tensin homolog [PTEN]). Inactivation of the second allele of these TSG would be an early event in this oncogenic pathway (Knudson's "two-hit" model). Sporadic breast cancers result from a serial stepwise accumulation of acquired and uncorrected mutations in somatic genes, without any germ-line mutation playing a role. Mutational activation of oncogenes, often coupled with non-mutational inactivation of TSG, is probably an early event in sporadic tumors, followed by more, independent mutations in at least four or five other genes, the chronological order of which is likely less important. Oncogenes that have been reported to play an early role in sporadic breast cancer are MYC, cyclin D1 (CCND1), and erbB2 (HER2/neu). In sporadic breast cancer, mutational inactivation of BRCA1/2 is rare, as inactivation requires both gene copies to be mutated or totally deleted. However, non-mutational functional suppression could result from various mechanisms, such as hypermethylation of the BRCA1 promoter or binding

of BRCA2 by EMSY. In sporadic breast tumorigenesis, at least three different pathway-specific mechanisms of tumor progression are recognizable, with breast carcinogenesis being different in ductal versus lobular carcinoma, and in well differentiated versus poorly differentiated ductal cancers. Thus, different breast cancer pathways emerge early in the process of carcinogenesis, ultimately leading to clinically different tumor types. As mutations acquired early during tumorigenesis will be present in all later stages, large-scale gene expression profiling using DNA microarray analysis techniques can help to classify breast cancers into clinically relevant subtypes [17].

C-erbB-2 (neu) oncogene was found on the cell membrane in primary breast cancers. No obvious relation was found between neu oncogene, age, and lymph node status and tumor size. There was a tendency towards smaller primary tumors and more ER-negative tumors in the oncogene-positive group. No case of distant metastasis during follow-up was found among the oncogene-negative patients, while several oncogene-positive patients developed such metastases. This suggests that the neu oncogene is an independent prognostic factor, which might predict the development of distant metastasis. Further studies including more patients and long-term survival analysis are, however, needed in order to evaluate the prognostic significance of the neu oncogene [18]. Pooling of microarray datasets seems to be a reasonable approach to increase sample size when a heterogeneous disease like breast cancer is concerned. Different methods for the adaption of datasets have been used in the literature. Influences of these strategies using a pool of Affymetrix U133A microarrays from breast cancer samples have been analyzed. Data on the resulting concordance with biochemical assays of well-known parameters highlight critical pitfalls. The cutoffs derived by a method for the inference of cutoff values directly from the data without prior knowledge of the true result displayed high specificity and sensitivity. Markers with a bimodal distribution like ER, progesterone-receptor (PgR), and HER2 discriminate different biological subtypes of disease with distinct clinical courses. In contrast, markers displaying a continuous distribution like proliferation markers as Ki67 rather describe the composition of the mixture of cells in the tumor [19].

The development of new chemotherapeutic agents and concepts of radiation therapy, administered as primary, adjuvant, and palliative therapy, has led to new perspectives in breast cancer therapy. Apart from conventional chemotherapy, recently developed novel agents interfere with molecular mechanisms that are altered in cancer cells. Those targets are not necessarily breast cancer-specific. Promising strategies include inhibition of GF receptors, blocking of tumor angiogenesis and signal transduction pathways, modulation of apoptosis, cancer vaccination, and inhibition of invasion and metastasis [20]. Approximately one fourth of all women diagnosed with early breast cancer present with tumors that are characterized by erbB2 amplification. While the associated HER-2/neu receptor over-expression results in a high risk of relapse and poor prognosis, these tumors also represent a target for a selective monoclonal antibody therapy with trastuzumab (Herceptin). The combination of trastuzumab with chemotherapy has led to a considerable reduction of recurrences and to a significant reduction in breast cancer mortality both in the adjuvant and metastatic setting. Unfortunately, despite HER-2/neu over-expression, not all patients equally benefit from trastuzumab treatment, and almost all women with metastatic breast cancer eventually progress during antibody therapy. Moreover, trastuzumab is burdened with cardiotoxicity, thus increasing the risk of symptomatic congestive heart failure. In addition, the marginal costs for one year therapy of trastuzumab-based therapy, which is currently considered to be the most effective treatment regimen in the adjuvant setting, may amount for up to US\$ 40.000. Testing for erbB2

oncogene amplification by fluorescence ISH (FISH) and chromogenic ISH (CISH), respectively, and staining for HER-2/neu receptor over-expression by immunohistochemistry (IHC) represent the current standard for determining patient eligibility for trastuzumab-based therapy. However, while the negative predictive value of these assays for predicting the absence of benefit from trastuzumab-based therapy is sufficiently high, their positive predictive value remains insufficient, i.e., only a proportion of patients selected by these tests substantially benefit from trastuzumab-containing regimen. Accordingly, over the last years a number of biomarkers have been evaluated in their potential to predict response to trastuzumab-based therapies. These include markers of activation of HER-2/neu (e.g. tyrosine phosphorylated HER-2/neu in tissue and cleaved HER-2/neu extracellular domain in serum) and its dimerization partners (e.g., EGFR), respectively, but also components of HER-2/neu-induced downstream signaling pathways that are crucial for the growth inhibitory effects of trastuzumab (e.g., PTEN and phosphatidylinositol 3-kinase [PI3K]). Other parameters, such as topoisomerase-II alpha and c-myc co-amplifications, have also been identified as potentially useful predictors of response to trastuzumab-based chemotherapy regimen. While the benefit of these predictive biomarkers in the metastatic setting is currently explored, their usefulness in the adjuvant setting is still largely unknown. It is, however, undisputable that, within the group of HER-2/neu over-expressing tumors, further response predictors are needed in order to minimize trastuzumab-associated side effects, and to reduce the considerable societal costs that are associated with trastuzumab-based treatment regimen [13].

As a consequence of the success with the HER2 targeting antibody trastuzumab (Herceptin) in the treatment of early stage and metastatic breast cancer, several antibodies inhibiting cellular signaling of vascular EGF (VEGF) and EGFR were tested with respect to their efficacy in breast cancer. In phase II and III clinical trials the humanized anti-VEGF antibody bevacizumab (Avastin), alone or in combination with capecitabine, exhibited responses in patients with metastatic breast cancer. Recent developments focus on small molecules interfering with different signal transduction pathways in tumor cells. Numerous inhibitors of EGF and VEGF TK receptor (TKR) and farnesyl transferases are in early stages of clinical development for breast cancer. Another promising approach is the targeting of endothelins and their two G-protein coupled receptors (ET(A)R and ET(B)R). Due to the heterogeneity of disease and varying response to conventional systemic therapies, these new perceptions may lead to substantial patient benefit and provide a promising basis for future clinical application [21]. Other further developed compounds show promising results in clinical studies as a second generation of GF inhibitors. Different approaches in anti-angiogenic therapy are under preclinical and clinical phase-II trials. Pro-apoptotic agents show synergistic effects with docetaxel in a clinical phase-I trial. Other compounds that target heat shock protein 90 (HSP 90), histone deacetylase, and 3-hydroxy-3-methyl coenzyme A (HMG-CoA) reductase target atypical apoptotic pathways being lethal to tumor cells only but not to normal tissue, suggesting a tumor-specific way of action. Matrix metalloproteinases (MMP) inhibitors have been demonstrating promising results in patients with refractory malignant pleural effusion in a phase-I trial. Several TK inhibitors (TKI) currently under clinical investigation preliminarily show hopeful results in patients with advanced breast cancer. Furthermore, recent progress in defining the immunogenic epitopes of tumor antigens has rejuvenated the interest in cancer vaccines. Typical dose escalation studies leading to the highest clinically still tolerated dose do not appear to be equally appropriate for the estimation of efficiency of those compounds as for conventional cytotoxic regimes. Rather, escalation up to an amount of therapeutic agent that is

sufficient for maximum target inhibition should be promoted, where classical measures of cytoreduction such as complete or partial remission are replaced both by time to progression and treatment failure as an appropriate measure of the efficacy of an agent [20]. Clinical trial data indicate that epirubicin-based adjuvant treatment of breast cancer is associated with marked improvement in relapse-free and overall survival compared with traditional cyclophosphamide (CTX)/methotrexate (MTX)/5-fluorouracil (5-FU). The outcomes are comparable to those achieved with sequential use of doxorubicin/CTX and paclitaxel. Dose intensification of epirubicin is feasible, with tolerable side effects and no increased risk of cardiotoxicity beyond that expected. Clinical trial data coupled with pharmacokinetic evidence provide a strong foundation and rationale for current and future investigation of epirubicin in combination with the taxanes in the adjuvant setting. An ongoing German study is evaluating epirubicin/CTX in combination with trastuzumab as first-line therapy of metastatic breast cancer in patients whose tumors overexpress HER2/neu protein. These results, particularly the tolerability of this regimen, will form the basis for future adjuvant and neoadjuvant studies of epirubicin/trastuzumab-based regimens [22].

Ovarian Cancer

Epithelial ovarian cancer is the leading cause of death from gynecological cancers, largely owing to the development of recurrent intractable disease. Surgical and chemotherapeutic advances have been made, yet the cause of ovarian oncogenesis is poorly understood. Only a small number of distinct genetic mutations are known to contribute to ovarian carcinogenesis. Furthermore, understanding mechanistic genotype-phenotype links is complicated by frequent aneuploidy. Recognition of the familial clustering has focused investigators in the direction of isolating genetic susceptibility loci for ovarian cancer. The familial clustering of breast and ovarian cancers, as well as the chromosomal alterations and oncogenes identified, have all contributed to our understanding of the genetic factors involved in the development of ovarian cancer. Epigenetic deregulation is even more prominent, and ovarian cancers are replete with such aberrations that repress tumor-suppressors and activate proto-oncogenes. Research on cytogenetic abnormalities have led to the identification of oncogenes and TSG, which have contributed to a multistep model of molecular oncogenesis. Epigenetic therapies are emerging as promising agents for resensitizing platinum-resistant ovarian cancers. These drugs may also have the potential to alter epigenetic programming in cancer progenitor cells and provide a strategy for improving therapy of ovarian cancer [23, 24].

As with many other tumors, the origin and development of ovarian cancer is constituted by several molecular mechanisms, many of which are still unknown. Our understanding of ovarian cancerogenesis has been hampered by the lack of a well-defined precursor lesion, the lack of knowledge about tumor progression, and by the relative inaccessibility of the ovaries in the abdominal cavity. Furthermore, data in the literature are incomplete and often contradictory, and they are mainly founded on results obtained on cell lines and not on observations based on the *in vivo* study of ovarian cancer. Despite this situation, the study of control mechanisms of proliferation and differentiation in normal ovarian functioning has enabled clinicians to identify certain growth factors and oncogenes which seem to have an important role in the neoplastic transformation of ovarian tissue. Recent studies using experimental models allow us to better

define the fundamental mechanisms of carcinogenesis from the serous ovarian cells and of invasion of the abdominopelvic cavity by proximity [25, 26].

Most ovary carcinomas are epithelial tumors. Approximately 80-90% of adult ovarian cancers are assumed to originate from ovarian surface cells. The morphology of the ovarian surface epithelium changes constantly, exhibiting features such as crypts, inclusion cysts, villous processes, and different forms of Müllerian epithelium. The unique nature of ovarian surface features and their absence in the immediately adjacent peritoneal mesothelium suggest that local factors may play an important part in modifying the growth and morphology of the ovarian surface epithelium. Recent studies tend to emphasize oncogene activation, along with concomitant cytogenetic changes, in the development of ovarian cancer [27]. It has been proposed that epithelial ovarian cancers are of unifocal origin and arise from a single cell. Many alterations occur during the multistep carcinogenesis including interaction of peptide growth factors, activation of proto-oncogenes, and loss of TSG. Increased activity of TGF- α and decreased activity of TGF- β may contribute to the development of many ovarian cancers. Loss of TGF- β responsiveness has been associated with the down-regulation of c-myc expression in the development of ovarian cancer. Alternative expression of many oncogenes including ras, erbB2, and c-myc, were detected in many studies. Protein 53 (p53) mutation was detected in 50% of advanced ovarian cancer, suggesting that loss of TSG function facilitates transformation. Serum parameters like alpha-fetoprotein (AFP), carcinoembryonic antigen (CEA), cancer antigen 125 (CA-125), immunosuppressive acidic protein (IAP), lactate dehydrogenase (LDH), sialic acid (SA), TGF- α , and macrophage-CSF (M-CSF) have been used as ovarian tumor markers. None of these biochemical markers is presently consistent and specific enough to be an early detection for ovarian cancers [28].

However, the development and progression of epithelial ovarian cancer can be correlated with various biologic and molecular factors. Tumor growth has been associated with aberrant and dysfunctional expression and mutation of various genes. These genetic defects include oncogene over-expression, amplification or mutation, aberrant TSG expression or mutation, and the inappropriate expression of cytokines and GF and/or the cellular receptors for these molecules. Dysregulation of host immune responses may also play a permissive role in the pathogenesis of the disease. Since ovarian cancer has been associated with the frequency of ovulation, the repeated proliferation of epithelial cells may increase the chance of a genetic accident that could contribute to the activation of an oncogene or inactivation of a suppressor gene. These events, combined with the inherent ability of ovarian epithelial cells to respond to and produce various cytokines and GF, could promote oncogenesis [29]. Recent evidence indicates that inherited and acquired genetic mutations are the driving force behind carcinogenesis and cellular transformation. A number of proto-oncogenes and TSG are associated with ovarian carcinomas, including p53, BRCA1, and BRCA2, mismatch repair genes such as human mutS homolog 2 (hMSH2) and human mutL homolog 1 (hMLH1), and PTEN, HER-2/neu, K-ras, fms, and active serine/threonine PK 2 (AKT2). Novel genes recently implicated in ovarian tumorigenesis include novel ovarian epithelial Y2 (NOEY2), ovarian cancer-associated gene 1 protein (OVCA1), and PI3K. Although no singular gene alteration has been shown to initiate transformation in the ovarian epithelium, elucidation of the complex molecular and cellular mechanisms involving these known gene mutations may result in new clinical management strategies [30]. Ovarian cancer is caused by genetic alterations that disrupt proliferation, apoptosis, senescence, and DNA repair. Approximately 10% of ovarian cancers arise in women who have inherited mutations in cancer susceptibility genes (BRCA1 or

BRCA2). The ability to perform genetic testing allows identification of women at increased risk who can be offered prophylactic oophorectomy or other interventions aimed at preventing ovarian cancer. The vast majority of ovarian cancers are sporadic, resulting from the accumulation of genetic damage over a lifetime in the absence of a familial or heritable component. Several specific genes involved in ovarian carcinogenesis have been identified, including the p53 TSG and HER2/neu and PIK3CA oncogenes. The recent availability of expression microarrays has facilitated the simultaneous examination of thousands of genes, and this promises to extend further our understanding of the molecular events involved in the development of ovarian cancers. Hopefully, this knowledge can be translated into effective screening, treatment, surveillance, and prevention strategies in the future [31, 32]. The inability to identify relevant markers for pre-symptomatic screening in early stage or “pre-invasive” ovarian cancer has plagued investigators and clinicians facing the problems of early detection. The characteristic late stage of disease at initial presentation has hindered our understanding of the biologic progression and stepwise molecular alterations that result in ovarian carcinoma. To date, most screening studies have focused on identifying early anatomic changes using ultrasound or fluctuations in serum biomarkers such as CA-125. These screening methodologies have proven inadequate in both sensitivity and specificity for early stage ovarian cancer detection. Molecular analysis of ovarian carcinomas has revealed alterations in oncogenes and TSG associated with these tumors. The HER-2/neu oncogene, a member of the EGF family, is amplified or over-expressed in approximately 25-30% of ovarian carcinomas. Significant data substantiate an important role for HER-2/neu in the pathophysiology of ovarian cancer. While potentially an attractive surrogate endpoint biomarker (SEB), serum HER-2/neu levels have not proven to be a useful screening modality. In response to the urgent need for improved early detection for ovarian cancer, current research efforts include:

- 1) differential hybridization studies between normal and malignant ovarian epithelium to define potentially unique ovarian cancer antigens which may ultimately have utility;
- 2) defining physical alterations that occur in malignant ovarian tissues using implanted telemetry systems;
- 3) studies using positron emission tomography (PET) to detect changes in glucose metabolism between normal and malignant ovarian tissues; and
- 4) screening studies using a 3-dimensional (3-D) ultrasound unit to improve the accuracy of this technique in recognizing early neoplastic changes.

By taking diverse approaches to tackle this problem, an improved understanding of ovarian carcinogenesis should translate into the identification of appropriate SEB for early detection [33].

The discovery of peptide GF and cancer-causing genes (oncogenes and TSG) has provided us with the exciting opportunity to begin to understand the molecular pathology of human ovarian cancer. Activation of several genes, including HER-2/neu, myc, ras, and p53 have been described in some ovarian cancers. In addition, some proto-oncogenes such as the EGFR (erbB) and the M-CSF receptor (fms) are expressed along with the respective ligands (peptide GF) in some ovarian cancers. Although the studies reviewed represent a promising beginning, we remain far from a comprehensive understanding of growth regulation and transformation of human ovarian epithelium [34]. Molecular genetic evaluation of ovarian cancer primarily has utilized mutation analysis, IHC techniques, and loss of heterozygosity (LOH) studies. Over-

expression of the HER-2/neu oncogene is present in approximately one third of ovarian cancers and is associated with poor prognosis. Mutations of the K-ras oncogene have been identified in a similar proportion of mucinous ovarian tumors, including borderline tumors. Some authors have frequently detected LOH on chromosome 17, including the p53 and BRCA1 loci, and at 17p3.3 and 1717q22-23. Genetic linkage analysis indicates that the majority of inherited ovarian cancers are caused by mutations in the BRCA1 gene. Mutations in mismatch repair genes have been identified in ovarian cancers that occur as part of the hereditary non-polyposis colon-rectal cancer (HNPCC) syndrome. Sporadic ovarian tumors are the end result of a complex pathway involving multiple oncogenes and TSG, including HER-2/neu, K-ras, p53, BRCA1, and additional TSG on chromosome 17. The majority of inherited ovarian cancers are due to mutations in the BRCA1 gene, which appears to be a TSG [35]. K-ras activation is specific for mucinous tumors including adenomas, borderline tumors, and carcinomas, suggesting that K-ras activation may be associated with the mucinous differentiation rather than malignant transformation. Inactivation of p53 is detected in 30-40% of ovarian carcinoma. Mutations are more frequently observed in serous carcinomas, but not found in adenomas or rarely found in borderline tumors, suggesting that p53 mutations may be directly involved in malignant transformation. TGFbeta-2 mutations are found in 50% of endometrioid adenocarcinoma (EAC), but rarely in other type. Loss of deleted in colorectal carcinoma (DCC) mRNA expression is found in 50% of serous carcinomas but less frequently in other type. Loss of DCC expression is rare in borderline tumors and adenomas, suggesting that inactivation of DCC may be directly involved in malignant transformation. Microsatellite instability (MI) is found in 17% of ovarian carcinomas, and is frequently found in EAC. Although inactivation of p16 by point mutation or deletion is rare, p16 inactivation by loss of expression is relatively common [36].

The classic prognostic parameters are insufficient for predicting the prognosis of the individual patient. Knowledge of molecular and biological factors which are responsible for the development and progression of ovarian cancer may improve the prediction of prognosis. Recent data both on factors associated with the development and control of ovarian cancer cells and on DNA ploidy suggest that steroid and peptide hormones have a role in disease etiology and progression, and that peptide GF and cytokines, oncogenes and TSG, by their impact on mitosis and cell number may influence the rate of mutations, which could confer malignant transformation. DNA ploidy is an objective independent prognostic factor. DNA aneuploidy indicates high risk, diploidy low risk. Only tumors shown to be DNA diploid by flow-cytometry and image cytometry are considered diploid. S-phase fraction is currently not reliable. Understanding the mechanisms involved in ovarian cancer development and growth will allow opportunities for the rational design of effective anti-tumor treatment modalities. More objective and reproducible prognostic variables will improve the predictiveness of prognosis [37]. To date, there are no prognostic factors in ovarian cancer that adequately account for tumor biology and the course of the disease. In recent years, some reports have described the prognostic significance of the amplification and over-expression of the oncogene c-erbB-2 (HER2/neu) in various human cancers, including ovarian cancer. As we have seen, the c-erbB-2 proto-oncogene is located on the long arm of chromosome 17. It encodes a 185 kD trans-membrane glycoprotein receptor (p185HER2) that has sequence similarities with the EGFR. In ovarian cancer, the percentage of c-erbB-2 positive cases varies from 9% to 32%. Correlation with tumor stage and the degree of histological differentiation was not observed. The over-expression of c-erbB-2 is a new and statistically independent prognostic factor. The over-

expression of oncogene c-erbB-2 in ovarian cancer can be detected by IHC staining for the protein p185 and characterizes a group with unfavorable tumor biology and a significantly worse prognosis. Elevated serum levels of the c-erbB-2 oncoprotein have been identified in patients with various cancers known to overexpress the c-erbB-2 oncogene. Antiproliferative effects of monoclonal antibodies directed against p185 have been demonstrated in breast cancer patients. This may lead to a new approach in ovarian carcinoma therapy too, over and above the diagnostic aspects [38]. Within past years, the measurement of serological, histochemical, and molecular genetic markers has had an increasing influence on clinical decisions about initial treatment and follow-up. The most studied and interesting markers in ovarian cancer, CA-125, CA-19.9, tumor-associated trypsin inhibitor (TATI), cancer associated serum antigen (CASA), CEA, tissue polypeptide antigen (TPA), tissue polypeptide specific antigen (TPS), and cytokeratin-19 fragment (CYFRA21-1) are now the most widely used serological tumor markers for management of ovarian cancer patients. Ras oncogenes, C-erb2 proto-oncogene, p53 suppressor gene, and B-cell lymphoma 2 (bcl-2) oncogene are examples of currently used molecular genetic markers. As histochemical markers-proliferation markers, flow cytometric analysis, thymidine labeling index, Ki-67 nuclear antigen or differentiation markers are nowadays the ones most often determined. Some of these markers might be useful adjuncts for monitoring response to therapy, including early detection of tumor reactivation to allow curative therapy and rapid detection of treatment failure. The study of these markers may also lead to a better understanding of the biological characteristics of ovarian cancer. Numerous tumor markers have been recognized as promising prognostic factors. The information derived from studies of these markers also represents the most promising avenue towards new treatment strategies. Nevertheless, to validate these factors, prospective studies of a large patient population are needed [39].

The biology of ovarian cancer covers essentially all aspects of the disease: from how it arises to how it responds to chemotherapy, often becomes refractory to treatment, and ultimately kills the patient. Ovarian cancer is often initially responsive to chemotherapy but ultimately becomes refractory [40]. Previous studies have suggested the safety of conservative surgery with unilateral salpingo-oophorectomy or cystectomy for patients with stage I borderline ovarian tumors. Laparoscopic treatment of adnexal masses has proved to be a safe and effective diagnostic and therapeutic tool in the hands of experienced laparoscopists. For women who are treated conservatively, follow-up is important. Surgery remains the most effective therapy for later stage lesions. Adjuvant therapy for advanced stage of borderline ovarian tumors remains controversial. Conservative management of borderline ovarian tumors is an appropriate therapeutic option for young women with early-stage lesions who wish to preserve their childbearing potential. Available data indicate that in these patients fertility, pregnancy outcome, and survival remain excellent [41]. Conventional therapy for epithelial ovarian cancer, including aggressive cytoreductive surgery followed by combination chemotherapy regimens, has failed to reduce the number of deaths caused by this disease, which remains the most lethal of gynecologic malignancies [42]. Surgery has reached its limits, and further aggressive surgery will result in an inordinate morbidity and mortality. Ovarian carcinoma is ideally treated by complete surgical removal of the cancer, followed by anti-cancer chemotherapy. Since it is often impossible to remove all of the cancer, adjunctive chemotherapy is playing an increasingly important role in the management of the cancer [43].

Drug discovery in the ovarian cancer arena has led to the activation of several important clinical trials. Many biologic agents have come down the pipeline and are being studied in

phase II trials for recurrent disease. These agents include anti-vascular compounds that disrupt angiogenesis through a variety of mechanisms (e.g., prevention of ligand-binding to the VEGF receptor-2 (VEGF-R2), high-affinity VEGF blockade, oral TKI stimulated by VEGF, inhibition of $\alpha 5 \beta 1$ integrin, neutralization of angioproteins, etc.). Other novel drugs include oral platinum compounds as well as those that antagonize the tumor proliferation genes in the Hedgehog pathway, and that target folic acid receptors which are expressed by ovarian cancer cells. In addition, studies are underway with oral agents that inhibit the TK activity associated with two oncogenes (EGFR and HER-2/neu). Finally, emerging technologies in clinical trials include nanotechnology to enhance delivery of chemotherapy to ovarian tumors, drug resistance/sensitivity assays to guide therapy, and agents that mobilize and induce proliferation of hematopoietic progenitor cells to aid in red blood cell, white blood cell, and platelet recovery following chemotherapy [44]. Monoclonal antibodies, which offer the promise of high selectivity for detection and therapy, may be targeted to tumor-associated antigens, GF, receptors, or oncogenes. They may be used alone as immunotherapeutic agents or conjugated to chemotherapeutic drugs, toxins, or radionuclides. Radioimmunoconjugates may also be used for preoperative or intraoperative tumor localization [42]. New anti-cancer drugs must be found or synthesized, and new combinations of current anti-cancer drugs with mechanisms to protect the bone marrow must be explored. The field of genetics and the identification of the patient at high risk because of a familial history of ovarian cancer must be expanded. The role of tumor markers and oncogenes requires more in-depth study so that these signs can play a greater role in monitoring and identifying the patient with early ovarian cancer. The emerging fields of genetic engineering and biologic response modifiers are opening up new avenues for additional modalities of therapy. The expanding areas of research in cancer are starting to dispel the doom and gloom of the last five decades with a spirit of optimism for the diagnosis and treatment of ovarian cancer [43].

In summary, ovarian cancer of epithelial origin is associated with the highest mortality rate of all gynecologic malignancies. Since no symptoms or signs are manifested at the early stages of the disease, it is no surprise that in 75% of patients peritoneal metastases are found during primary surgery. Despite advances in conservative treatment methods (invasive and non-invasive), screening for early detection of the disease is not yet available, and the overall survival rate is as low as 5-15%. Recent studies in molecular biology have drawn attention to different research directions in ovarian cancer and have contributed much to our understanding of this disease and its underlying pathologic mechanisms. Some of the new aspects of this research are, specifically: hereditary ovarian cancer, genetic background in terms of chromosomal changes, DNA anomalies, oncogenes, TSG, peptide GF and cytokines, invasiveness and metastasis, and finally, drug resistance. No breakthrough has as yet occurred in any of these subjects, but results are promising. The clinical application of the steadily increasing knowledge in the biology of ovarian cancer may assist in the development of new treatment modalities that will improve survival [45]. Significant progress has been made in understanding the molecular biology of ovarian cancer and the role that single-nucleotide polymorphisms, TSG, and oncogenes play in promoting tumor cell growth and proliferation. Strategies have been developed to correct gene defects or single out ovarian cancer cells for destruction. Molecular-based therapies are now under development to specifically target receptors and signal transduction pathways that control cell proliferation and apoptosis, angiogenesis, cellular adhesion, and cell motility in ovarian tumors. The end product of this

intense investigation will be new targeted therapies that offer the hope of improving the medical management of ovarian cancer while being significantly less toxic to normal cells [46].

Endometrial Cancer

Endometrial carcinoma is today among the most common malignancies of the female genital tract in industrialized countries, with a lifetime risk among women of 2-3%, and occurs predominantly after the menopause. Recently, the prolonged life expectancy, post-menopausal use of HRT, and the availability of easily applied diagnostic techniques have led to increasing incidence of endometrial cancer. Although during the past several decades, the histopathology, spread patterns, and prognostic factors of endometrial cancers have been better defined, the clinic-pathologic and biologic prognostic parameters should be further evaluated for the better treatment results in endometrial cancer [47]. In order to improve the treatment and follow-up of these patients, various prognostic factors have been extensively studied. Patient age, stage of disease, histologic type, and histologic grade have been shown to influence survival significantly, and the prognostic impact of these traditional clinic-pathologic variables is well established. In addition, parity, hormone receptor concentration in the tumor, DNA ploidy, and morphometric nuclear grade have all been found to influence prognosis. Information about DNA ploidy has especially been used in the clinical situation to determine individualized treatment. Several prognostic markers for tumor cell proliferation, cell cycle regulation (p53, p21, and p16), and angiogenesis have been identified. It is likely that the information derived from these tumor biomarkers will reduce the need for extensive surgical staging and adjuvant treatment in endometrial carcinoma [48].

Approximately 75% of cases are diagnosed at an early stage with a tumor confined to the uterine corpus. Although most patients are cured by surgery alone, about 15-20% with no signs of locally advanced or metastatic disease at primary treatment recurs, with limited responsiveness to systemic therapy. The most common basis for determining the risk of recurrent disease has been classification of endometrial cancers into two subtypes [49]. Increasing evidence, in fact, suggests that the majority of cases can be divided into two different types of endometrial cancer based on clinic-pathological and molecular characteristics. Type I, associated with an endocrine milieu of estrogen predominance, accounts for the majority of cases and is associated with a low-stage, low-grade and endometrioid histology and develop from endometrial hyperplasia. They have good prognosis and are sensitive to endocrine treatment. In contrast, type II endometrial cancers are not associated with a history of unopposed estrogens and develop from the atrophic endometrium of elderly women, are characterized by a high-stage, high-grade and non-endometrioid histology. Mainly, they are of serous papillary or clear cell morphology, have a poor prognosis and do not react to endocrine treatment. However, the prognostic value of this distinction is limited, as up to 20% of type I endometrial cancers recur, while half of type II cancers do not [50]. Both types of endometrial cancer probably differ markedly with regard to the molecular mechanisms of transformation. The transition from normal endometrium to a malignant tumor is thought to involve a stepwise accumulation of alterations in cellular mechanisms leading to dysfunctional cell growth [49].

Molecular techniques have been used to identify specific genetic alterations in endometrial cancers. Overexpression of the HER-2/neu oncogene occurs in 10% of endometrial cancers and correlates with poor survival. Alterations in other TKR (c-fms and EGFR) also occur in some

cases. The c-myc oncogene, which encodes a nuclear transcription factor, also may be over-expressed in some invasive cancers. Mutations in the K-ras oncogene occur in 10% and in 20-30% of American and Japanese endometrial cancers, respectively. K-ras mutations also have been observed in endometrial hyperplasias, and this may represent an early event in the development of some cancers. Mutation of the p53 TSG, with resultant over-expression of mutant p53 protein, occurs in 20% of endometrial adenocarcinomas. Over-expression of p53 is associated with advanced stage and poor survival. Because p53 mutations do not occur frequently in endometrial hyperplasias, this may be a relatively late event in endometrial carcinogenesis. Recent studies have shown that mutations occur in microsatellite sequences in some endometrial cancers. Because microsatellite instability in HNPCC has been found to be caused by mutations in DNA repair genes, similar mutations are being sought in endometrial cancers. Although several molecular alterations have been identified, the molecular pathogenesis of endometrial cancer remains poorly understood [51]. Some endometrial cancers and endometrial adenocarcinoma cell lines show amplified expression of proto-oncogenes (fos, fms, myc, myb, neu, and erb-B) and augmented production of GF (CSF-1), EGF, TGF-alpha, and TGF-beta), and EGFR. Oncogene expression, the presence of ER and PgR, and the fraction of cells in S phase are useful biochemical prognostic indicators of clinical outcome, and markers recognized by monoclonal antibodies are available for use in following the clinical course of the disease and responses to treatment. In vivo and in vitro studies on normal and neoplastic tissues are providing evidence of paracrine influences on epithelial cell proliferation. Long-term administration of tamoxifen as adjuvant therapy for breast cancer has recently been found to increase the risk for development of endometrial cancer [52].

Uterine cancer is often diagnosed at an early stage and is therefore considered one of the most curable gynecologic malignancies. Despite this, a substantial number of women who present at more advanced stage or with unfavorable histologies suffer significant morbidity and death from this disease. Research continues along several fronts in an attempt to improve the prognosis for this group of women. Basic scientific research has continued to evaluate mechanisms of carcinogenesis in the hope that better targets for treatment and prevention of disease will be found. Epidemiologic studies have attempted to further define risk factors as well as elucidate risk in those patients receiving combination estrogen and progestin HRT. Clinical studies have further defined prognostic factors, and examined new surgical staging techniques and the need for adjuvant therapy after primary surgery. However, treatment options for advanced and recurrent disease remain limited [53, 54].

In summary, adenocarcinoma of the endometrium is the most common gynecologic malignancy in the United States, accounting for some 36,000 cases of invasive cancer each year. Although most endometrial carcinomas are detected at low stage, there is still a significant mortality from the disease. Hyperplastic lesions of the endometrium follow a continuum, with the risk of progression to carcinoma being related to the severity of the disorder. Risk factors associated with the development of adenocarcinoma include hyperplasia, obesity, menstrual abnormalities, diabetes, hypertension, prior pelvic irradiation, sequential oral contraceptive (OC) use, diet, and exogenous estrogen use. In post-menopausal women, prolonged life expectancy, changes in reproductive behavior and prevalence of overweight and obesity, as well as HRT use, may partially account for the observed increases of incidence rates in some countries. There is also some evidence of genetic predisposition, and some data indicating the possibility of specific genetic abnormalities and activation of oncogenes as factors determining the etiology of the disease. At this time there is no accepted screening test for endometrial

carcinoma, though the role of immunochemistry techniques for screening and follow-up has just begun to be realized. Dilatation and curettage (DandC) along with hysteroscopy remain the major means of diagnosis. In order to improve treatment and follow-up of endometrial carcinoma patients, the importance of various prognostic factors has been extensively studied. A variety of prognostic variables including tumor cell type (serous carcinoma and clear cell carcinomas [CCC] are poor prognostic types), histological grade, stage of disease, depth of myometrial invasion, lymph node status, peritoneal cytology, presence of disease in preformed vascular spaces, presence of adnexal metastases, lymphovascular space involvement, and presence of cervical involvement have been defined. Other factors currently being investigated are ER and PgR status, p53 status, flow cytometric analysis for ploidy and S-phase fraction, and oncogenes such as HER-2/neu (c-erbB-2). The identification of high-risk groups would make it possible to avoid unnecessary adjuvant treatment among patients with a good prognosis. Although the treatment plan for each patient must be individualized, the mainstay of treatment remains total abdominal hysterectomy with bilateral salpingo-oophorectomy. Metastatic and recurrent disease is usually treated with hormonal therapy and systemic chemotherapy. Radiation therapy like surgery in recurrent disease is only applicable for the treatment of local recurrences [55, 56].

Cervical Cancer

Cervical carcinoma is one of the major causes of death in women worldwide. It is difficult to foresee a dramatic increase in cure rate even with the most optimal combination of cytotoxic drugs, surgery, and radiation. Therefore, testing of molecular targeted therapies against this malignancy is highly desirable. Cervical cancer is a multistep process with accumulation of genetic and epigenetic alterations in regulatory genes, leading to activation of oncogenes and inactivation or loss of TSG. In the last decade, in addition to genetic alterations, epigenetic inactivation of TSG by promoter hypermethylation has been recognized as an important and alternative mechanism in tumorigenesis. In cervical cancer, epigenetic alterations can affect the expression of HPV as well as host genes in relation to stages representing the multistep process of carcinogenesis [57]. Significant advances have been made toward the understanding of the initiation and progression of cervical dysplasia and neoplasia at the molecular level. To date, this has not translated into improvements in diagnosis or treatment although it is a realistic expectation that this will occur. Significant variation in the proportion of tissue specimens that exhibit genetic alterations is striking. This may be attributed to different methods of analysis, different methods of tissue fixation, which influence antigen preservation, and the analysis of small numbers of samples per report, which introduces the possibility of sampling error. In spite of the variation among published reports, it is clear that several genetic alterations occur in pre-neoplastic and early-stage invasive cervical neoplasms. The prognostic applicability of oncogene mutations is a particularly interesting area of investigation that is the closest to clinical application, although additional research involving larger numbers of patients is critical. The development of convenient methods of tissue fixation that preserve the myc oncoprotein, the synthesis of specific antibodies that provide consistent results, and the application of computer-assisted image analysis to quantitate results will be particularly important in this regard [58].

Invasive cervical cancer remains a leading cause of morbidity and mortality, especially among women in the developing world where screening is either deficient or absent. Of all agents linked to the causation of this disease, high-risk HPV appears to be the strongest factor. However, not all women with HPV develop cervical cancer. Steroid contraception has been postulated to be one mechanism whereby HPV exerts its tumorigenic effect on cervical tissue. Steroids are thought to bind to specific DNA sequences within transcriptional regulatory regions on the HPV DNA to either increase or suppress transcription of various genes. Although some earlier studies were reassuring as no increased incidence of cervical cancer was observed, subsequent research has shown a causative association, especially among long-term users. The role of steroids was further enhanced by the discovery of hormone receptors in cervical tissue. Some earlier studies of OC steroids found no increased risk, even after controlling for other risk factors, including smoking and number of partners. However, prospective studies have shown a greater progression of dysplasia to CIS with more than 6 years of steroid OC use. Similar findings were also evident from other work, including the Royal College of General Practitioners (RCGP) Oral Contraception Study. The World Health Organization (WHO) Collaborative Study of Neoplasia and Steroid Contraceptives showed a relative risk of 1.2 for invasive cancer in users of the long-acting progestational contraceptive, depot-medroxyprogesterone acetate (DMPA). However, in users of more than 5 years duration, an estimate of 2.4 was reported. The upstream regulatory region (URR) of the HPV type 16 viral genome mediates transcriptional control of the HPV genome and is thought to contain enhancer elements that are activated by steroid hormones. It has been shown that steroid hormones bind to specific glucocorticoid-response elements within HPV DNA. Experimental evidence has revealed that high-risk type HPV 16 are able to stimulate the development of vaginal and cervical squamous cell carcinomas (SCC) in transgenic mice exposed to slow-release pellets of 17 beta-estradiol (E2) in the presence of human keratin-14 promoter. SCC developed in a multi-stage pathway only in transgenic mice and not in non-transgenic mice. The E6 oncoprotein of HPV 16 has been shown to bind to the p53 TSG and stimulate its degradation by a ubiquitin-dependent protease system. Steroid hormones are thought to increase the expression of the E6 and E7 HPV 16 oncogenes, which in turn bind to and degrade the p53 gene product, leading to apoptotic failure and carcinogenesis. However, the molecular basis of this remains to be proven [59]. The mere presence of the HPV is not sufficient for the development of neoplasia. Genetic and other co-factors seem to be necessary for the expression of the invasive phenotype. The expression of HPV 16 E6-E7 oncogenes results in chromosomal aneuploidy, favoring the integration of high-risk HPV genomes into cellular chromosomes. The integration of HPV 16 may not always be required for the progression to the invasive phenotype, unlike HPV 18 DNA. Such integration sites are randomly distributed over the whole genome. The genetic susceptibility of codon 98 of the fragile histidine triad has been elucidated. The interaction between Human Immunodeficiency Virus (HIV) and HPV are complex and favor the persistence and progression of cervical disease. Future research should pave the way for therapeutic vaccine development [60].

Gestational Trophoblastic Disease

Gestational trophoblastic diseases (GTD) are interrelated conditions characterized by abnormal growth of chorionic tissues with various propensities for local invasion and

metastasis. Complete mole seems to be more like choriocarcinoma and is a unique conception in that all nuclear DNA is paternally derived and all cytoplasmic DNA is maternally derived. In contrast, partial mole generally appears to be more like normal placenta and has a triploid karyotype, where the extra haploid set of chromosomes is paternally derived: GTD are characterized by altered expression of several regulatory GF and oncogenes. These results may have both prognostic and therapeutic consequences and provide insight into the relationship between normal placenta and gestational trophoblastic diseases. While differences in expression of oncoproteins may be important to the development of GTD, the precise molecular changes that are critical to pathogenesis remain unknown [61, 62]. Tumour invasion and trophoblastic invasion share the same biochemical mediators: the MMP and their inhibitors (tissue inhibitor of metalloproteinases [TIMP]). MMP are a family of enzymes capable of digesting the extracellular matrices of the host tissues. Human cytotrophoblastic cells are constitutively invasive and produce MMP. That MMP are causally related to trophoblast invasion in the endometrium is shown by the fact that TIMP inhibit cytotrophoblastic invasion in vitro. In contrast to tumor invasion of a host tissue, trophoblastic invasion during implantation and placentation is stringently controlled both in space and time. The factors responsible for these important regulatory processes are unknown, but in-vitro studies point to autocrine (trophoblastic) and paracrine (endometrial) controls by cytokines and GF. These regulators exert their effects directly or indirectly by activating nuclear transcription factors. Transcription factors are proteins or protein complexes (often the products of oncogenes) that activate genes by binding to specific sites of the DNA located in the regulatory (5' flanking) region of genes. Some authors have speculated about a potential role of these transcription factors (particularly the oncogenes Jun and Fos) in regulating trophoblast invasion [63, 64].

Oncogenes and Cancer Therapy

In the past few years, many encouraging advancements have been made in understanding the molecular mechanisms underlying carcinogenesis and tumor progression. These improvements have led to the identification of promising new targets for cancer therapy [21]. Antisense technology has emerged as an exciting and promising strategy in the fight against cancer. The antisense concept is to selectively bind short, modified DNA or RNA molecules to mRNA in cells and prevent the synthesis of the encoded protein. As anticancer agents, these molecules can be targeted against a myriad of genes involved in cell transformation, cell survival, metastasis, and angiogenesis. Indeed, the list of possible antisense targets increases as the knowledge of the genetic basis of oncogenesis expands. Antisense cancer drugs have entered human clinical trials. At least four of these compounds are currently in phase II trials, including those targeting PKC alpha, bcl-2, c-raf, and the R1-alpha subunit of PKA. A new development in antisense chemistry (peptide nucleic acids) is achieved, along with alternative antisense-related strategies (ribozymes and 2-5A-antisense) designed to overcome some of the challenges of this already encouraging technology [65].

Heparin-binding (HB)-EGF, a member of the EGF family, exerts its biological activity through activation of the EGFR and other erbB receptors. HB-EGF participates in diverse biological processes, including heart development and maintenance, skin wound healing, eyelid formation, blastocyst implantation, progression of atherosclerosis, and tumor formation, through the activation of signaling molecules downstream of erbB receptors and interactions

with molecules associated with HB-EGF. Recent studies have indicated that HB-EGF gene expression is significantly elevated in many human cancers and its expression level in a number of cancer-derived cell lines is much higher than those of other EGFR ligands. Several lines of evidence have indicated that HB-EGF plays a key role in the acquisition of malignant phenotypes, such as tumorigenicity, invasion, metastasis, and resistance to chemotherapy. Studies *in vitro* and *in vivo* have indicated that HB-EGF expression is essential for tumor formation of cancer-derived cell lines. CRM197, a specific inhibitor of HB-EGF, and an antibody against HB-EGF are both able to inhibit tumor growth in nude mice. These results indicate that HB-EGF is a promising target for cancer therapy, and that the development of targeting tools against HB-EGF could represent a novel type of therapeutic strategy, as an alternative to targeting *erbB* receptors [66].

ONCOGENES IN REPRODUCTIVE ENDOCRINOLOGY

The union of a healthy egg and a healthy sperm is required for propagation of mammalian species, and thus any factor that disrupts the normal production of female or male gametes is a potential threat to reproductive performance. These hazards to gonadal function are derived from both clinical and environmental sources, and can affect either somatic cell or germ cell lineages, or in some cases both, with equal consequences, i.e., the loss of fertility. Females of the species are particularly at risk to gonadal toxicants since, unlike males, females are born with an irreplaceable stockpile of germ cells in their ovaries at the time of birth. Natural selection processes further dwindle this precious reserve such that by the time of puberty, when eggs could actually be used for fertilization and pregnancy, the number of remaining oocytes has been depleted to less than three-quarters of the starting cohort. In the human female, this completely normal loss of oocytes eventually leads to near-exhaustion of the germ cell reserve around the fifth decade of life, and the menopause ensues. Consequently, exposure of women to potentially damaging agents, such as anti-cancer drugs, industrial chemicals, or even cigarette smoke, can have a dramatic and irreparable effect on the ovary by accelerating the natural process of germ cell depletion and, as a direct consequence, advance the time to menopause. Many gonadal toxicants exert their effects via modulation of discrete signaling pathways linked to apoptotic cell death in the female germ-line [67]. For decades, the mechanisms responsible for germ cell depletion from the ovary, either directly during the perinatal period or indirectly via follicular atresia during post-natal life, have remained relatively obscure. The recent application of sensitive biochemical techniques for the study of cell death to the analysis of ovarian function has revealed that these two events, as well as a third instance of ovarian cell degeneration (luteolysis), are dependent upon the activation of physiological cell death mechanisms. It is therefore hypothesized that the controlled deletion of ovarian cell populations is accomplished via activation of a “universal” pathway of cellular suicide involving altered expression of a conserved cohort of genes [68]. Apoptosis is a process of single-cell deletion requiring active participation of the cell in its own demise. First described in 1972, it is now known to play a major role in embryogenesis, tissue homeostasis, and neoplasia. Apoptosis can be initiated when DNA damage occurs causing the cell to pause in its reproductive cycle. If the DNA damage is beyond repair, the cell proceeds to apoptotic cell death. When the genetic mechanism involved in the pathway of apoptosis is altered, the cell

does not die. Further mutations occur by proliferation and such multiple mutational events can lead to a malignant phenotype and cancer growth. The TSG p53 causes a DNA-damaged cell to rest and attempt repair. If damage is irreparable, p53 levels will continue to increase, initiating apoptosis. Mutation of p53, found in approximately 50% of cancers, can stop the apoptotic process. Increased bcl-2 expression, an apoptosis inhibitor, also plays a role in cellular transformation and cancer growth. Its altered expression occurs in the presence of oncogene expression. Apoptosis has a role in malignant transformation, cancer growth, and response to therapy for gynecological cancers. For cervical cancer and its precursors, data on apoptotic index, bcl-2, and Bax expression have a relationship to HPV expression. In ovarian epithelial malignancies, apoptosis plays a role in chemotherapeutic responses. The data for endometrial cancer are currently limited to apoptotic index [69].

Through the study of naturally occurring mutations in humans, the creation of mutations by site-directed mutagenesis, and the production of transgenic knockout mice, further understanding of molecular reproductive endocrinology had been achieved. Mutations in the aromatase gene in females have confirmed that its deficiency results in a previously unrecognized form of sexual ambiguity with a 46,XX karyotype, and delayed puberty with multicystic ovaries. It has long been known that estrogen is necessary for skeletal growth and epiphyseal closure in the female, but aromatase and ER gene mutations in men have demonstrated for the first time, that estrogen is important for epiphyseal closure in the male. Mutations in the steroidogenic acute regulatory protein have been recently described that demonstrate the cause for lipoid congenital adrenal hyperplasia, a disorder characterized by the complete lack of steroid production. New gene mutations in human chorionic gonadotropin beta-subunits (beta-hCG), pituitary hormones, G-protein coupled receptors, G-proteins, steroid enzymes and their receptors have also been characterized recently. Site-directed mutagenesis experiments and transgenic knockout mice have been increasingly used to study the effects of normal endocrine function. Normal functions of steroid receptor genes (steroidogenic factor-1, ER, Pgr), the glycoprotein alpha-subunit, luteinizing hormone (LH) beta, and proto-oncogenes such as rearranged during transfection (RET) have been better characterized by creating knockout models. Molecular biology techniques permit these types of studies which may be difficult, if not impossible, to perform otherwise in physiologic settings [70, 71].

E2 and ER signaling have been implicated in the development and progression of several cancers. Emerging evidence suggests that the status of ER co-regulators in tumor cells plays an important role in hormonal responsiveness and tumor progression. Proline, glutamic acid, and leucine-rich protein-1 (PELP1), also known as modulator of non-genomic actions of the ER (MNAR), a novel ER co-activator that plays an essential role in the ER's actions and its expression, is de-regulated in several hormonal responsive cancers, including breast, endometrial, prostate, and ovarian cancer. The precise function of PELP1/MNAR in cancer progression remains unclear, but PELP1 appears to function as a scaffolding protein, coupling ER with several proteins that are implicated in oncogenesis. Emerging evidence suggests that PELP1/MNAR increases E2-mediated cell proliferation and participates in E2-mediated tumorigenesis and metastasis [72]. PELP1 has been shown to participate in both genomic and non-genomic functions of ER. The expression and localization of PELP1/MNAR are deregulated in a wide variety of tumors and have been implicated in the development of hormonal resistance in cancer cell lines. Emerging data suggest that PELP1/MNAR interacts with many proteins and activates several oncogenes, including Src kinase, PI3K, and signal

transducers and activators of transcription 3 (STAT3). These new results suggest that PELP1/MNAR may act as an oncogene as well as cooperating with other oncogenes [73, 74].

The skin expresses ER, PgR, and androgen receptors (AR). In the presence of steroid hormones, such as those contained in OC, the skin likely responds to hormonal signals that control the cell cycle, apoptosis, DNA replication, and other cellular functions. Some estrogen-responsive pathways have the potential to promote tumor development, including the augmentation of EGF signaling, the expression of proto-oncogenes, and inhibition of apoptosis. The question of whether OC increase the risk for the development of skin cancer, particularly melanoma, is still an area of concern. The available evidence suggests that while the skin responds to estrogens, progestins, and androgens, these responses do not significantly increase the risk of developing skin cancer when estrogen exposure is not excessive [75].

Endometriosis

Endometriosis, a disease affecting 3% to 10% of women in the reproductive-age group, is one of the most frequent benign gynecological diseases, with an unclear pathophysiology characterized by the ectopic growth of endometrial tissue under the influence of estrogen causing endometrium-like inflammatory lesions outside the uterine cavity. It is also becoming recognized as a condition in which ectopic endometrial cells exhibit abnormal proliferative and apoptotic regulation in response to appropriate stimuli. Apoptosis plays a critical role in maintaining tissue homeostasis and represents a normal function to eliminate excess or dysfunctional cells. Accumulated evidence suggests that, in healthy women, endometrial cells expelled during menstruation do not survive in ectopic locations because of programmed cell death, while decreased apoptosis may lead to the ectopic survival and implantation of these cells, resulting in the development of endometriosis. Both the inability of endometrial cells to transmit a “death” signal and the ability of endometrial cells to avoid cell death have been associated with increased expression of anti-apoptotic factors and decreased expression of pre-apoptotic factors. Further investigations may elucidate the role of apoptosis-associated molecules in the pathogenesis of endometriosis. Medical treatment with apoptosis-inducing agents may be novel and promising therapeutic strategy for endometriosis [76].

Endometriosis is well established as a condition showing heritable tendencies. Polygenic/multi-factorial etiology appears far more likely to be the etiology than Mendelian inheritance. The current task is to determine the number and location of genes responsible for endometriosis. Endometriosis is a genetic disorder of polygenic/multi-factorial inheritance and bears similarity to neoplasia and, hence, is a multistep phenomenon of clonal origin [77]. Recently, a number of studies have investigated genetic polymorphisms as a possible factor contributing to the development of endometriosis. Current data regarding genes with nucleotide polymorphisms investigated with regard to endometriosis found a strikingly large amount of conflicting results. About 50% of the reviewed studies demonstrated positive correlations between different polymorphisms and endometriosis. This relation is most clearly seen in groups 1 (cytokines and inflammation), 2 (steroid-synthesizing enzymes and detoxifying enzymes and receptors), 4 (estradiol metabolism), 5 (other enzymes and metabolic systems), and 7 (adhesion molecules and matrix enzymes). Group 8 (apoptosis, cell-cycle regulation, and oncogenes) seemed to be negatively correlated with the disease, whereas group 3 (hormone receptors), 6 (GF systems), and especially 9 (human leukocyte antigen [HLA] system

components) showed a relatively strong correlation. Polymorphisms may have a limited value in assessing possible development of endometriosis. There are some single nucleotide polymorphisms (SNP) relationships that are clinically stronger than others [78]. In a series of studies, it has been hypothesized that endometriotic proliferation is, in part, precipitated by mutations in oncogenes or deletions in TSG that have been shown to be important steps in the transformation from a benign to a malignant epithelium. It has been reported previously that no mutations in the p53 and k-ras genes in cases of endometriosis were found. However, having shown that endometriotic deposits were monoclonal, some authors showed loss of heterozygosity on chromosomes 9p (18%), 11q (18%), and 22q (15%); in total 28% of endometriotic lesions showed loss of heterozygosity at one or more sites. Any loss of heterozygosity in normal endometrium could be demonstrated. Adjacent endometriosis, atypical endometriosis, and EAC of the ovary have been examined and showed common genetic alterations that are consistent with a common lineage. These common alterations were not seen in lesions that were distant from each other. In EAC, an increased frequency of mutations in the PTEN/methylmalonic aciduria cobalamin deficiency (MMAC) TSG has been reported that was not seen in CCC or serous carcinoma, suggesting distinct developmental pathways for these tumors [79]. Similarly to tumor metastasis, endometriotic implants require neo-vascularization to proliferate, invade the extracellular matrix, and establish an endometriotic lesion. Despite its high prevalence and incapacitating symptoms, the exact pathogenic mechanism of endometriosis remains unsolved. A relationship between endometriosis and gynecological cancer, especially ovarian cancer, has been reported. Endometriosis is a multifactorial and polygenic disease, and emerging data provide evidence that a dysregulation of miRNA expression may be involved. miRNA appear to be potent regulators of gene expression in endometriosis, raising the prospect of using miRNA as biomarkers and therapeutic tools in this disease [11].

There is evidence that endometriosis, as well as drugs used in the process of in vitro fertilization (IVF), appear to associate with increased risk for gynecological cancer. There are data to support that ovarian endometriosis could have the potential for malignant transformation. Epidemiologic and genetic studies support this notion. It seems that endometriosis is associated with specific types of ovarian cancer (EAC and CCC). There is no clear association between endometriosis and breast or endometrial cancer. More studies are needed to establish the risk factors that may lead to malignant transformation of this condition and to identify predisposed individuals who may require closer surveillance. Currently, there is no proven relationship between any type of gynecological cancer and drugs used for infertility treatment. In principle, infertile women have increased risk for gynecologic malignancies. Nulligravidas who received treatment are at increased risk for malignancy compared with women who had conceived after treatment. There is limited evidence that clomiphene citrate use for more than six cycles or 900 mg or treatment of women over the age of 40 could increase their risk for ovarian and breast cancer. More studies with the appropriate statistical power and follow-up time are required to evaluate accurately the long-term effects of these drugs and procedures [80]. Thus, although population-based studies have unequivocally reported an increased risk of ovarian cancer in women with endometriosis, the biological evidence supporting the idea of endometriosis as a pre-neoplastic condition is scanty and not well substantiated. The fundamental features of human neoplasms (monoclonal growth, genetic changes, mutations in TSG and replicative advantage) have been evaluated in endometriotic lesions but results obtained are discordant. It is plausible that ectopic glands may

expand monoclonally but the entity of this phenomenon is debated. According to some allelotyping studies, from one-third to one-half of endometriosis lesions would harbor somatic genetic changes in chromosomal regions supposed to contain genes involved in ovarian tumorigenesis, especially for the endometrioid histotype. These findings would be consistent with the progression model for carcinogenesis from the benign precursor to ovarian cancer but they could not be unequivocally replicated. Gene mutational studies are rare in this context. A single group has found missense mutations and deletions of PTEN gene in about 20% of ovarian endometriotic cysts. Moreover, in a model of genetically engineered mice harboring an oncogenic allele of k-ras resulting in benign lesions reminiscent of endometriosis, a conditional deletion of PTEN caused the progression towards the EAC. Based on these data, the causal link between endometriosis and ovarian EAC/CCC remains to be defined both in terms of entity of association and of underlying molecular mechanisms [81]. In conclusion, although endometriosis generally remains a benign condition, it demonstrates somatically acquired genetic alterations. CCC and EAC are the most frequent types of epithelial ovarian carcinoma (EOC) associated with endometriosis. Retrograde menstruation or ovarian hemorrhage carries highly pro-oxidant factors, such as iron, into the peritoneal cavity or ovarian endometrioma. CCC and EAC should be considered separately in studies of endometriosis-associated EOC. The repeated events of hemorrhage in endometriosis can contribute to carcinogenesis and progression via three major processes:

- 1) increasing oxidative stress promotes DNA methylation;
- 2) activating anti-apoptotic pathways supports tumor promotion; and
- 3) aberrant expression of stress signaling pathways contributes to tumor progression [82, 83].

CONCLUSION

Improved molecular techniques have led to the identification of many genetic mutations in gynecologic malignancies [5]. In gynecologic oncologic fields, there are many investigations to explore the basic pathogenesis of gynecologic cancer, such as cervical cancer, ovarian cancer, and endometrial cancer. It is now known that specific types of human papilloma virus (HPV) are the principal etiologic agents for both cervical cancer and its precursors. However, the various kinds of alterations in oncogenes and tumor suppressor genes may play additional roles in carcinogenesis of cervical cancer. Although ovarian carcinoma is the most frequent cause of death from gynecologic malignancies, the histogenesis and biological characteristics of these tumors are not well understood. During the last several years, many key observations have been made concerning the genetic alterations associated with ovarian cancer. Recent researches including some dominant oncogenes and tumor suppressor gene mutations common to these malignancies are providing bases to elucidate the mechanisms underlying this cancer. The most important basis of endometrial cancer is that K-ras and p53 mutations are also frequently observed [1]. The molecular characterization of cancer has provided a better understanding of tumor formation and the clinical behavior of different tumor types, with important implications for developing screening tests and prognostic markers. Applications of these findings have led to novel targeted gene therapies that correct the critical genetic defects

seen in gynecologic cancers. Future research will focus on the clinical translation of these genetic alterations as targets of cancer prevention, screening, and treatment [5].

REFERENCES

- [1] Whang, JD; Lee, JH. Molecular genetics of gynecologic cancer. *J. Korean Med. Sci.*, 1997 12, 383-389.
- [2] Lynch, DH. Oncogenes and cancer. *Am. J. Reprod. Immunol. Microbiol.*, 1987 15, 24-28.
- [3] Taylor, RR; Teneriello, MG; Nash, JD; Park, RC; Birrer, MJ. The molecular genetics of gyn malignancies. *Oncology*, (Williston Park), 1994 8, 63-70.
- [4] Ueda, M; Toji, E; Nunobiki, O; Izuma, S; Okamoto, Y; Torii, K; Noda, S. Germline polymorphism of cancer susceptibility genes in gynecologic cancer. *Hum. Cell*, 2008 21, 95-104.
- [5] Cass, I; Baldwin, RL; Karlan, BY. Molecular advances in gynecologic oncology. *Curr. Opin. Oncol.*, 1999 11, 394-400.
- [6] Sekine, M; Tanaka, K. Linkage analysis for identifying genes implicated in hereditary cancers. *Nihon Rinsho*, 2000 58, 1206-1210.
- [7] Sakamoto, M; Sakamoto, H; Suehiro, Y; Akiya, T; Iwabuchi, H; Sakunaga, H; Muroya, T; Noda, T; Sugishita, T; Tenjin, Y. CGH (comparative genomic hybridization). *Nihon Rinsho*, 1996 54, 933-943.
- [8] Matsuta, M; Nishiya, I. Comparative genomic hybridization (CGH). *Hum. Cell*, 1993 6 231-236.
- [9] Maxwell, GL; Carlson, JW. Oncogenes in gynecologic oncology. *Obstet. Gynecol. Surv.*, 1996 51, 710-717.
- [10] Cirisano, FD; Karlan, BY. The role of the HER-2/neu oncogene in gynecologic cancers. *J. Soc. Gynecol. Investig.*, 1996 3, 99-105.
- [11] Gilabert-Estelles, J; Braza-Boils, A; Ramon, LA; Zorio, E; Medina, P; Espana, F; Estelles, A. Role of microRNAs in gynecological pathology. *Curr. Med. Chem.*, 2012 19, 2406-2413.
- [12] Huang, Y. A mirror of two faces: Lin28 as a master regulator of both miRNA and mRNA. *Wiley Interdiscip. Rev. RNA*, 2012 3, 483-494.
- [13] Singer, CF; Köstler, WJ; Hudelist, G. Predicting the efficacy of trastuzumab-based therapy in breast cancer: current standards and future strategies. *Biochim. Biophys. Acta*, 2008 1786, 105-113.
- [14] Sapi, E. The role of CSF-1 in normal physiology of mammary gland and breast cancer: an update. *Exp. Biol. Med. (Maywood)*, 2004 229, 1-11.
- [15] Schneider, HP; Jackisch, C. Potential benefits of estrogens and progestogens on breast cancer. *Int. J. Fertil. Womens Med.*, 1998 43, 278-285.
- [16] Beckmann, MW; Niederacher, D; Schnürch, HG; Gusterson, BA; Bender, HG. Multistep carcinogenesis of breast cancer and tumour heterogeneity. *J. Mol. Med. (Berl)*, 1997 75, 429-439.
- [17] Kenemans, P; Verstraeten, RA; Verheijen, RH. Oncogenic pathways in hereditary and sporadic breast cancer. *Maturitas*, 2004 49, 34-43.

- [18] Makar, AP; Desmedt, EJ; De Potter, CR; Vanderheyden, JS; Schatteman, EA. Neu (C-erbB-2) oncogene in breast cancer and its possible association with the risk of distant metastases. A retrospective study and review of literature. *Acta. Oncol.*, 1990 29, 931-934.
- [19] Karn, T; Metzler, D; Ruckhäberle, E; Hanker, L; Gätje, R; Solbach, C; Ahr, A; Schmidt, M; Holtrich, U; Kaufmann, M; Rody, A. Data-driven derivation of cutoffs from a pool of 3,030 Affymetrix arrays to stratify distinct clinical types of breast cancer. *Breast Cancer Res. Treat.*, 2010 120, 567-579.
- [20] Sauer, G; Deissler, H; Kurzeder, C; Kreienberg, R. New molecular targets of breast cancer therapy. *Strahlenther. Onkol.*, 2002 178, 123-133.
- [21] Fischgräbe, J; Wülfing, P. Targeted therapies in breast cancer: established drugs and recent developments. *Curr. Clin. Pharmacol.*, 2008 3, 85-98.
- [22] Untch, M; von Koch, F; Kahlert, S; Konecny, G. Overview of epirubicin-based adjuvant therapy in breast cancer. *Clin. Breast. Cancer*, 2000 1 Suppl 1, S41-S45.
- [23] Murphy, SK. Targeting the epigenome in ovarian cancer. *Future Oncol.*, 2012 8, 151-164.
- [24] Karlan, BY; Krakow, D. Genetic aspects of ovarian cancer. *Curr. Opin. Obstet. Gynecol.*, 1994 6, 105-110.
- [25] van den Brûle, FA; Lambotte, R; Castronovo, V. Epithelial cancers of the ovary. Recent trends. *J. Gynecol. Obstet. Biol. Reprod.*, (Paris), 1995 24, 241-252.
- [26] Jassoni, VM; Amadori, A; Gentile, G; Alesi, L. Potential role of growth factors in ovarian cancer. *Front Biosci.*, 1996 1, g14-g19.
- [27] Dietl, J; Marzusch, K. Ovarian surface epithelium and human ovarian cancer. *Gynecol. Obstet. Invest.*, 1993 35, 129-135.
- [28] Chow, SN; Chien, CH; Chen, CT. Molecular biology of human ovarian cancer. *Int. Surg.*, 1996 81, 152-157.
- [29] Berek, JS; Martínez-Maza, O. Molecular and biologic factors in the pathogenesis of ovarian cancer. *J. Reprod. Med.*, 1994 39, 241-248.
- [30] Li, AJ; Karlan, BY. Genetic factors in ovarian carcinoma. *Curr. Oncol. Rep.*, 2001 3, 27-32.
- [31] Wenham, RM; Lancaster, JM; Berchuck, A. Molecular aspects of ovarian cancer. *Best. Pract. Res. Clin. Obstet. Gynaecol.*, 2002 16, 483-497.
- [32] Baker, VV. Molecular biology and genetics of epithelial ovarian cancer. *Obstet. Gynecol. Clin. North Am.*, 1994 21, 25-40.
- [33] Karlan, BY. Screening for ovarian cancer: what are the optimal surrogate endpoints for clinical trials? *J. Cell Biochem. Suppl.*, 1995 23, 227-232.
- [34] Berchuck, A; Kohler, MF; Boente, MP; Rodriguez, GC; Whitaker, RS; Bast, RC Jr. Growth regulation and transformation of ovarian epithelium. *Cancer*, 1993 71, 545-551.
- [35] Gallion, HH; Pieretti, M; DePriest, PD; van Nagell, JR Jr. The molecular basis of ovarian cancer. *Cancer*, 1995 76, 1992-1997.
- [36] Fujita, M; Enomoto, T; Murata, Y. Genetic alterations in ovarian carcinoma: with specific reference to histological subtypes. *Mol. Cell Endocrinol.*, 2003 202, 97-99.
- [37] Berek, JS; Martínez-Maza, O; Hamilton, T; Tropé, C; Kaern, J; Baak, J; Rustin, GJ. Molecular and biological factors in the pathogenesis of ovarian cancer. *Ann. Oncol.*, 1993 4 Suppl 4, 3-16.

- [38] Meden, H; Kuhn, W. Overexpression of the oncogene c-erbB-2 (HER2/neu) in ovarian cancer: a new prognostic factor. *Eur. J. Obstet. Gynecol. Reprod. Biol.*, 1997 71, 173-179.
- [39] Mazurek, A; Nikliński, J; Laudański, T; Pluygers, E. Clinical tumour markers in ovarian cancer. *Eur. J. Cancer Prev.*, 1998 7, 23-35.
- [40] Auersperg, N; Edelson, MI; Mok, SC; Johnson, SW; Hamilton, TC. The biology of ovarian cancer. *Semin. Oncol.*, 1998 25, 281-304.
- [41] Tinelli, R; Tinelli, A; Tinelli, FG; Cicinelli, E; Malvasi, A. Conservative surgery for borderline ovarian tumors: a review. *Gynecol. Oncol.*, 2006 100, 185-191.
- [42] Muto, MG; Kassis, AI. Monoclonal antibodies used in the detection and treatment of epithelial ovarian cancer. *Cancer*, 1995 76, 2016-2027.
- [43] Barber, HR. New frontiers in ovarian cancer diagnosis and management. *Yale J. Biol. Med.*, 1991 64, 127-141.
- [44] Chase, DM; Mathur, N; Tewari, KS. Drug discovery in ovarian cancer. *Recent. Pat. Anticancer. Drug. Discov.*, 2010 5, 251-260.
- [45] Rabinerson, D; Kaplan, B; Levavi, H; Neri, A. The biology of ovarian cancer of epithelial origin. *Isr. J. Med. Sci.*, 1996 32, 1128-1133.
- [46] DiSaia, PJ; Bloss, JD. Treatment of ovarian cancer: new strategies. *Gynecol. Oncol.*, 2003 90, S24-S32.
- [47] Kim, JW; Kim, SH; Kim, YT; Kim, DK. Clinicopathologic and biological parameters predicting the prognosis in endometrial cancer. *Yonsei. Med. J.*, 2002 43, 769-778.
- [48] Salvesen, HB; Akslen, LA. Molecular pathogenesis and prognostic factors in endometrial carcinoma. *APMIS*, 2002 110, 673-689.
- [49] Gründker, C; Günthert, AR; Emons, G. Hormonal heterogeneity of endometrial cancer. *Adv. Exp. Med. Biol.*, 2008 630, 166-188.
- [50] Engelsens, IB; Akslen, LA; Salvesen, HB. Biologic markers in endometrial cancer treatment. *APMIS*, 2009 117, 693-707.
- [51] Berchuck, A; Boyd, J. Molecular basis of endometrial cancer. *Cancer*, 1995 76, 2034-2040.
- [52] Gurpide, E. Endometrial cancer: biochemical and clinical correlates. *J. Natl. Cancer Inst.*, 1991 83, 405-416.
- [53] Yamada, SD; McGonigle, KF. Cancer of the endometrium and corpus uteri. *Curr. Opin. Obstet. Gynecol.*, 1998 10, 57-60.
- [54] Takai, N; Ueda, T; Nishida, M; Nasu, K; Miyakawa, I. The relationship between oncogene expression and clinical outcome in endometrial carcinoma. *Curr. Cancer Drug. Targets*, 2004 4, 511-520.
- [55] Noumoff, JS; Faruqi, S. Endometrial adenocarcinoma. *Microsc. Res. Tech.*, 1993 25, 246-254.
- [56] Uharcek, P. Prognostic factors in endometrial carcinoma. *J. Obstet. Gynaecol. Res.*, 2008 34, 776-783.
- [57] Lu, Q; Ma, D; Zhao, S. DNA methylation changes in cervical cancers. *Methods Mol. Biol.*, 2012 863, 155-176.
- [58] Baker, VV. Oncogene expression in cervical cancer. *Cancer Treat Res.*, 1994 70, 43-51.
- [59] Moodley, M; Moodley, J; Chetty, R; Herrington, CS. The role of steroid contraceptive hormones in the pathogenesis of invasive cervical cancer: a review. *Int. J. Gynecol. Cancer*, 2003 13, 103-110.

- [60] Moodley, M. Update on pathophysiologic mechanisms of human papillomavirus. *Curr. Opin. Obstet. Gynecol.*, 2005 17, 61-64.
- [61] Fulop, V; Mok, SC; Gati, I; Berkowitz, RS. Recent advances in molecular biology of gestational trophoblastic diseases. A review. *J. Reprod. Med.*, 2002 47, 369-379.
- [62] Fulop, V; Mok, SC; Berkowitz, RS. Molecular biology of gestational trophoblastic neoplasia: a review. *J. Reprod. Med.*, 2004 49, 415-422.
- [63] Bischof, P; Campana, A. A putative role for oncogenes in trophoblast invasion? *Hum. Reprod.*, 2000 15 Suppl 6, 51-58.
- [64] Bajoria, R; Sooranna, SR; Ward, BS; Chatterjee, R. Prospective function of placental leptin at maternal-fetal interface. *Placenta*, 2002 23, 103-115.
- [65] Kushner, DM; Silverman, RH. Antisense cancer therapy: the state of the science. *Curr. Oncol. Rep.*, 2000 2, 23-30.
- [66] Miyamoto, S; Yagi, H; Yotsumoto, F; Kawarabayashi, T; Mekada, E. Heparin-binding epidermal growth factor-like growth factor as a novel targeting molecule for cancer therapy. *Cancer Sci.*, 2006 97, 341-347.
- [67] Tilly, JL. Molecular and genetic basis of normal and toxicant-induced apoptosis in female germ cells. *Toxicol. Lett.*, 1998 102-103, 497-501.
- [68] Tilly, JL. Apoptosis and ovarian function. *Rev Reprod*, 1996 1, 162-172.
- [69] Sheets, EE; Yeh, J. The role of apoptosis in gynaecological malignancies. *Ann. Med.*, 1997 29, 121-126.
- [70] Layman, LC. Molecular biology in reproductive endocrinology. *Curr. Opin. Obstet. Gynecol.*, 1995 7, 328-339.
- [71] McDonough, PG. Oncogenic role of the Y chromosome. *Ann. N Y Acad. Sci.*, 1997 816, 356-361.
- [72] Rajhans, R; Vadlamudi, RK. Comprehensive analysis of recent biochemical and biologic findings regarding a newly discovered protein-PELP1/MNAR. *Clin. Exp. Metastasis.*, 2006 23, 1-7.
- [73] Nair, S; Vadlamudi, RK. Emerging significance of ER-coregulator PELP1/MNAR in cancer. *Histol. Histopathol.*, 2007 22, 91-96.
- [74] Vadlamudi, RK; Kumar, R. Functional and biological properties of the nuclear receptor coregulator PELP1/MNAR. *Nucl. Recept. Signal.*, 2007 5, e004.
- [75] Leslie, KK; Espey, E. Oral contraceptives and skin cancer: is there a link? *Am. J. Clin. Dermatol.*, 2005 6, 349-355.
- [76] Nasu, K; Nishida, M; Kawano, Y; Tsuno, A; Abe, W; Yuge, A; Takai, N; Narahara, H. Aberrant expression of apoptosis-related molecules in endometriosis: a possible mechanism underlying the pathogenesis of endometriosis. *Reprod. Sci.*, 2011 18, 206-218.
- [77] Simpson, JL; Bischoff, FZ. Heritability and molecular genetic studies of endometriosis. *Ann. N Y Acad. Sci.*, 2002 955, 239-251.
- [78] Falconer, H; D'Hooghe, T; Fried, G. Endometriosis and genetic polymorphisms. *Obstet. Gynecol. Surv.*, 2007 62, 616-628.
- [79] Thomas, EJ; Campbell, IG. Molecular genetic defects in endometriosis. *Gynecol. Obstet. Invest.*, 2000 50 Suppl 1, 44-50.
- [80] Vlahos, NF; Economopoulos, KP; Fotiou, S. Endometriosis, in vitro fertilisation and the risk of gynaecological malignancies, including ovarian and breast cancer. *Best Pract. Res. Clin. Obstet. Gynaecol.*, 2010 24, 39-50.

- [81] Viganó, P; Somigliana, E; Chiodo, I; Abbiati, A; Vercellini, P. Molecular mechanisms and biological plausibility underlying the malignant transformation of endometriosis: a critical analysis. *Hum. Reprod. Update*, 2006 12, 77-89.
- [82] Kobayashi, H; Kajihara, H; Yamada, Y; Tanase, Y; Kanayama, S; Furukawa, N; Noguchi, T; Haruta, S; Yoshida, S; Naruse, K; Sado, T; Oi, H. Risk of carcinoma in women with ovarian endometrioma. *Front Biosci.*, (Elite Ed), 2011 3, 529-539.
- [83] Sekizawa, A; Amemiya, S; Otsuka, J; Saito, H; Farina, A; Okai, T; Tachikawa, T. Malignant transformation of endometriosis: application of laser microdissection for analysis of genetic alterations according to pathological changes. *Med. Electron. Microsc.*, 2004 37, 97-100.

Chapter 102

**THE CLASSIFICATION, MECHANISMS
OF ACTIVATION AND ROLES IN CANCER
DEVELOPMENT OF ONCOGENES**

***Patricio Barros-Núñez*, Mónica Alejandra Rosales-Reynoso
and Clara Ibet Juárez-Vázquez***

Centro de Investigación Biomédica de Occidente, Instituto Mexicano del Seguro Social,
Guadalajara, Jalisco, México

ABSTRACT

Cancer is an uncontrolled cell growth caused by accumulation of genetic and epigenetic mutations in genes that normally play a role in the regulation of cell proliferation, survival, apoptosis and cell cycle. Mutations are occurring mainly in oncogenes, tumor-suppressor genes, microRNA genes, or as DNA repair defects and aberrant DNA methylation. Oncogenes encode proteins that control cell proliferation and/or apoptosis. Products of oncogenes can be classified into 6 groups by its biological activity: transcription factors, growth factors, growth factor receptors, signal transducers, chromatin remodeling and apoptotic regulators. The main changes related to oncogene activation are chromosomal translocations and mutations that can occur as early events or during tumor progression; whereas amplification usually occurs during the tumor progression. These alterations are usually somatic events, although germ-line mutations can also predispose to familial cancer. A single genetic change is rarely insufficient for developing a malignant tumor. Most evidences point to a multistep process of sequential alterations in a wide number of oncogenes, tumor-suppressor genes, or microRNA genes related with cancer.

Recently, other mechanisms as the inflammation or alterations in the cellular metabolism have been associated with cancer. This chapter describes some of the key molecular mechanisms involved in the development and progression of cancer.

* Corresponding Author's Email: pbarros_gdl@yahoo.com.mx (Genetic Research. Centro de Investigación Biomédica de Occidente. Instituto Mexicano del Seguro Social. Sierra Mojada 800. Colonia Independencia. CP: 44340. Guadalajara, Jalisco, México, Tel: 52-33 3668 3000 Ext. 31922).

INTRODUCTION

During already long time, it has been accepted that the cancer is caused by the accumulation of genetic and epigenetic changes that lead to abnormal regulation of the cell growth control [1]. Discovery that abnormal genes, named oncogenes, play a crucial role in developing the malignant disease has been decisive in understanding the genetic nature of cancer. More recently, six biological capabilities acquired during the multistep development of human tumors were established as the principal hallmarks of cancer. These features rationalize the complexities of the neoplastic disease and comprise: sustaining proliferative signaling, evading growth suppressors, resisting cell death, enabling replicative immortality, inducing angiogenesis, and activating invasion and metastasis.

In addition, two essential components of the tumorigenic disease, which guides the tumor progress, are: genome instability, which generates the genetic diversity that accelerates the hallmarks acquisition, and inflammation, which promotes multiple hallmark functions. Two other emerging tumor features have been added recently: reprogramming of energy metabolism and the microRNA genes that can initiate tumorigenesis, enhance its progression and evade the immune destruction. Finally, neoplastic cells exhibit another type of complexity: a repertoire of recruited and apparently normal cells that contribute to create a “tumor microenvironment” [2-5]. Only the whole understanding of this biological complexity will permit us to develop more efficient cancer treatments. In this chapter we intend review the most important features related with the oncogenes, their activation mechanisms and roles in cancer, and the malignant metabolic reprogramming.

ONCOGENES

Are altered forms of normal cellular genes called proto-oncogenes. In human cancers, proto-oncogenes are frequently located adjacent to chromosomal breakpoints and are targets for mutation.

Products of proto-oncogenes are highly conserved in evolution and serve to regulate the cascade of events that maintains the ordered progression through the cell cycle, reproduction, and differentiation. In cancer cells, this ordered progression is partially lost when one or more of the components of this pathway are altered [6].

Our understanding of the molecular mechanisms leading to cancer has considerably advanced from the study of oncogenes. The application of techniques from many cancer research disciplines has led to the discovery of both dominantly acting transforming genes and of tumor suppressor genes [6].

Discovery of oncogenes. The majority of oncogenes were initially isolated as altered forms of proto-oncogenes acquired (transduced) by RNA tumor viruses (*v-onc*). In 1990, Payton Rous [7] discovered that transplantable sarcomas in chickens could be induced by a cell free agent. The transforming agent was a retrovirus that had transduced part of a normal cellular gene called *src* (sarcoma).

The virally transduced *src* gene (*v-src*) was altered by mutation compared with its normal cellular counterpart (*c-src*), rendering it constitutively activated. This discovery demonstrated

that our cells harbor genes that, when abnormally activated, are capable of inducing tumorigenesis [8, 9].

Over 70 proto-oncogenes activated by proviral insertion of a non-transforming retrovirus have been identified. This number includes some genes first identified as viral oncogenes [10-12].

Detection of oncogenes in human tumors. Evidence for a genetic role in cancer comes from multiple sources. Many of the cancer prone syndromes, such as Fanconi syndrome, Bloom syndrome, and ataxia telangiectasia, show greatly increased chromosome instability [13]. Studies of colon cancer demonstrate that many cancers have accumulated multiple chromosome deletions and mutations [14].

Identification that the bacterial *mutS* and *mutL* DNA mismatch repair genes as genetic lesions that predispose individuals to colon cancer further supports the role of mutation in the generation of cancer [15, 16].

Tumoral transformation. Transforming events in cancer development includes three stages: initiation, promotion and progression [17].

Table 1. Chromosome translocation in human cancers [1]

Affected gene	Rearrangements	Disease	Protein type
Oncogenes juxtaposed with IG loci			
c-MYC	t(8:14) (q24;q32) t(2:8) (p12;q24) t(8:22) (q24;q11)	Burkitt Lymphoma; BL-ALL	HLH domain
BCL1 (Cyclin D1)	t(11:14) (q13;q32)	B-cell chronic lymphocyte leukemia	PRADI-G1 Cyclin D1
BCL-2	t(14:18) (q32;q21)	Follicular Lymphoma	Inner mitochondrial membrane
BCL-3	t(14:19) (q32;q13.1)	Chronic B cell leukemia	CDC10 motif
IL-3	t(5:14) (q31;q32)	Acute pre-B cell leukemia	Growth factor
Oncogenes juxtaposed with TRC loci			
c-MYC	t(8:14) (q24;q11)	Acute T cell leukemia	HLH domain
LYLA	t(7:19) (q35;p13)	Acute T cell leukemia	HLH domain
TALI/SCL/TCL-5	t(1:14) (q32;q11)	Acute T cell leukemia	HLH domain
TAL-2	t(7:9) (q35;q34)	Acute T cell leukemia	HLH domain
Rhombotin 1/ttg-1	t(11:14) (p15;q11)	Acute T cell leukemia	HLH domain
Rhombotin 2/ttg-2	t(11:14)(p13;q11) t(7:11) (q35;p13)	Acute T cell leukemia	HLH domain
HOX11	t(10:14) (q24;q11)	Acute T cell leukemia	Homeodomain
TAN-1	t(7:9)(q34;q34.3)	Acute T cell leukemia	Notch homologue
TCL-1	t(14:14) (q11;q32.1), t(7q35-14q32.1) or inv14(q11;q32.1)	T- cell prolymphocytic leukemia	
Hematopoietic tumor			
Gene fusion			
c-ABL (9q34) BCR (22q11)	t(9:22) (q34;q11)	Chronic and acute myelogenous leukemia	Tyrosine Kinase activated by BCR
PBX-1 (1q23)	t(1:19) (q23;p13.3)	Acute pre-B cell leukemia	Homeodomain

Table 1. (Continued)

Affected gene	Rearrangements	Disease	Protein type
E2A (19p13.3)			HLH
PML(15q21)	t(15:27) (q21;q11-22)	Acute promyelocytic leukemia	Zn finger
RAR (17q21)			
CAN (6-23)	T(6:9) (p23;q34)	Acute myeloid leukemia	No homology
DEK (9q34)			
REL	Ins (2:12) (p11.2-14)	Non- Hodgkin lymphoma	NF-KB family
ALL1(MLL)*	11q23	ALL and AML	Chromatin modifiers
Solid tumors			
Gene fusions in sarcomas			
FL11, EWS	t(11:22) (q24; q12)	Ewing's sarcoma	Ets transcription factor family
ERG, EWS	t(21:22) (q22;q12)	Ewing's sarcoma	Ets transcription factor family
ATV1, EWS	t(7:21) (q22;q12)	Ewing's sarcoma	Ets transcription factor family
ATF1, EWS	t(12:22) (q13;q12)	Soft-tissue clear cell sarcoma	Transcription factor
CHN, EWS	t(9:22) (q22.31;q12)	Myxoid chondrosarcoma	Steroid receptor family
WT1,EWS	t(11:22) (p13;q12)	Desmoplastic small round cell tumor	Wilm's tumor gene
SSX1, SSX2, SYT	t(X:18) (p11.2;q11.2)	Synovial sarcoma	HLH domain
PAX3, FKHR	t(1:13) (q36;q14)	Alveolar	Homeobox homologue
PAX7, FKHR	t(1:13) (q36;q14)	Rhabdomyosarcoma	Homeobox homologue
CHOP, TLS	t(12:16) (q13;p11)	Myxoid liposarcoma	Transcription factor
var, HMG1-C	t(var:12) (var;q13-15)	Lipomas	HMG DNA binding protein
HMG2-C	t(12:14) (q13;q15)	Leiomyomas	HMG DNA binding protein
Gene fusions in thyroid carcinomas			
RET/ptc1	Inv(10) (q11.2;q2.1)	Papillary thyroid carcinomas	Tyrosine kinase activated by H4
Ret/ptc2	t(1:17) (q11.2;q23)	Papillary thyroid carcinomas	Tyrosine kinase activated by Rla (Pka)
RET/ptc3	Inv(10) (q11.2)	Papillary thyroid carcinomas	Tyrosine kinase activated by ELE1
TRK	Inv(1) (q31;q22-23)	Papillary thyroid carcinomas	Tyrosine kinase activated by TPM3
TRK-t1 (T2)	Inv(1) (q31;q25)	Papillary thyroid carcinomas	Tyrosine kinase activated by TPR
TRK-T3	T(1q31;3)	Papillary thyroid carcinomas	Tyrosine kinase activated by TFG
Gene fusions in prostate cancer			
TMPR552	ERG	T(21q22;21q22.3)	Ets transcription factor family activated by TMPR552
	ETV1	7p21.2	
Deregulation of oncogenes			
Parathyroid tumors			
PTH deregulates PRAD1 Cyclin D1	Inv(11) (p15q; q13)	Parathyroid adenoma	PRAD1/Cyclin D1
*ALL1 can fuse with more than 50 different genes. More frequently it fuses with AF4 in ALL and AF9 in AML.			

Oncogenes encode proteins that control cell proliferation, apoptosis or both and chromatin modification [1, 18]. They can be activated by chromosomal alterations resulting from mutation or gene fusion, by juxtaposition to enhancer elements, or by amplification. Genomic changes as chromosomal rearrangements, point mutations, deletions and gene amplification, activation of proto-oncogenes and inactivation of tumor suppressor genes are required for cancer initiation [19]. Amplification usually occurs during tumor progression [1, 20, 21].

Tissue specific mutations. Oncogene activation in human tumors is specific for some tissues; for example, gene amplification of *N-myc* occurs in neuroblastoma and small cell lung cancer but is extremely rare in other tissues; bcr-abl translocation is observed in most patients with chronic leukemia; mutations in the *RAS* gene are found in a high percentage of pancreatic, colorectal and lung cancers [22]. The activation of oncogenes by chromosomal rearrangements, mutations and gene amplification confers to the cells, carrying these alterations, greater growth and survival [1, 23].

Cytogenetic rearrangements. Chromosome abnormalities commonly found in cancer cells are inversions and translocations. In hematopoietic cancers and solid tumors, translocations and inversions deregulating the oncogenic transcription are present. In prostate cancer, the mechanism of oncogenic activation is a fusion between a gene with a very active promoter as *TMPRSS2*, and other with oncogenic activity as *ERG1* [24]. In cancers of T and B cells, the most common alterations giving *MYC* gene deregulation are translocations; while gene fusion is more common in myeloid cancers and soft tissue sarcomas [1]. Table 1 summarizes the most important chromosome abnormalities giving human cancer.

Oncogene response. After the oncogenic activation by mutations, the structure of the encoded protein enhances its transforming activity [25]. An example is the *RAS* oncogenes family: *KRAS*, *HRAS* and *NRAS*. Mutations in *RAS* genes has been associated with exposure to environmental carcinogens; when mutation occurs in codon 12, 13 or 61, the *RAS* gene encodes a protein which remains in the active state, so that induces continuous cell growth. *KRAS* mutations are common in carcinomas of the lung, colon and pancreas, whereas *NRAS* mutations occur primarily in acute myelogenous leukemia and myelodysplastic syndrome [1, 26].

Activating mutations in the *BRAF* gene occur in 59% of melanomas, 18% of colorectal cancers, 14% of hepatocellular carcinomas and 11% of gliomas [27]. Most *BRAF* mutations occur in the protein kinase domain by changing the valine residue to glutamic acid at position 599 (V599E), producing a protein with uncontrolled constitutive activity that stimulates the MAP kinase cascade, allowing proliferation, differentiation and cell survival [1, 27, 28]. Table 2 summarizes some oncogenes and tumor suppressors with their metabolic changes and related diseases.

Oncogene Classification

The products of oncogenes can be classified into six broad groups: growth factors, growth factor receptors, transcription factors, chromatin remodelers, signal transducers, and apoptosis regulators (Table 3).

Table 2. Biochemical functions of oncogenes and tumor suppressor genes [29]

Gene	Function	Disease
Oncogenes		
PI3K	Activates Akt (via PIP3); reduces β -oxidation (via Akt) enzyme carnitine palmitoyltransferase 1A (CPT1A)	Ovarian and gastrointestinal cancer
Akt	Upregulates fatty acid synthase (FASN); activates mTOR complex 1	Breast and ovarian cancer
Her2	Increases, through activation of PI3K, Akt, and mTOR expression of FASN and acetyl-CoA carboxylase α (ACC α) at the translational level	Mammary carcinoma
Tyrosine kinases	Generate phosphotyrosines that can bind to pyruvate kinase isoform PKM2, converting it from a tetramer to a less active dimer	Multiple cancers
E7 from HPV16	Binds PKM2, converting it from a tetramer to a less active dimer	Cervical carcinoma
Tumor Suppressor Genes		
p53	Required for expression of SCO2 and hence optimal OXPHOS; enhances the expression of TIGAR; as glycolysis inhibitor, reduces the expression of the glycolytic enzyme phosphoglyceromutase	Multiple cancers
VHL	Ubiquitin ligase required for degradation of HIF-1 α	Clear cell renal carcinoma
TSC1 (hamartin) and TSC2 (tuberin)	Negative regulators of Rheb (Which inhibit mTOR)	Tuberous sclerosis and lymphangioleiomyomatosis
PTEN	Negative regulator of class 1 PI3K	Cowden syndrome and prostate cancer
LKB1	Required for activation of AMPK	Peutz-Jeghers and sporadic lung adenocarcinoma
NF1	Negative regulator of RAS and PI3K-Akt pathway	Neurofibromatosis
PML	Negative regulator of mTOR complex 1	Promyelocytic leukemia and lung cancer
Succinate dehydrogenase (SDH). Subunits B,C and D.	Accumulated succinate competitively inhibits HIF-1 α prolylhydroxylases (PHDs)	Paraganglioma and pheochromocytoma.
Fumarate hydratase (fumarase)	Accumulated fumarate inhibits PHDs	Leiomyomatosis and papillary renal carcinoma

Growth Factors and Growth Factor Receptors

Activation of a single growth factor gene can result in malignant transformation. Platelet-derived growth factor (PDGF) is released from platelets during coagulation, induce proliferation of various cell types and stimulate fibroblasts to participate in wound healing [1, 30]. Over-expression of PDGF induces the *in vitro* transformation of fibroblasts containing

PDGF receptors, but it does not influence fibroblasts lacking these receptors [1]. This autocrine loop entails over-expression of *PDGF-β*, an antibody against its receptor, or small molecules that block the receptor and inhibit growth of the transformed fibroblasts.

Table 3. Classification of oncogenes [1]

Oncogene	Chromosome	Neoplasm	Mechanism of Activation
Transcription factors			
v-myc	8q24.1 (MYC)	Carcinoma myelocytomatosis	Deregulated activity
N-MYC	2p24	Neuroblastoma: lung carcinoma	Deregulated activity
L-MYC	1p32	Carcinoma of lung	Deregulated activity
v-myb	6q22-24	Myeloblastosis	Deregulated activity
v-fos	14q21-22	Osteosarcoma	Deregulated activity
v-jun	1p32-p31	Sarcoma	Deregulated activity
v-ski	1q22-24	Carcinoma	Deregulated activity
v-rel	2p12-14	Lymphatic leukemia	Mutant NFkappa B
v-est-1	11p23-24	Erythroblastosis	Deregulated activity
v-erbA1	17p11-21	Erythroblastosis	T3 transcription factors
v-erbA2	3p22-24.1	Erythroblastosis	T3 transcription factors
Related to apoptosis and others			
BCL2	18q21.3	B cell lymphomas	Antiapoptotic protein
MDM2	12q14	Sarcomas	Complexes with p53
Chromatin remodelers			
ALL1 (MLL)	11q23	Chromosome translocation	Chromatin modifier
Growth factors			
v-sis	22q12.3-13.1	Glioma/fibrosarcoma	B chain PDGF
Int2	11q13	Mammary carcinoma	Member of FGF family
KS3	11q13.3	Kaposi's sarcoma	Member of FGF family
HST	11q13.3	Stomach carcinoma	Member of FGF family
Growth factor receptors			
Tyrosine kinases: integral membrane protein			
EGFR	7p1.1-1.3	Squamous cell carcinoma	EFG receptor
v-fms	5q33-34	Sarcoma	CSF1 receptor
v-KIT	4q11-21	Sarcoma/GIST	Stem cell factor receptor
v-ros	6q22	Sarcoma	?
MET	7p31	MNNG-treated human osteosarcoma cell line	HGF/SF receptor
TRK	1q21-q22	Colon/thyroid carcinomas	NGF receptor
NEU	17q11.2-12	Neuroblastoma/breast	?
RET	10q11.2	Carcinomas of thyroid Men 2A Men 2B	GFNG/NTT/ART/PSP receptor activation/ fusion proteins
Receptors lacking protein kinase activity			
Mas	6q24-27	Epidermoid carcinoma	Angiotensin receptor
Signal transducers			
Cytoplasmic tyrosine kinases			
SRC	20q12-q13	Colon carcinoma	Protein tyrosine Kinase
v-yes	18q21.3	Sarcoma	Protein tyrosine Kinase

Table 3. (Continued)

Oncogene	Chromosome	Neoplasm	Mechanism of Activation
v-fgr	1p36.1-36.2	Sarcoma	Protein tyrosine Kinase
v-fes	15q25-26	Sarcoma	Protein tyrosine Kinase
ABL	9q34.1	CML	Protein tyrosine Kinase
Membrane-associated G proteins			
H-RAS	11p15.5	Colon, lung, pancreas carcinomas	GTPase
K-RAS	12p11.1-12.1	AML, thyroid carcinoma	GTPase
N-RAS	1p11-13	Carcinoma, melanoma	GTPase
BRAF	7q34	Melanoma, thyroid, colon ovary	
Gsp	20q13.3	Adenomas of thyroid	Gs alpha
Gip	17q21.3-q22	Ovary, adrenal carcinoma	Gi alpha
GTPase exchange factor (GEF)			
Dbl	Xq27	Diffuse B cell lymphoma	GEF for Rho and Cdc42Hs
Vav	19p13.2	Hematopoietic cells	GEF for Ras ?
Serine/threonine kinases: cytoplasmic			
v-mos	8q11	Sarcoma	Protein Kinase (ser/thr)
v-raf	3p25	Sarcoma	Protein Kinase (ser/thr)
Pim-1	6p21	T-cell lymphoma	Protein Kinase (ser/thr)
Cytoplasmic regulators			
v-crk	17p13.3		SH-2/SH-3 adaptor

ALL: Acute lymphoblastic leukemia; AML: acute myeloid leukemia; CML: chronic myelogenous leukemia; GTPase: guanosine triphosphatase; PDGF: Platelet-derived growth factor

Another example is the WNT glycoprotein family that inhibits the phosphorylation of β -catenin, which participate in cell-cell adhesion and activation of several signal transduction pathways [31-33]. The APC protein actively participates controlling the activity of β -catenin. In familial adenomatous polyposis, inactivating mutations of *APC* block the degradation of β -catenin by inhibiting its phosphorylation. As a result, free β -catenin in the cytoplasm translocates into the nucleus, where activates genes involved in cell proliferation and invasion [33] (Figure 1). Growth factor receptors are altered in several cancer types (figure 2) [1, 34]. In many tumors, a deletion of the ligand-binding domain of epidermal growth factor receptor (EGFR), a trans-membrane protein with tyrosine kinase activity, causes constitutive activation of the receptor in absence of ligand binding. The activated receptor phosphorylates tyrosines in the intracellular domain of the receptor, providing interactions sites for cytoplasmic proteins containing the SRC homology domain and other binding domains. These interactions deregulate signaling in several pathways. Activating mutations occur in three other members of the EGFR family (*ERBB2*, *ERBB3* and *ERBB4*) and within the kinase domains of the HER2/neu and KIT signaling receptors [35].

Such mutations occur in lung and breast cancer and gastrointestinal stromal tumors. Two classes of clinically active anti-EGFR agents have been developed: a monoclonal antibody against the extracellular domain of the receptor (e.g., cetuximab) and competitive inhibitors of the tyrosine kinase activity of the receptor (e.g., erlotinib and gefitinib) [35].

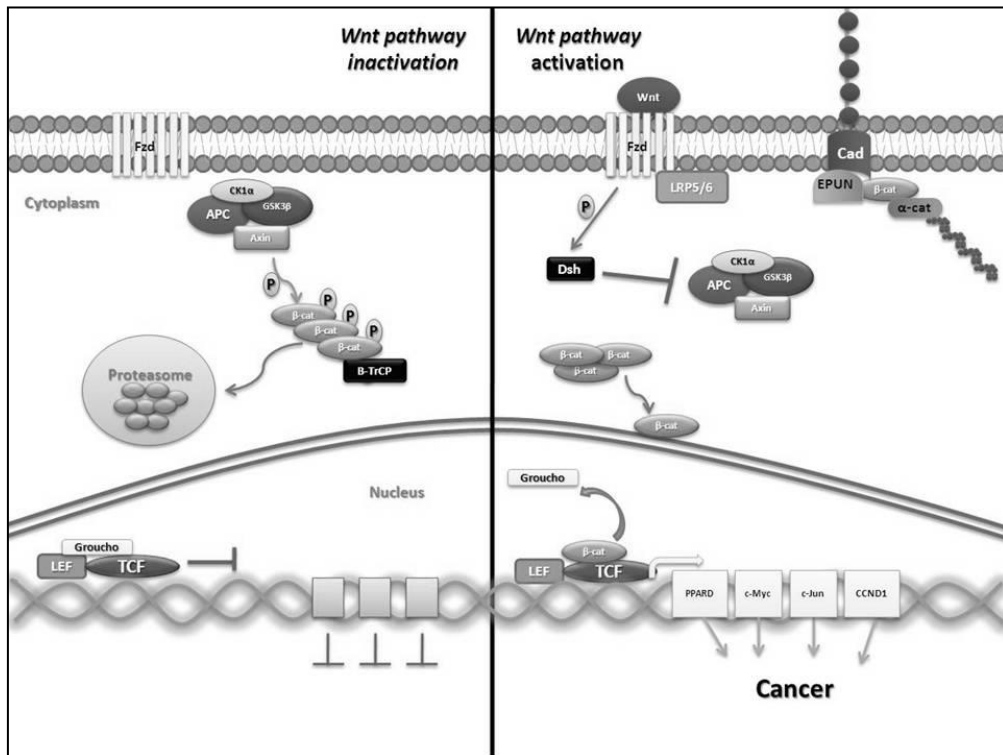


Figure 1. Dual functions of β catenina in cell adhesion and transcription activation. β catenina is in a destructive cytoplasmic complex composed of activated protein C (APC), axin, glycogen synthase kinase 3 beta (GSK-3 β), and casein kinase (CK1).

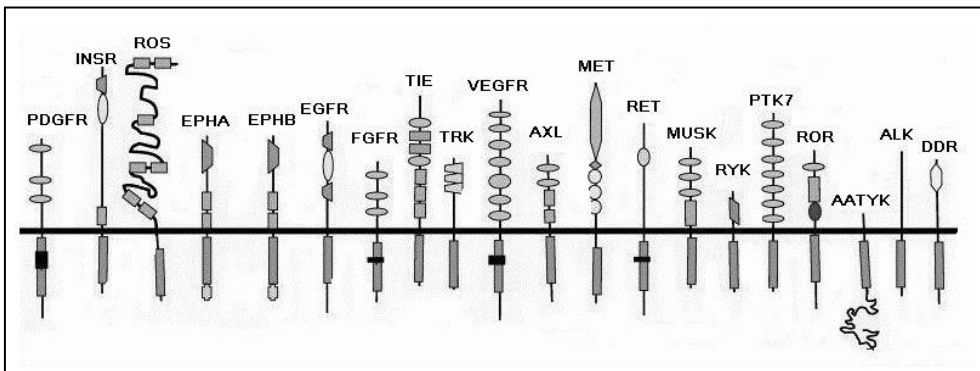


Figure 2. Examples of receptor tyrosine kinases. The epidermal growth factor (EGF), platelet-derived growth factor (PDGF) and fibroblast growth factor (FGF) receptors have been found to be involved in a variety of human cancers. Modified to [36].

Vascular endothelial growth factor (VEGF) regulates hypoxia-dependent control of gene transcription (Figure 3). The activity of VEGF is mediated by three receptor tyrosine kinases: VEGFR1 (FLT1), VEGFR2 (FLK1-KDR) and VEGFR3 (FLT4). VEGF stimulates angiogenesis in a variety of cancers, and inhibitors of VEGF and VEGFRs have been developed. Bevacizumab is a monoclonal anti-VEGF antibody, and SU5412, a small molecule, binds the receptor tyrosine kinases of the VEGFR1 and VEGFR2 as well as the kinases of the

PDGF receptor and KIT. In addition to inhibiting the ABL kinase, imatinib also inhibits the PDGF and KIT receptor kinases. Gastrointestinal stromal tumors that carry activating mutations of KIT respond to imatinib or other inhibitors of these receptor kinases [37, 38].

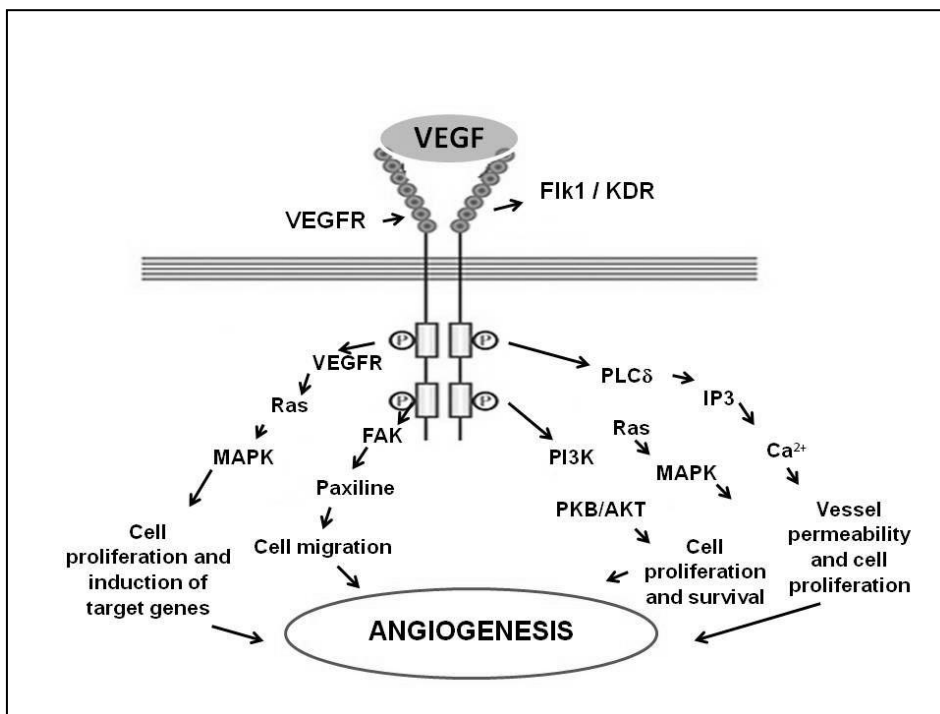


Figure 3. Role of VEGF-VEGFR interaction in angiogenesis. Several pathways are activated by the interaction of vascular endothelial growth factor (VEGF) and VEGF receptor (VEGFR) FAK denotes fatal adhesion kinase, Flk fetal liver kinase, IP3 inositol triphosphate, KDR kinase insert domain containing receptor, MAPK mitogen-activated protein kinase, PI3K phosphoinositol 3 kinase, PKB protein kinase B, and PLC phospholipase C. (Modified to [1].

Transcription Factors

Usually, transcription factors are members of multigene families sharing common structural domains. Many transcription factors require interaction with other proteins; for example, the Fos transcription protein dimerizes the Jun transcription factor to form the AP1 transcription factor, which increases the expression of several genes that control the cell division [39].

Chromosomal translocations often activate transcription factor genes in lymphoid cancers and sometimes in solid tumors (e. g., prostate cancer, see table 2). In certain sarcomas, chromosomal translocations resulting in fused proteins occur consistently; in Ewing's sarcoma, for example, the *EWS* gene is fused with one of a number of partner genes, resulting in aberrant transcriptional activity of the fused proteins.

The *EWS* protein is an RNA binding molecule with a domain that, when fused to a heterologous DNA binding domain, can greatly stimulate gene transcription. Prostate carcinomas carry translocations of the *TMPR552* gene that fuse with and activate *ERG1* or

ETV1. These genes are members of the Ets (E-26) family of transcription regulators, which can activate or repress genes involved in cellular proliferation, differentiation, and apoptosis.

The Ets family of transcription factors characterized by an evolutionarily-conserved DNA-binding domain regulates expression of a variety of viral and cellular genes by binding to a purine-rich GGAA/T core sequence in cooperation with other transcriptional factors and co-factors.

Most Ets family proteins are nuclear targets for activation of Ras-MAP kinase signaling pathway and some of them affect proliferation of cells by regulating the immediate early response genes and other growth-related genes. Some of them also regulate apoptosis-related genes.

Several Ets family proteins are preferentially expressed in specific cell lineages and are involved in their development and differentiation by increasing the enhancer or promoter activities of the genes encoding growth factor receptors and integrin families specific for the cell lineages. The fusion of TMPR552, which has androgen-responsive promoter elements, with an ETS related gene creates a fusion protein that increases proliferation and inhibits apoptosis of cells in the prostate gland, thereby facilitating their transformation into cancer cells [1, 40, 41].

Chromatin Remodelers

Chromatin compaction plays a critical role in the control of gene expression, replication and repair, and chromosome segregation. Two kinds of enzymes participate in the remodeling of chromatin: 1. ATP dependent enzymes that move the positions of nucleosomes, the repeating subunits of histones in chromatin around which DNA winds, and 2. Enzymes that modify the N-terminal tails of histones [42, 43].

The pattern of histone modifications constitutes an epigenetic code that determines the interaction between nucleosomes and chromatin associated proteins. These interactions, in turn, determine the structure of chromatin and its transcriptional capacity [44].

In acute lymphocytic leukemia and acute myelogenous leukemia, the *ALL1* (also named *MLL*) gene can fuse with 1 of more than 50 genes. *ALL1* is part of a very large, stable multiprotein complex. Most of the proteins in the complex are components of transcription complexes; others are involved in histone methylation and RNA processing [45].

The entire complex remodels, acetylates, and methylates nucleosomes and free histones. The fusion of *ALL1* with 1 of more than 50 proteins results in the formation of the chimeric proteins that underlie acute lymphoblastic leukemia and acute myelogenous leukemia. *ALL1* fusion proteins deregulate homeobox genes (which encode transcription factors) the *EPHA7* gene (which encodes a receptor tyrosine kinase), and microRNA genes such as *miR191* [45].

Signal Transducers

Binding of receptor tyrosine kinases to the appropriate ligand causes reorganization of the receptors and autophosphorylation of tyrosines in the intracellular portion of the molecules [46]. Autophosphorylation enhances the kinase activity of the receptor or promotes the interaction of the receptor with domains of cytoplasmic proteins that are effectors and

regulators of intracellular signaling [47]. In humans, there are approximately 120 SRC homology 2 domains in 100 different proteins that mediate responses to signals initiated by phosphorylated tyrosines. Some of these proteins share domains with enzymatic activity, whereas others link activated receptors to downstream targets. Many oncogenes encode members of signal transduction pathways. They fall into two main groups: nonreceptor protein kinases and guanosine-triphosphate-binding proteins. The nonreceptor proteins kinases are of two types: tyrosine kinases (e.g., ABL, LCK, and SRC) and serine and threonine kinases (e.g., AKT, RAF1, MOS and PIM1). Proteins involved in signal transduction become oncogenic if they bear activating mutations. An important example is PI3K and some of its downstream targets, such as AKT and SGK, which are critical in tyrosine kinase signaling and can be mutated in cancer cells [48, 49].

Apoptosis Regulators

The *BCL2* gene, which is involved in the initiation of roughly all follicular lymphomas and some diffuse large B-cell lymphomas [50, 51], encode a cytoplasmic protein that localizes in mitochondria and increases cell survival by inhibiting apoptosis. The *BCL2* family members inhibit apoptosis and are up-regulated in several other cancer types as chronic lymphocytic leukemia and lung cancer [52]. Two main pathways lead to apoptosis: the stress pathway and the death-receptor pathway (Figure 4). The stress pathway is triggered by proteins that contain the BCL2 homology 3-domain, which inactivates BCL2 and BCL-XL and inhibits apoptosis. Drugs that mimic the BCL2 homology 3-domain and can bind to BCL-XL or BCL2 (peptides or small organic molecules that bind in a groove of these proteins) are under development. The death-receptor pathway is activated by binding of Fas ligand, TRAIL, and tumor necrosis factor α , to their corresponding (death) receptors on the cell surface. Activation of death receptors activates caspases that cause cell death [53, 54] (Figure 4).

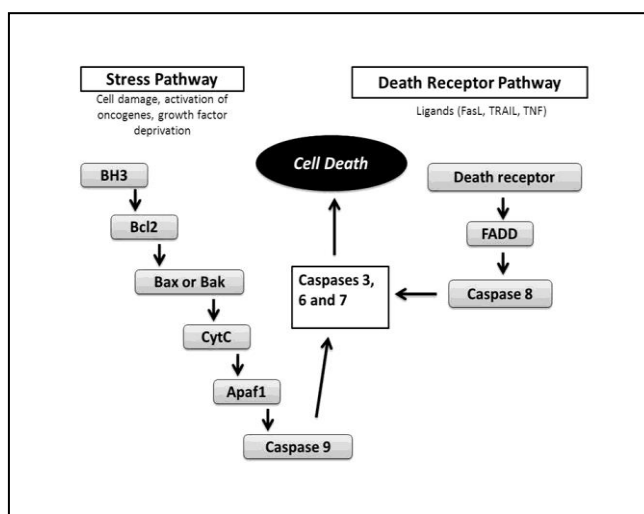


Figure 4. The two main pathways to programmed cell death, or apoptosis. The effectors of cell death are the downstream caspases, proteolytic enzymes activated by caspases 8 and 9, which are capable of clearing many of the cellular proteins causing cell death. FADD denotes Fas-associated death domain. Modified to [1].

Gene amplification usually occurs before the tumor progression; an example is the *DHFR* amplification in methotrexate-resistant acute lymphoblastic leukemia [55]. Amplification of *DHFR* is accompanied by cytogenetic abnormalities that reflect the oncogene amplification [56, 57]. The amplified DNA segment contains many genes and usually involves hundreds of kilobases. Amplification phenomenon is frequently involved in oncogenes families as *MYC*, *CCND1*, *EGFR* and *RAS*. In several cancer types as small-cell lung, breast, esophageal, cervical, ovarian and head and neck, *NMYC* amplification correlates with advanced tumor stage [58]. *CCND1* amplification occurs in breast, esophageal, hepatocellular, and head and neck cancer. *EGFR* is amplified in glioblastoma and head and neck cancer. *ERBB2* amplification in breast cancer correlates with a poor prognosis [1, 59].

Chronic Inflammation as Activation Mechanism

Inflammation is an important mechanism that can remove the agent responsible for the injury and initiate tissue repair and regeneration by a coordinated release of immune response [19]. The mechanism inflammatory involves innate and adaptive immune response. After elimination of the pathogen and wound healing, inflammation decreases [60].

However, an unsolved inflammatory process can disrupt cellular microenvironment, which leads to alterations in genes related to cancer and posttranslational modification of key proteins in the cell cycle, DNA repair and apoptosis [60].

In early stages of development and progression of tumor, mononuclear cells, macrophages, mast cells and neutrophils are present through up-regulation of pro-inflammatory cytokines such as interferon- γ , tumor necrosis factor (TNF), interleukin (IL)-1 α / β or IL-6. Also the activated nuclear factor-kB (NF-kB) is a transcription factor that relates inflammation and tumorigenesis, allowing pre-neoplastic and malignant cells evade apoptosis. Therefore, all these factors may act as initiators and promoters of carcinogenesis [60].

The role of chronic inflammation in carcinogenesis has been evaluated in several epidemiological studies, where pro-inflammatory and anti-inflammatory cytokines, viral infections and genetic markers involved in the inflammatory response were analyzed [60, 61]. It is estimated that the underlying infections and inflammatory reactions are related to 15-25% of all cancer cases [60-62].

Association between inflammatory processes and cancer are well known in the case of intestinal disease and colorectal cancer, virus hepatitis B (or C) and alcoholic liver cirrhosis or hepatocellular carcinoma, chronic esophageal reflux resulting in Barrett's esophagus and esophageal carcinoma, infection by human papilloma virus and cervical cancer, prostatitis and prostate cancer and infection by helicobacter pylori with gastric cancer [60, 63, 64]. In other cases, cancer is related with chronic irritation caused by long exposure to cigarette, silica and asbestos [63, 64].

The mechanism by which inflammation predispose cancer depends on whether it is secondary to chronic irritation or infection. In the case of HPV malignant transformation is mediated by E6 and E7 oncoproteins [65]. In other cases, the oncogenic activation occurs by the classical mode; for example, *H. pylori* contain protein factors that affect host cell signaling [66].

The chronic inflammatory response can lead to genomic injury and initiation of malignant transformation. A defense mechanism is the production of free radicals such as reactive oxygen

intermediated (ROI), hydroxyl radical ($\bullet\text{OH}$) and superoxide ($\text{O}_2\text{-}\bullet$) and reactive nitrogen intermediates (RNI), nitric oxide ($\text{NO}\bullet$) and peroxynitrite (ONOO-), which are formed by the enzymatic reaction in the host (myeloperoxidase, NADPH oxidase and nitric oxide) governed by different signaling pathways [67].

Intrinsic cellular mechanisms to prevent the unregulated proliferation or mutations in DNA are diverse, among them tumor suppressors involved in DNA repair, cell cycle arrest, apoptosis and senescence [19]. During the face of massive cell death, by infectious or noninfectious injury, cell repopulation occurs from the undifferentiated precursor cells, for which two sequential steps are required: first, some cells must survive, and second, must be a clonal expansion of this undifferentiated precursor cells in order to maintain the tissue functioning. The development of new cells is regulated by inflammatory pathways [68, 69] as part of the repair process and in defense of the infection. In initiated cells, the inflammatory response, providing cell survival and proliferation, possibly is leading to tumor promotion [19].

Evident association between tumorigenesis and host defense and / or tissue repair have been reported. Most of the studies described are based in tissue injuries and wounds that support tumor growth and neoplastic progression. For example, injection of the Rous sarcoma virus causes sarcoma in the injection site [70], probably mediated by transforming growth factor- β (TGF- β) and fibroblast growth factors (FGFs) [71]. Some metabolic signaling pathways are involved; among them the Wnt- β -catenin pathway [72] and molecules COX-1 and 2 [73].

Other studies have focused on the role NF- κB (transcription factors) in tumorigenesis. Inactivation of the classical NF- κB in colonic epithelial cells due to deletion of the I κB protein kinase β (IKK β) decreases the frequency of visible tumors [19, 74]. So in the colonic epithelium injured and that was added as a mutagen azoxymethane (AOM), the NF- κB provides a survival signal for cells initiated [75]. Otherwise, the IKK β acts protection injury infectious or non-infectious and host defense in intervening in survival of colon epithelial cells [76, 77].

Many inflammatory mediators such as cytokines, chemokines and eicosanoids stimulate the proliferation of normal and transformed cells [63]. After the administration of the phorbol ester TPA (12-O-tetradecanoylphorbol-13-acetate) and the mutagen DMBA (7, 12-dimethylbenza-anthracene) in TNF-deficient mice, fewer skin tumors were observed. Based on these experiments, the authors suggest that inflammatory mediators act as tumor promoters [78].

In the presence of tumor initiation and tissue injury and apoptosis, activation of inflammation depend tissue repair/compensatory proliferation leading to tumor promotion [19].

REPROGRAMMING OF METABOLIC PATHWAYS IN CANCER

Cancer cells require greater amounts of nutrients from the bioenergy reserves, for which different metabolic pathways are activated or modified. The signals that stimulate cell proliferation are also involved in the reorganization of the metabolic activities that allow the resting cells, initiate proliferation. The new conditions of the metabolism of cells differ from normal cells metabolism mainly by the high rate of glycolysis, lactate production, biosynthesis of lipids and other macromolecules [79].

The first observations on metabolic alterations in tumors were reported in 1920, highlighting that tumor cells of ascites in conditions of rapid proliferation consume large amount of glucose, and excrete most of the carbon as lactate, rather than oxidize it completely as do the normal cells. This phenomenon is called "Warburg effect" [80]. Warburg proposed that tumor cells use the increase of the glycolytic flow as protection before permanent damage of oxidative metabolism [79, 81]. The Warburg effect is not observed in the process of cell proliferation from all tumors [29]. This increase in the consumption of glucose, which increases the production of ATP and anabolic reactions, offers some advantages for tumor growth. *First*, for obtaining ATP, tumor cells use glucose as the most abundant extracellular nutrient in anaerobic glycolysis, and although the production of ATP per glucose is low, the high glycolytic flow generated exceeds the production of ATP from oxidative phosphorylation (OXPHOS) [29, 81]. *Second*, degradation of glucose provides the necessary intermediate compounds for different anabolic reactions, thus: ribose for synthesis of nucleotides; glucose-6-phosphate to form glycogen and ribose-5-phosphate; dihydroxyacetone phosphate for synthesis of triglycerides and phospholipids; pyruvate for synthesis of alanine and malate; and, through oxidative pentose phosphate pathway (PPP) produce nicotinamide adenine dinucleotide phosphate (NADPH) [29, 79]. *Third*, the cancer cells produce lactic and bicarbonic acid; lactate is the main end product of anaerobic glycolysis. The acidic conditions create a favorable environment for the tumor, inhibiting autoimmune effects of anti-cancer. In this atmosphere the anaerobic components (cancer cells) and aerobic (non-transformed stromal cells) are involved in metabolic pathways, complementary, recycling products of anaerobic metabolism to maintain the survival and growth of cancer cells. *Fourth*, tumors can metabolize glucose by the pentose phosphate pathway to generate NADPH and thus ensure antioxidant defenses against a hostile environment with chemotherapeutic agents [29].

In this way, the entire metabolism is reorganized to increase anabolic processes that enable growth and cell proliferation [29]. We will summarize some of the mechanisms modified in cellular signaling and key aspects of the metabolism of both normal physiological and tumorigenic proliferation. Likewise, the PI3K/Akt/mTOR pathway and the *MYC* and *HIF-1 α* transcription factors, which appear to regulate complementary aspects of cell metabolism, will be analyzed [79].

Metabolic Activity in Cell Proliferation

Normal mammalian cells do not proliferate in an autonomous way, entering the cell cycle by indication of growth factors and signaling pathways that influence gene expression and cellular physiology. The proliferating cells depend on growth factors to generate this metabolic flux and enhance the uptake of nutrients from the extracellular space [79]. In the absence of growth factors, mammalian cells quickly lose expression of transporters of nutrients and cannot keep cell autonomy for the synthesis of basal bioenergetics and macromolecular replacement. In this case, the cells carry out autophagy, which provides a limited supply of substrates generated from the macromolecular to maintain the production of ATP for cell survival [82].

The transduction of signals for cell proliferation has direct effects on the metabolic flux [79]. The mechanisms that integrate the signal transduction and cell metabolism are highly conserved between normal cells and tumor cells. The biggest difference is that in normal cells, the initiation of signaling requires extracellular stimulation, while cancer cells present

mutations chronically favoring these routes, maintaining metabolic biosynthesis phenotype, regardless of the physiological limitations of the normal cell. In other words, the cancer cells achieve an increase in metabolic autonomy. For example, in cell proliferation the enzyme lactate dehydrogenase (LDH-A) is induced by oncogenes like *HER2/neu* and *MYC* promoting cell proliferation [79]. Other enzymes highly expressed in tumors are the embryonic isoforms of pyruvate kinase (PKM2), fatty acids synthase (FASN), choline kinase (ChoK) [29, 83], and the lipogenic enzymes ATP citrate lyase and fatty acid synthase [79, 81, 84].

Rarely, tricarboxylic acid (TCA) cycle enzymes as succinate dehydrogenase (SDH) and fumarate hydratase (FH) behave as tumor suppressor; thus, mutations in the SDHB, SDHC and SDHD subunits can cause familial paraganglioma or pheochromocytoma [85], mutations in FH produce a dominant syndrome of uterine fibrosis, leiomyomata and renal carcinoma of papillary cells [79, 86]. Proliferating cells use intermediate compounds derived from the TCA cycle to synthesize lipids, proteins and nucleic acids, so use this cycle as a center of biosynthesis [79].

PI3K/Akt/mTOR Signaling Pathway

The activation of the PI3K/Akt/mTOR pathway both in growth factors dependent cells as in tumour cells, improves many of the metabolic activities that keep the cell biosynthesis by several processes. *First*, the cells increase expression of transporters and the consumption of glucose, amino acids and other nutrients [87, 88].

Second, Akt (v-akt murine thymoma viral oncogene) increases glycolysis and the production of lactate, which is sufficient to induce the Warburg effect in normal and cancer cells [89]. *Third*, PI3K and Akt stimulate the synthesis of lipids in many cell types, mTOR (mechanistic target of rapamycin) is key in regulating the protein translation [81, 90, 91].

In normal cells, the activation of the PI3K system is controlled by the phosphatase PTEN (phosphatase and tensin homolog). In malignant cells, the pathway can be triggered by mutations that activate PI3K or increase the stimuli of the system (BCR-ABL, *HER2/neu* amplification, etc.) or eliminate negative regulators of the same PTEN [79] (Table 4).

Table 4. Key mutations involved in PI3K pathway [79]

Gene	Mutation	Cancer	Frequency
PIK3CA	Gene activation	Breast	25%
		Colon	>30%
	Gene amplification	Head and neck	>35%
Akt2	Gene amplification	Ovary	12%
PTEN	Loss of heterozygosity	Glioma	≤40%
BCR-ABL	Fusion by chromosomal translocation	Chronic myelogenous leukemia	>90%
		Acute lymphocytic leukemia	20%
HER2/neu	Gene amplification	Breast	25%
EGFR	Gene amplification	Lung (non-small cell)	>50

Regulation of Glycolysis by HIF-1

One of the most important mechanisms involved in the aerobic glycolysis is the activation of hypoxia inducible factor (*HIF*); it is a transcription factor induced by hypoxic and oxidative stress, as well as oncogenic, inflammatory and metabolic factors [81, 92]. HIF-1 is a heterodimer composed of a stable β subunit and a unstable α subunit, both synthesized in normoxia by sequential action of the enzymes prolyl hydroxylase dependent of oxygen (PHDs) and ubiquitin ligase VHL [29, 81]. The active form of HIF-1 (HIF-1 α subunit) is expressed under the control of growth factor signaling pathways, mainly PI3K/Akt/mTOR [93]. In hypoxia, prolyl hydroxylation is inhibited by reactive oxygen species (ROS) from mitochondria, resulting in stable transcriptional activity of HIF-1 complex [79]. In tumor cells, stimulation by growth factors is required to express HIF-1 α , necessary for regulating the intracellular fate of carbon derived from glucose [79] (Figure 5).

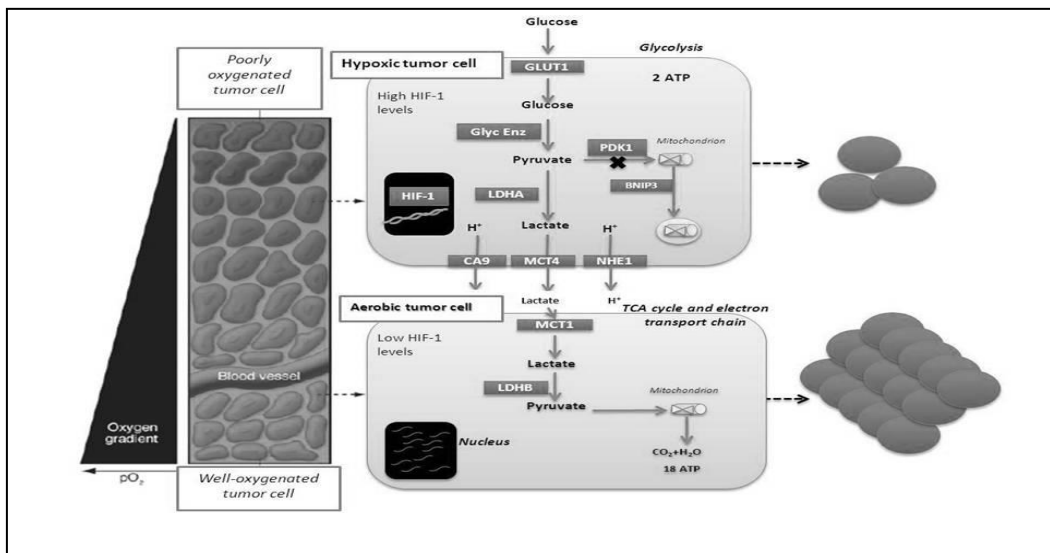


Figure 5. Tumor cells near to blood vessels are well oxygenated, whereas more distant tumor cells are poorly oxygenated and express high levels of HIF-1, which induces the expression of proteins that increase: uptake of glucose (glucose transporter 1 (GLUT1); conversion of glucose to pyruvate (glycolytic enzymes [Glyc. Enz.]); generation of lactate and H^+ (LDHA); and efflux of these molecules out of the cell (carbonic anhydrase IX (CA9), sodium-hydrogen exchanger 1 (NHE1), encoding the transporter 4 monocarboxylate (MCT4). In these hypoxic cells, two moles of lactate are produced for each mole of glucose, associated with a reduced substrate delivery to the mitochondria through the action of pyruvate dehydrogenase kinase 1 (PDK1). Hypoxic cells generate 2 mol of ATP and 2 mol of lactate for each mol of glucose consumed, whereas aerobic cells generate 36 mol of ATP per 2 mol of lactate consumed. Modified to [94].

HIF-1 stimulates the conversion of glucose to pyruvate and lactate by means of glucose transporter type 1 (GLUT1), hexokinase (HK2 and HK1), LDHA and monocarboxylate transporter type 4 (MCT4). In addition, HIF-1 reduces the conversion of pyruvate to acetyl-CoA by action of the pyruvate dehydrogenase (PDH) [79, 94]. HIF-1 cooperates with the proto-oncogene *MYC* to aerobic glycolysis promotion by induction of *HK2* and pyruvate dehydrogenase kinase 1 (*PDK1*) [29, 79].

Tumor cells show cyclical changes in phases of oxygenation, which involves a dynamic regulation of metabolic symbiosis, where cells can change from a state of lactate production to one of consumption of the same. Intratumoral hypoxia is associated with increased risk of metastasis and mortality [94] (Figure 5).

Metabolic Changes and Cancer Cells

Cancer cells differ from normal cells by a number of changes in their physiological processes, including autonomous proliferation, angiogenesis, apoptosis evasion, limitless replicative potential, tissue invasion or metastasis and immune response (Figure 6) [5, 29].

- **Autonomous proliferation:** Growth factors regulate two transduction signals in the Ras/Raf/MAP kinase pathway: ERK and PI3K. Both activate mTOR to stimulate cell growth. Most cancers have mutations in the main regulators of this pathway: K-Ras, H-Ras, N-Ras, B-Raf, subunit p110a of PI3K, receptor tyrosine kinase (RTKs) or its effectors downstream (Akt and PDK1), or inactivating mutations of negative regulators of these proteins [95]. Over activation of the PI3K/Akt system in autonomous cells provides increase in the flow of glucose and amino acids that may be attributable to the activation of HIF-1a. Akt stimulates the expression of GLUT1 and translocation to the plasma membrane of GLUT4. In addition, Akt stimulates glycolysis by activation of 6-phosphofructo-2-kinase (PFK2) and synthesis of fatty acids by phosphorylation of ATP citrate lyase [29].
- **Apoptosis evasion:** Alterations in the OXPHOS can induce resistance to apoptosis. Likewise, inhibition of the respiratory chain can inhibit the activation of proteins pro-apoptotic Bcl-2, Bax and Bak [96]. In studies published by Bonnet et al. [97], the pharmacological inhibition of PDK1, an enzyme that catalyzes the phosphorylation and inactivation of PDH, produces the reactivation of PDH, induces apoptosis in cancer cells, which is an interesting example of reversal of resistance to apoptosis and metabolic reprogramming [29].
- **Continuous replication:** To immortalize the replicative power, tumor cells often mutate or inactivate inducers of senescence as the tumor suppressor p53 [98]. In glycolysis, enzyme phosphoglycerate mutase (PGM), which is negatively regulated by p53, catalyzes the conversion of 3-phosphoglycerate (3PG) to 2-phosphoglycerate (2PG) [99].
- **Angiogenesis:** In response to hypoxia and HIF-1, many tumors over-express the vascular endothelial growth factor (VEGF) by activation of the ERK and PI3K signaling pathways [29, 100]. On the other hand, mitochondrial protein F1F0-ATPase, whose function is to provide the necessary protons for aerobic glycolysis, is inhibited by angiostatin, which prevents the endogenous angiogenesis by intracellular acidification [29]. Chi et al. [101], reported an antibody similar to angiostatin with angiogenesis inhibitor effect.
- **Tissue invasion and metastasis:** E-cadherin is involved in the intercellular union [29]. Activation of HIF-1a causes loss of E-cadherin and expression of proto-oncogenes *met* and *Twist* that induce metastasis through the chemokines receptor (CXCR4) and lysyl

- Immune response: The antitumor effects of cytotoxic T lymphocytes (CTLs) can be inhibited by the metabolic micro-environment of the tumor cells [29]. The macrophages associated with tumors are markers of poor prognosis: facilitate angiogenesis as well as promotion and migration of tumor cells [104]. In advanced stage cancer, the production of lactate inhibits the production of cytokines, the CTLs cytolytic activity and the cell proliferation. Alternative therapies modifying the tumor metabolism have been proposed inhibiting the protons exportation and the production of lactate or indolamine 2, 3-dioxygenase, which could restore the defective antitumor immune response [29].

The diagram illustrates the Hallmarks of Cancer, centered around ten key features arranged in a circle:

- Autonomous proliferation**: Influenced by Akt, PI3K, mTOR, and Anabolic reactions.
- Immune response**: Influenced by Kynurenins.
- Insensitivity to antigrowth**: Influenced by Lactate, Anaerobic glycolysis, Oxygen independence, and Proton extrusion.
- Tissue invasion and metastasis**: Influenced by Extracellular acidification and HIF-1.
- Continuous replication**: Influenced by Inhibited OXPHOS, p53, PGM, Glycolysis, and F₁F₀ATPase.
- Angiogenesis**: Influenced by Ang-2, VEGF, and HIF-1.
- Apoptosis evasion**: Influenced by HK-VDAC association and Inhibited OXPHOS.

Additional pathways shown include:

- Glycolysis leading to Inhibited OXPHOS.
- Akt leading to PI3K, which leads to mTOR.
- mTOR leading to Anabolic reactions.
- Lactate being produced from Anaerobic glycolysis and influencing Proton extrusion.
- Oxygen independence leading to Anaerobic glycolysis.
- p53 leading to Inhibited OXPHOS and Glycolysis.
- HIF-1 leading to Ang-2, VEGF, and Extracellular acidification.

Figure 6. Cancer cell changes and their links to tumor metabolism. Hypothetical links between different metabolic alterations and the non-metabolic characteristics of neoplasia [circle] are depicted. Centripetal arrows indicate how the changes in cancer cells can impinge on metabolism. Centrifugal arrows illustrate how neoplasia-associated metabolic reprogramming can contribute to the acquisition of cancer changes. Ang-2: angiopoietin-2; GLUT: glucose transporter; HIF: hypoxia-inducible factor; HK: hexokinase; OXPHOS: oxidative phosphorylation; PGM: phosphoglycerate mutase; PI3K: phosphatidylinositol 3-kinase; SCO2: synthesis of cytochrome c oxidase 2; VDAC: voltage-dependent anion channel; VEGF: vascular endothelial growth factor. Modified to [29].

Table 5. Metabolic pathways in cancer as therapeutic target [29]

Target	Desired Effects	Examples of Compounds
Glycolysis		
Glucose uptake	Glucose transport or initial glycolysis inhibition.	2-Deoxyglucose has radiosensitizing and chemosensitizing effects
Hexokinase [HK1 and HK2]	Inhibition of enzymatic activity and dissociation from mitochondria	3-Bromopyruvate has potent antitumor effects in vitro and in vivo
Pyruvate dehydrogenase kinase 1 [PDK1]	Inhibition of PDK1 for deinhibition of pyruvate dehydrogenase	Dichloroacetate [DCA]
Lactate dehydrogenase A [LDHA]	Inhibition	SiRNA
Pyruvate kinase [PK] isoenzyme PKM2	Translocation of PKM2 to the nucleus for induction of apoptosis	Somatostatin and its derivative TT-232 [in vitro]
Fatty Acid Synthesis		
ATP citrate lyase [ACL]	Inhibition	SB-2049990 inhibits pancreatic cancer growth in nude mice
Acetyl-CoA Carboxylase [ACC]	Inhibition	Sorafenib A induces apoptosis or autophagy in vitro
Fatty acid Synthase [FASN]	Inhibition	Cerulenin and its derivative C57 inhibit human ovarian cancer
Choline kinase [ChoK]	Inhibition	MN58b reduces phosphomonoesters in human cancer xenografts
HIF		
HIF-1 α prolylhydroxylases [PHDs]	Activation of PHDs for inhibition of HIF	Cell-permeating α -ketoglutarate derivatives reverse HIF activation in SDH- or FH-deficient cells
Hypoxia-inducible factor 1 [HIF-1]	Inhibition of DNA binding	Echinomycin
Reactive oxygen species [ROS]	Antioxidants neutralize ROS and reduce HIF-1 function via PDHs and VHL	N-acetylcysteine [NAC]; vitamin C
Hypoxia	Cytotoxic effects of components enriched in hypoxic cells	Tirapazamine [TPZ], a hypoxia activated prodrug.
Proton Extrusion		
Na ⁺ /H ⁺ exchanger	Inhibition	Cariporide
Bicarbonate/Cl ⁻ exchanger	Inhibition	S3705
MCT1 lactate /H ⁺ symporter	Inhibition	α -cyano-4-OH-cinnamate
Carbonic anhydrases 9 and 12 [CA9 - CA12]	Inhibition	Sulfonamide indisulam
F ₁ F ₀ ATP synthase	Inhibition	Angiostatin; antibodies
Other		
AMPK	Activation	Biguanides and thiazolidinediones activate AMPK through inhibition of OXPHOS, reducing the risk of cancer in diabetic patients
eIF4E	Inhibition of translation initiation by eIF4E	Antisense oligonucleotide inhibits growth of human breast cancer
L-type amino acid transporter 1 [LAT1]	Inhibition to reduce amino acid transport	2-aminobicyclo [2.2.1]-heptane 2-carboxylic acid inhibits tumor growth in a xenograft model

CANCER INITIATION AND PROGRESSION

Cancer is a very heterogeneous disease. Tumors that arise in the same tissue can even exhibit an array of cellular pathologies, ranging from benign hyperplasia to highly invasive malignancies. On the other hand, cancer is a complex disease which involves the deregulation of multiple pathways. Microarray analysis has identified thousands of genes that are transcriptionally deregulated in cancer [107]; however, it remains unclear which deregulated genes play a causative role in tumor initiation and maintenance and which ones represent passerby with no selective advantage. Regardless of the role that specific genes may play in cancer progression, when a tumor is histological or clinically identified, a large number of molecular lesions have been accumulated [108].

A literature survey on all published cancer genes identified 291 genes for which there were molecularly characterized mutations and evidence of a causative role in tumorigenesis; these genes represent more than 1 percent of the human genome [109]. The large number of mutations found in tumor samples raises the question of whether there exists a common thread that underlies most, if not all, human malignancies, or biological rules that govern cancer initiation and progression. Despite the heterogeneity observed in cancer, most tumors share certain characteristics: self-sufficiency in growth signals, insensitivity to anti-growth signals, evasion of apoptosis, acquisition of a limitless replicative potential, sustained angiogenesis, and tissue invasion and metastasis [5, 108]. The similarity in the cellular processes in all cancer cells, regardless of their tissue of origin, likely indicates common tumor-initiation mechanisms.

When chronic myelogenous leukemia converts to acute leukemia, the malignant clone acquires an additional t(9;22) translocation, an isochromosome 17, or trisomy of chromosome 8. Likewise, when follicular lymphoma becomes aggressive, the lymphoma cells often bear a t(8;14) translocation in addition to the original t(14;18) translocation [1]. These findings support the hypothesis that most hematopoietic tumors and soft-tissue sarcomas are initiated by the activation of an oncogene, followed by alterations in tumor-suppressor genes and other oncogenes. In contrast, most carcinomas are initiated by the loss of function of a tumor-suppressor gene, followed by alterations in oncogenes and additional tumor-suppressor genes [110].

Comparative genomic hybridization has revealed a number of genes that can be amplified or deleted in cancer. Breast cancer genome analyses indicate that there are only a few genes that are frequently mutated but many that are rarely mutated, providing an explanation for the heterogeneity in cancer. Complex somatic DNA rearrangements, mostly intra-chromosomal duplications, have been found in the breast cancer genomes [111].

Genome sequences of primary breast cancer and cancer metastasis show restricted de novo mutations in metastasis, but significantly shared mutations in the primary tumor. These studies provide information on tumor evolution and identify pathways critical to breast cancer metastasis [112].

The *ErbB2*, *PI3KCA*, *MYC*, and *CCND1* oncogenes are frequently deregulated in breast cancer. Loss-of-function mutations of *RB* in breast cancer cell lines and primary tumors were reported since 1988 [113]. At least ten tumor suppressor genes, all of which are involved in the regulation of genomic integrity, have been associated with hereditary breast cancer [114].

Cellular transformation is thought to take place through the accumulation of mutations, as well as epigenetic changes, that activate oncogenes or down regulate tumor-suppressor genes

and lead to uncontrolled clonal expansion. Oncogenes originally were identified as the transforming agents of tumor viruses; later oncogenes were recognized as mutated versions of normal cellular genes, or proto-oncogenes, which had been incorporated into the viral genome by recombination [115].

The positive control of cellular growth is associated with mutations in oncogenes, which are generally dominant. In contrast, tumor suppressor genes function as negative regulators of cellular growth and are generally recessive. Thus, inactivation of both copies of a tumor-suppressor gene is usually necessary for tumor development. Several lines of evidence indicate that mutations in a single oncogene or tumor suppressor are insufficient to give rise to cancer [115]. First, most cancers develop late in life and the incidence of disease increases dramatically with age. Statistical analysis of epidemiological data shows that a cell needs to accumulate four to five sequential genetic lesions in key regulatory pathways in order to become malignant [116]. Second, *in vitro* experiments using cell lines, as well as *in vivo* models of cancer, confirm the multiple-hit hypothesis (retinoblastoma and certain types of leukemia are exceptions). Experiments using retroviruses containing activated versions of growth-controlling genes could lead to transformation of rodent cells in culture [117, 118]. However, the cells used in these initial cancer studies were immortal and could therefore proliferate indefinitely. When these experiments were repeated with primary cell lines, it was found that activation of at least one pair of oncogenes was required for transformation [119].

Tumors are histologically classified as presenting with different “grades,” which correspond to a set of physiological markers as loss of differentiation, abnormal ploidy, and morphology, and correlate with patient outcome. Higher grade tumors have a more negative prognosis, while low-grade tumors are often considered early lesions and may progress to more invasive, high-grade disease. These observations have led to the hypothesis that cancer progression can be dissected into a small number of crucial steps whose sequential deregulation is critical for the clinical progression from low- to high-grade cancer. At least four pathways must be altered in order for tumor progression to occur: maintenance of telomere length (achieved by expressing human telomerase), deregulated cell-cycle entry (inactivation of *Rb*), deregulated cell growth arrest and apoptosis (inactivation of *p53*), and growth-factor independence (by oncogenic *Ras* overexpression). It remains to be explored how different oncogenes and tumor-suppressors found in tumor samples contribute to these cancer pathways and how they interact with each other to reinforce their tumorigenic potential [120].

Multistep process observed in human cancer has also been found in mouse models carrying activated oncogenes or inactivated tumor-suppressor genes, in which the duration and aggressiveness of the disease can be changed by introducing into the mouse genome the same sequential genetic alterations observed in human tumors.

ONCOGENES AS THERAPEUTIC TARGETS

Cancer mouse models have been used to test the hypothesis that, if tumor growth remains dependent on their original transforming oncogenic mutations, oncogene inactivation could lead to tumor regression [121]. Several studies evaluating the over expression of the *MYC* oncogene in lymphoid and epidermal tissues showed that the inactivation of *MYC* led to sustained tumor regression with concomitant promotion of either differentiation or apoptosis

[121-123]. However, in other models, a fraction of tumor cells were refractory to *MYC* inactivation; these cells presumably had acquired new mutations that allowed Myc-independent growth [124-126].

The main cause of treatment failure for cancer patients are the metastasis. According to the model of metastatic progression only a small subset of cells from the primary tumor acquire the requisite mutations to metastasize to distant sites, where new mutations are accumulated as a response to the different selective pressures of a novel environment [127]. However, recent data suggest that most cells in primary tumors with metastatic potential already contain the lesions necessary for metastasis and, possibly, for survival in a foreign environment. Microarray analysis compared patterns of gene expression in breast cancer patients with their known five-year survival and recurrence rates. Seventy genes were identified that could predict clinical outcome with a combined 83 percent accuracy [128]. Moreover, solid tumors of different origin shared the same metastatic signature, implying there is a common set of molecules regulating metastasis in a variety of primary tumors [129, 130].

Oncogenic proteins in cancer cells can be targeted by monoclonal antibodies when expressed on the cell surface, or by small molecules acting on specific molecules in particular metabolic ways. For example, imatinib targets the initial step of the multistep process in chronic myelogenous leukemia [131]. The same drug can affect the KIT and PDGFR receptor kinases. [132, 133]. Of particular interest are inhibitors of the BCL2 family, which can induce the apoptotic death of cancer cells. In acute promyelocytic leukemia, which is initiated by a t(15;17) chromosome translocation that fuses the *PML* gene to *RAR α* (a nuclear receptor for retinoic acid), addition of retinoic acid can induce terminal differentiation and death of APL cells. This modality is called differentiation therapy [134, 135].

MicroRNA Genes

Recent findings in molecular biology show that the participation of small regulatory non-coding RNA is required for cellular diversity in developing and mature organisms. These molecules act as sequence-specific post-transcriptional regulators in the expression of other RNA transcripts [136]. Small regulatory non-coding RNA molecules include microRNAs (miRNAs), small interfering RNAs (siRNAs) and repeat-associated siRNAs (rasiRNAs), which are unified by their association with the argonaute (AGO) family of proteins and by their function. All these RNAs direct the binding of protein complexes to specific nucleic acid sequences [137- 139].

Unlike other genes involved in cancer, miRNAs do not encode proteins. Their products are a single RNA strand of about 21 to 23 nucleotides implicated in to regulate the gene expression. A miRNA molecule can anneal to a messenger RNA (mRNA) containing a complementary nucleotide sequence and blocking the protein translation or causing degradation of the mRNA [140]. Mapping of numerous miRNA genes has shown that many of them occur in chromosomal regions that undergo rearrangements, deletions, and amplifications in cancer cells. The genome regions that are consistently involved in chromosomal rearrangements in cancer cells but that lack oncogenes or tumor-suppressor genes appear to harbor miRNA genes [141].

Recent studies have shown that the RNA Pol III drives miRNA transcription from dense human clusters interspersed among repetitive Alu elements. These gene clusters are transcribed

as polycistronic primary transcripts and subsequently cleaved into multiple miRNAs, or from intergenic regions as independent transcriptional units, or from intronic sequences of protein-coding or non-coding transcription units or exonic sequences of non-coding genes [142] (Figure 7).

Primary miRNAs transiently receive a 7-methylguanosine (7mGpppG) cap and a poly(A) tail and is processed into a precursor miRNA (pre-miRNA) by the nuclear RNase III enzyme Drosha and its partner DiGeorge syndrome critical region gene 8 (DGCR8). Exportin-5 transports pre-miRNA into the cytosol, where it is processed by the Dicer RNaseIII enzyme into a mature double strand miRNAs. The RNA strand is recruited into the RNA-induced silencing (RISC) effector complex and assembled through processes that are dependent on Dicer and other double strand RNA binding domain proteins, as well as on members of the argonaute family. MiRNAs guide the RISC complex to the 3' untranslated regions (3'-UTR) of the complementary messenger RNA (mRNA) targets for repress their expression by several mechanisms: repression of mRNA translation, destabilization of mRNA transcripts through cleavage, deadenylation, and localization in the processing body (P-body), where the miRNA-targeted mRNA can be sequestered from the translational machinery and degraded or stored for subsequent use. Nuclear localization of mature miRNAs has been described as a novel mechanism of action for miRNAs. Scissors indicate the cleavage on pri-miRNA or mRNA [143].

Expression profiling of miRNA genes has revealed signatures associated with tumor classification, diagnosis, staging, and progression, as well as prognosis and response to treatment. For example, miRNA expression profiling can distinguish between indolent and aggressive forms of chronic lymphocytic leukemia, and expression of a small panel of miRNA genes correlates with prognosis in lung cancer [144- 146].

Some miRNA genes that are deregulated in chronic lymphocytic leukemia have germ-line or somatic mutations in a miRNA precursor that affect the processing of short single-stranded miRNA molecules [144]. MiRNA genes can be up-regulated or down-regulated in cancer cells. The up-regulated miRNA genes function as oncogenes by down-regulating tumor-suppressor genes, whereas the down-regulated miRNA genes function as tumor-suppressor genes by down-regulating oncogenes [147]. Table 6 display a number of examples of up or down regulated miRNAs in cancer.

The function of miRNA genes depends on their targets in a specific tissue. A miRNA gene can be a tumor suppressor if its critical target is an oncogene and it can be an oncogene if its target is a tumor-suppressor gene. Up-regulation of miRNA genes can be due to amplification, deregulation of a transcription factor, or demethylation of CpG islands in the promoter regions of the gene. For example, the ALL1 (MLL) fusion proteins of acute lymphoblastic leukemia or acute myeloblastic leukemia carrying chromosome 11q23 translocations target the Drosha nuclease complex to specific miRNA genes, including *miR191*, thereby enhancing the processing of their miRNA precursors [148]. The *miR191* gene is also up-regulated in numerous types of solid cancers [149], suggesting that it is the downstream target of signal-translocation pathways involved in cancer. MiRNA genes functioning as tumor suppressors can be down-regulated because of deletions, epigenetic silencing, or loss of the expression of one or more transcription factors [150].

The *miR155* gene is over expressed in diffuse large B-cell lymphoma, the aggressive form of chronic lymphocytic leukemia, and in breast, lung, and colon cancers [151]. In transgenic mice under control of the E μ enhancer of the immunoglobulin genes, over expression of

miR155 causes acute lymphoblastic leukemia or high-grade lymphoma, indicating that deregulation of a single miRNA gene can cause malignant transformation. Since it takes several months for the tumors in these mice to become aggressive, it is likely that additional genetic alterations are needed for the development of truthful neoplasia [152].

The *LET7* miRNA family, which are deleted or under expressed in lung cancer, target *RAS*; loss of *LET7* results in overexpression of *RAS* [153]. *MIR15a* and *miR-16-1*, the microRNAs those are deleted or down-regulated in chronic lymphocytic leukemia, cause overexpression of *BCL2*, which protects cells from apoptosis [154].

The expression of a set of 21 miRNAs is altered in at least three types of solid tumors [149]. Among these, *miR21* is of particular interest because inhibits expression of the tumor suppressor PTEN, which encodes a phosphatase involved in the PI3K kinase signaling pathway and is mutated in advanced breast, lung, gastric, and prostate cancers [155].

Table 6. MiRNAs aberrantly expressed in cancers [146]

Cancer type	Up regulated	Down regulated
Breast cancer	miR-10b, miR-21, miR-22, miR-27a, miR-155, miR-210, miR-221, miR-222, miR-328, miR-373, miR-520c	let-7, miR-7, miR-9-1, miR-17/miR-20, miR-31, miR-125a, miR-125b, miR-146, miR-200 family, miR-205, miR-206, miR-335
Chronic lymphocytic leukemia	miR-21, miR-155	miR-15, miR-16, miR-29b, miR-29c, miR-34a, miR-143, miR-145, miR-181b, miR-223
Lung cancer	miR-17-92 cluster, miR-21, miR-106a, miR-155	miR-1, let-7 family, miR-7, miR-15a/miR-16, miR-29 family
Lymphoma	miR-17-92 cluster, miR-155	miR-143, miR-145
Prostate cancer	miR-221, miR-222	miR-15a-miR-16-1 cluster, miR-101, miR-127, miR-449 ^a
Glioblastoma	miR-21, miR-221, miR-222	miR-7
Hepatocellular carcinoma	miR-17-92 cluster, miR-21, miR-143, miR-224	miR-1, miR-101, miR-122a
Colorectal cancer	miR-17-92 cluster, miR-21	miR-34a, miR-34b/c, miR-127, miR-143, miR-145, miR-342
Gastric cancer	miR-21, miR-27 ^a	miR-143, miR-145
Ovarian cancer	miR-214	miR-34b/c, miR-200 family
Melanoma	miR-221, miR-222	let-7 ^a , miR-34 ^a
Head and neck squamous cell carcinoma	miR-21	let-7d, miR-138, miR-205

Primary miRNAs (pri-miRNAs) receive a 7-methylguanosine (7mGpppG) cap and a poly(A) tail. The pri-miRNA is processed into a precursor miRNA (pre-miRNA) and exported by the protein exportin-5 into cytosol, where is processed by the Dicer RNaseIII enzyme, into 22 nt-long double strand miRNAs. Then, miRNAs guide the RISC complex to the untranslated regions of the mRNA targets for repressing their expression.

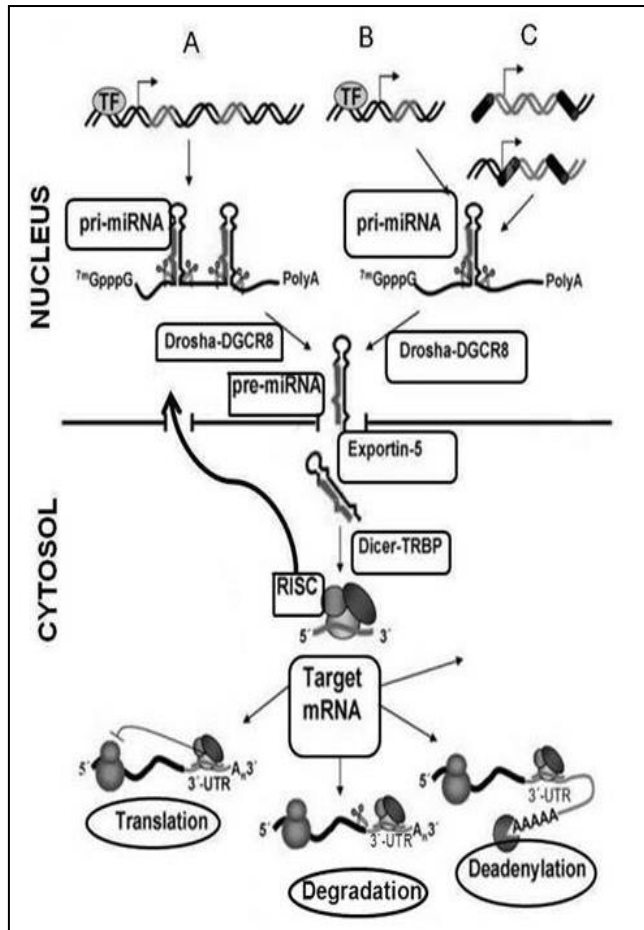


Figure 7. On the top line: (A) Gene clusters giving polycistronic primary transcripts and cleaved into multiple miRNAs; (B) intergenic regions transcribed as independent transcriptional units; (C) intronic sequences (in grey) of non-coding transcription units or exonic sequences (black cylinders) of non-coding genes. TF: transcription factor.

REFERENCES

- [1] Croce CM. Oncogenes and Cancer. *New Engl. J. Med.* 2008 Jan; 358(5):502-511.
- [2] Carmeliet P, Jain RK. Angiogenesis in cancer and others diseases. *Nature.* 2000 Sep; 407(6801):249-257.
- [3] Bergers G, Benjamin LE. Tumorigenesis and the angiogenic switch. *Nat. Rev. Cancer.* 2003 Jun; 3(6):401-410.

- [4] Bindea G, Mlecnik B, Fridman WH, Pagès F, Galon J. Natural immunity to cancer in humans. *Curr. Opin. immunol.* 2010 Apr; 22(2):215-222.
- [5] Hanahan D, Weinberg RA. Hallmarks of cancer: the next generation. *Cell.* 2011 Mar 4; 144(5):646-674.
- [6] Morag P. Oncogenes. En: *The metabolic basis of inherited disease*: Scriver CR, Beaudet AL, Sly WS, Valle D. Ed. Mc Graw-Hill. Inc. 7ma. Edition New York. 2001; pp 645-644.
- [7] Rous P. A sarcoma of the fowl transmissible by an agent separable from the tumor cells. *J. Exp. Med.* 1911 Apr; 13:397–411.
- [8] Temin HM. On the origin of genes for neoplasia: Clowes Memorial Lecture. *Cancer Res.* 1974 Nov; 34(11):2835-41.
- [9] Stehlin D, Varmus H, Bishop J, Vogt P. DNA related to the transforming genes of ovarian sarcoma viruses is present in normal avian. *Nature.* 1976 Mar; 260 (5547):170-3.
- [10] Bishop JM. Enemies within: The genesis of retrovirus oncogenes. *Cell.* 1981 Jan; 23(1):5-6.
- [11] Aaronson SA. Growth factors and cancer. *Science.* 1991 Nov; 254(5035):1146-53.
- [12] Butel JS. Viral carcinogenesis: revelation of molecular mechanism and etiology of human disease. *Carcinogenesis.* 2000(21):405-426.
- [13] Hanawalt P, Sarasin A. Cancer prone hereditary diseases with DNA processing abnormalities. *Trends Genet.* 1986; 2:124.
- [14] Vogelstein B, Fearon ER, Scott EK, Hamilton SR, Preisinger AC, Nakamura Y, et al. Allotype of colorectal carcinomas. *Science.* 1989 Apr; 244 (4901):207-11.
- [15] Leach FSC, Nicolaides NC, Papadopoulos N, Liu B, Jen J, Parsons R. et al. Mutations of a mutS homolog in hereditary non-polyposis colorectal cancer. *Cell.* 1993 Dec; 75 (6):1215-25.
- [16] Fishel R, Lescoe M, Rao MR, Copeland N, Jenkins NA, et al. The human mutator gene homolog MSH2 and its association with hereditary non polyposis colon cancer. *Cell.* 1993 Dec; 75 (5):1027-38.
- [17] Kinzler KW, Vogelstein B. Lessons from hereditary colorectal cancer. *Cell.* 1996 Oct; 87(2):159-70.
- [18] Konopka JB, Watanabe SM, Singer JW, Collins SJ, Witte ON. Cell lines and clinical isolates derived from Ph1 positive chronic myelogenous leukemia patients express c-abl proteins with a common structural alteration. *Proc. Natl. Acad. Sci. USA.* 1985 Mar; 82(6):1810-1814.
- [19] Rakoff-Nahoum S. Why Cancer and Inflammation? *Yale J. Biol. Med.* 2006 Dec; 79 (3-4):123-130.
- [20] Tsujimoto Y, Gorham J, Cossman J, Jaffe E, Croce CM. The t(14, 18) chromosome translocations involved in B-cell neoplasm result from mistakes in VDJ joining. *Science.* 1985 Sep; 229(4720):1390-1393.
- [21] Finger LR, Harvey RC, Moore RC, Showe LC, Croce CM. A common mechanism of chromosomal translocation in T and B cell neoplasia. *Science.* 1986 Nov; 234(4779):982-985.
- [22] Meza-Junco J, Montaña-Loza A, Aguayo-González A. Bases moleculares del cáncer. *Rev. Invest. Clin.* 2006 Jan-Feb; 58(1); 56-70.
- [23] Bishop JM. Molecular themes in oncogenesis. *Cell.* 1991 Jan 25; 64(2):235-248.

- [24] Tomlins SA, Rhodes DR, Perner S, Dhanasekaran SM, Mehra R, Sun XW, et al. Recurrent fusion of TMPRSS2 and ETS transcription factor genes in prostate cancer. *Science*. 2005 Oct; 310:644-8.
- [25] Rodenhuis S. ras and human tumors. *Semin Cancer Biol*. 1992 Aug; 3(4):241-247.
- [26] Beaupre DM, Kurzrock R. RAS and leukemia: from basic mechanisms to gene-directed therapy. *J. Clin. Oncol*. 1999 Mar; 17(3):1071-1079.
- [27] Davies H, Bignell GR, Cox C, Stephens P, Edkins S, Clegg S, et al. Mutations of the BRAF gene in human cancer. *Nature* 2002 Jun; 417(6892):949-54.
- [28] Frattini M, Ferrario C, Bressan P, Balestra D, De Cecco L, Mondellini P, et al. Alternative mutations of BRAF, RET and NTRK1 are associated with similar but distinct gene expression patterns in papillary thyroid cancer. *Oncogene*. 2004 Sep; 23(44): 7436-40.
- [29] Kroemer G & Pouyssegur J. Tumor Cell Metabolism: Cancer's Achilles' Heel. *Cell Metabolism*. 2008 Jun;13(6), 472-82.
- [30] Heldin CH, Westermark B. Mechanisms of activation and in vivo role of platelet-derived growth factor. *Physiol. Rev*. 1999 Oct; 79(4):1283-1316.
- [31] Giles RH, van Es JH, Clevers H. Caught up in a Wnt storm: wnt signaling in cancer. *Biochim. Biophys. Acta*. 2003 Jun; 1653(1):1-24.
- [32] MacDonald BT, Tamai K, He X. Wnt/beta-catenin signaling: components, mechanisms and diseases. *Dev. Cell*. 2009 Jul; 17(1):9-26.
- [33] Valkenburg KC, Williams BO. Mouse models of prostate cancer. *Prostate Cancer*. 2011;2011:895238.
- [34] Heldin CH. Dimerization of cell surface receptors in signal transduction. *Cell*. 1995 Jan; 80(2):213-223.
- [35] Arteaga CL. Epidermal growth factor receptor dependence in human tumors: more than just expression?. *Oncologist*. 2002; 7 Suppl 4:31-39.
- [36] Schiller MR. Coupling receptor tyrosine kinases to Rho GTPases-GEFs what's the link. *Cell signal*. 2006 Nov; 18(11):1834-1843.
- [37] Joensuu H, Dimitrijevic S. Tyrosine Kinase inhibitor imatinib (STI571) as an anticancer agent for solid tumours. *Ann. Med*. 2001 Oct; 33(7):451-455.
- [38] Tamborini E, Bonadiman L, Greco A, Albertini V, Negri T, Gronchi A et al. A new mutation in the KIT ATP pocket causes acquired resistance to imatinib in a gastrointestinal stromal tumor patient. *Gastroenterology*. 2004 Jul; 127(1):294-299.
- [39] Shaulian E, Karin M. AP-1 in cell proliferation and survival. *Oncogene*. 2001 Apr; 20(19); 2390-2400.
- [40] Oikawa T, Yamada T. Molecular biology of the Ets family of transcription factors. *Gene* 2003 Jan;16(303):11-34.
- [41] Tomlins SA, Rhodes DR, Perner S, Dhanasekaran SM, Mehra R, Sun XW, et al. Recurrent fusion of TMPRSS2 and ETS transcription factor genes in prostate cancer. *Science*. 2005 Oct; 310(5748):644-648.
- [42] Peterson CL, Workman JL. Promoter targeting and chromatin remodeling by SWI/SNF complex. *Curr. Opin. Genet. Dev*. 2000 Apr; 10(2):187-192.
- [43] Jenuwein T, Allis CD. Translating the histone code. *Science*. 2001 Aug; 293(5532):1074-1080.
- [44] Strahl BD, Allis CD. (2000). The language of covalent histone modifications. *Nature*. 2000 Jan; 403(6765):41-45.

- [45] Nakamura T, Mori T, Tada S, Krajewski W, Rozovskaia T, Wassell R, et al. ALL-1 is a histone methyltransferase that assembles a supercomplex of proteins involved in transcriptional regulation. *Mol Cell*. 2002 Nov; 10(5):1119-1128.
- [46] Pawson T, Warner N. Oncogenic re-wiring of cellular signaling pathways. *Oncogene*. 2007 Feb; 26(9):1268-1275.
- [47] Sicheri F, Moarefi I, Kuriyan J. Crystal structure of the Src family tyrosine kinase Hck. *Nature*. 1997 Feb; 385(6617):602-609.
- [48] Kaziro Y, Itoh H, Kozasa T, Nakafuku M, Satoh T. Structure and function of signal-transducing GTP-binding proteins. *Annu. Rev. Biochem*. 1991; 60:349-400.
- [49] Salgia R, Skarin AT. Molecular abnormalities in lung cancer. *J. Clin. Oncol*. 1998 Mar; 16(3):1207-1217.
- [50] Tsujimoto Y, Yunis J, Onorato-Showe L, Nowell PC, Croce CM. Molecular cloning of the chromosomal breakpoints of the B-cell lymphomas and leukemias with the t(11;14) chromosome translocation. *Science*. 1984 Jun; 224(4656):1403-1406.
- [51] Tsujimoto Y, Cossman J, Jaffe E, Croce CM. Involvement of the bcl-2 gene in human follicular lymphoma. *Science*. 1985 Jun; 228(4706):1440-1443.
- [52] Tsujimoto Y, Croce CM. Analysis of the structure, transcripts, and protein products of bcl-2, the gene involved in human follicular lymphoma. *Proc. Natl. Acad. Sci. USA*. 1986 Jul; 83(14):5214-5218.
- [53] Tsujimoto Y, Ikegaki N, Croce CM. Characterization of the protein product of bcl-2, the gene involved in human follicular lymphoma. *Oncogene*. 1987; 2(1):3-7.
- [54] Adams JM, Cory S. The Bcl-2 apoptotic switch in cancer development and therapy. *Oncogene*. 2007 Feb; 26(9):1324-1337.
- [55] Alt FW, Kellems RE, Bertino JR, Schimke RT. Selective multiplication of dihydrofolate reductase genes in methotrexate-resistant variants of cultured murine cells. *J. Biol. Chem*. 1978 Mar; 253(5):1357-70.
- [56] Cowell JK. Double minutes and homogeneously staining regions: gene amplification in mammalian cells. *Annu. Rev. Genet*. 1982; 16:21-59.
- [57] King CR, Kraus MH, Aaronson SA. Amplification of a novel v-erbB-related gene in a human mammary carcinoma. *Science*. 1985 Sep; 229(4717):974-6.
- [58] Schwab M, Alitalo K, Klempnauer KH, Varmus HE, Bishop JM, Gilbert F, et al. Amplified DNA with limited homology to myc cellular oncogene is shared by human neuroblastoma cell lines and a neuroblastoma tumour. *Nature*. 1983 Sep; 305(5931):245-8.
- [59] Press MF, Bernstein L, Thomas PA, Meisner LF, Zhou JY, Ma Y, et al. HER-2/neu gene amplification characterized by fluorescence in situ hybridization: poor prognosis in node-negative breast carcinomas. *J. Clin. Oncol*. 1997 Aug; 15(8): 2894-904.
- [60] Eiró N, Vizoso FJ. Inflammation and cancer. *World J. Gastrointest. Surg*. 2012 Mar; 4(3); 62-72.
- [61] Lorusso G, Rüegg C. The tumor microenvironment and its contribution to tumor evolution toward metastasis. *Histochem. Cell Biol*. 2008 Dec; 130(6): 1091-1103.
- [62] Kuper H, Adami HO, Trichopoulos D. Infections as a major preventable cause of human cancer. *J. Intern. Med*. 2000 Sep; 248(3):171-83.
- [63] Balkwill F, Mantovani A. Inflammation and cancer: back to Virchow? *Lancet*. 2002 Feb; 357(9255):539-45.
- [64] Coussens LM, Werb Z. Inflammation and cancer. *Nature*. 2002 Dec; 420(6917):860-67.

- [65] Munger K, Howley PM. Human papillomavirus immortalization and transformation functions. *Virus Res.* 2002 Nov; 89(2):213-28.
- [66] Peek RM Jr, Blaser MJ. Helicobacter pylori and gastrointestinal tract adenocarcinomas. *Nat. Rev. Cancer.* 2002 Jan; 2(1):28-37.
- [67] Hussain SP, Hofseth LJ, Harris CC. Radical causes of cancer. *Nat. Rev. Cancer.* 2003 Apr; 3(4):276-85.
- [68] Chen LW, Egan L, Li ZW, Greten FR, Kagnoff MF, Karin M. The two faces of IKK and NF-kappaB inhibition: prevention of systemic inflammation but increased local injury following intestinal ischemia-reperfusion. *Nat. Med.* 2003 May; 9 (5):575-81.
- [69] Wang D, Mann JR, DuBois RN. The role of prostaglandins and other eicosanoids in the gastrointestinal tract. *Gastroenterology.* 2005 May; 128 (5):1445-61.
- [70] Dolberg DS, Hollingsworth R, Hertle M, Bissell MJ. Wounding and its role in RSV mediated tumor formation. *Science.* 1985 Nov; 230 (4726):676-8.
- [71] Martins-Green M, Boudreau N, Bissell MJ. Inflammation is responsible for the development of wound-induced tumors in chickens infected with Rous sarcoma virus. *Cancer Res.* 1994 Aug; 54(16):4334-41.
- [72] Beachy PA, Karhadkar SS, Berman DM. Tissue repair and stem cell renewal in carcinogenesis. *Nature.* 2004 Nov; 432(7015):324-31.
- [73] Houchen CW, Stenson WF, Cohn SM. Disruption of cyclooxygenase-1 gene results in an impaired response to radiation injury. *Am. J. Physiol. Gastrointest. Liver Physiol.* 2000 Nov; 279 (5): G858-65.
- [74] Greten FR, Eckmann L, Greten TF, Park JM, Li ZW, Egan LJ, et al. IKKbeta links inflammation and tumorigenesis in a mouse model of colitis-associated cancer. *Cell.* 2004 Aug; 118(3):285-96.
- [75] Elewaut D, Di Donato JA, Kim JM, Truong F, Eckmann L, Kagnoff MF. NF-kappa B is a central regulator of the intestinal epithelial cell innate immune response induced by infection with enteroinvasive bacteria. *J. Immunol.* 1999 Aug; 163(3):1457-66.
- [76] Egan LJ, Eckman L, Greten FR, Chae S, Li ZW, Myhre GM, et al. IkappaB-kinasebetadependent NF-kappaB activation provides radioprotection to the intestinal epithelium. *Proc. Natl. Acad. Sci. USA.* 2004 Feb; 101(8):2452-7.
- [77] Chae S, Eckmann L, Miyamoto Y, Pothoulakis C, Karin M, Kagnoff MF. Epithelial cell I kappa B-kinase beta has an important protective role in Clostridium difficile toxin A-induced mucosal injury. *J. Immunol.* 2006 Jul; 177(2):1214-20.
- [78] Moore RJ, Owens DM, Stamp G, Arnott C, Burke F, East N, et al. Mice deficient in tumor necrosis factor-alpha are resistant to skin carcinogenesis. *Nat. Med.* 1999 Jul; 5(7):828-31.
- [79] DeBerardinis RJ, Lum JJ, Hatzivassiliou G, Thompson CB. The biology of Cancer: Metabolic Reprogramming Fuels Cell Growth and Proliferation. *Cell Metabolism.* 2008 Jan; 7 (1); 11-20.
- [80] Warburg O. On respiratory impairment in cancer cells. *Science.* 1956 Aug; 124(3215):269-70.
- [81] Soga T. cancer metabolism: Key players in metabolic reprogramming. *Cancer Sci.* 2013 Mar;104(3):275-81.
- [82] Lum JJ, Bauer DE, Kong M, Harris MH, Li C, Lindsten T, et al. Growth factor regulation of autophagy and cell survival in the absence of apoptosis. *Cell.* 2005 Jan; 120(2):237-48.

- [83] Christofk HR, Vander Heiden MG, Harris MH, Ramanathan A, Gerszten RE, Wei R, et al. The M2 splice isoform of pyruvate kinase is important for cancer metabolism and tumour growth. *Nature*. 2008 Mar; 452(7184):230-33.
- [84] Bauer DE, Hatzivassiliou G, Zhao F, Andreadis C, Thompson CB. ATP citrate lyase is an important component of cell growth and transformation. *Oncogene*. 2005 Sep; 24(41):6314–22.
- [85] Astuti D, Latif F, Dallol A, Dahia PL, Douglas F, George E, et al. Gene mutations in the succinate dehydrogenase subunit SDHB cause susceptibility to familial pheochromocytoma and to familial paraganglioma. *Am. J. Hum. Genet.* 2001 Jul; 69(1):49–54.
- [86] Tomlinson IP, Alam NA, Rowan AJ, Barclay E, Jaeger EE, Kelsell D, et al. Germline mutations in FH predispose to dominantly inherited uterine fibroids, skin leiomyomata and papillary renal cell cancer. *Nat. Genet.* 2002 Apr; 30(4):406–10.
- [87] Roos S, Jansson N, Palmberg I, Säljö K, Powell TL, Jansson T. Mammalian target of rapamycin in the human placenta regulates leucine transport and is down-regulated in restricted fetal growth. *J. Physiol.* 2007 Jul; 582(Pt 1):449-59.
- [88] Wieman HL, Wofford JA, Rathmell JC. Cytokine stimulation promotes glucose uptake via phosphatidylinositol-3 kinase/Akt regulation of Glut1 activity and trafficking. *Mol. Biol. Cell.* 2007 Apr; 18(4):1437-46.
- [89] Elstrom RL, Bauer DE, Buzzai M, Karnauskas R, Harris MH, Plas DR, et al. Akt stimulates aerobic glycolysis in cancer cells. *Cancer Res.* 2004 Jun; 64(11):3892-99.
- [90] Chang Y, Wang J, Lu X, Thewke DP, Mason RJ. KGF induces lipogenic genes through a PI3K and JNK/SREBP-1 pathway in H292 cells. *J. Lipid. Res.* 2005 Dec; 46(12):2624-35.
- [91] Gingras AC, Raught B, Sonenberg N. Regulation of translation initiation by FRAP/mTOR. *Genes Dev.* 2001 Apr; 15(7):807-26.
- [92] Semenza GL. Hypoxia-inducible factor 1 (HIF-1) pathway. *Sci STKE*. 2007 Oct; 2007(407):cm8.
- [93] Majumder PK, Febbo PG, Bikoff R, Berger R, Xue Q, McMahon LM, et al. mTOR inhibition reverses Akt-dependent prostate intraepithelial neoplasia through regulation of apoptotic and HIF-1-dependent pathways. *Nat. Med.* 2004 Jun; 10(6):594-601.
- [94] Semenza GL. Tumor metabolism: cancer cells give and take lactate. *J. of Clinical Investigation*. 2008 Dec; 118(12); 3835-37.
- [95] Shaw RJ, Cantley LC. Ras, PI(3)K and mTOR signalling controls tumour cell growth. *Nature*. 2006 May; 441(7092):424-30.
- [96] Tomiyama A, Serizawa S, Tachibana K, Sakurada K, Samejima H, Kuchino Y, et al. Critical role for mitochondrial oxidative phosphorylation in the activation of tumor suppressors Bax and Bak. *J. Natl. Cancer Inst.* 2006 Oct;98(20):1462-73.
- [97] Bonnet S, Archer SL, Allalunis-Turner J, Haromy A, Beaulieu C, Thompson R, et al. A mitochondria-K⁺ channel axis is suppressed in cancer and its normalization promotes apoptosis and inhibits cancer growth. *Cancer Cell*. 2007 Jan; 11(1):37-51.
- [98] Culmsee C, Mattson MP. p53 in neuronal apoptosis. *Biochem. Biophys. Res. Commun.* 2005 Jun; 331(3):761-77.
- [99] Kondoh H, Lleonart ME, Gil J, Wang J, Degan P, Peters G, et al. Glycolytic enzymes can modulate cellular life span. *Cancer Res.* 2005 Jan; 65(1):177-85.

- [100] Kim JW, Gao P, Liu YC, Semenza GL, Dang CV. Hypoxia inducible factor 1 and dysregulated c-Myc cooperatively induce vascular endothelial growth factor and metabolic switches hexokinase 2 and pyruvate dehydrogenase kinase 1. *Mol. Cell Biol.* 2007 Nov; 27(21):7381-93.
- [101] Chi SL, Wahl ML, Mowery YM, Shan S, Mukhopadhyay S, Hilderbrand SC, et al. Angiostatin-like activity of a monoclonal antibody to the catalytic subunit of F1F0 ATP synthase. *Cancer Res.* 2007 May; 67(10):4716-24.
- [102] Igney FH, Krammer PH. Death and anti-death: tumour resistance to apoptosis. *Nat. Rev. Cancer.* 2002 Apr; 2(4): 277-88.
- [103] Erler JT, Banneth KL, Nicolau M, Dornhofer N, Kong C, Le QT, et al. Lysyl oxidase is essential for hypoxia-induced metastasis. *Nature.* 2006 Apr; 440(7088):1222-26.
- [104] Condeelis J, Pollard JW. Macrophages: obligate partners for tumor cell migration, invasion, and metastasis. *Cell.* 2006 Jan; 124(2):263-66.
- [105] Evans JM, Ogston SA, Emslie-Smith A, Morris AD. Risk of mortality and adverse cardiovascular outcomes in type 2 diabetes: a comparison of patients treated with sulfonylureas and metformin. *Diabetologia.* 2006 may; 49(5):930-36.
- [106] Faivre S, Kroemer G, Raymond E. Current development of mTOR inhibitors as anticancer agents. *Nat. Rev. Drug Discov.* 2006 Aug; 5(8):671-88.
- [107] Perou CM, Sorlie T, Eisen MB, van de Rijn M, Jeffrey SS, Rees CA, et al. Molecular portraits of human breast tumours. *Nature.* 2000 Aug; 406(6797):747-52.
- [108] Pedraza-Fariña LG. Mechanisms of oncogenic cooperation in cancer initiation and metastasis. *Yale J. Biol. Med.* 2006 Dec; 79(3-4):95-103.
- [109] Futreal PA, Coin L, Marshall M, Down T, Hubbard T, Wooster R, et al. A census of human cancer genes. *Nat. Rev. Cancer.* 2004 Mar; 4(3):177-83.
- [110] Ohta M, Inoue H, Cotticelli MG, Kastury K, Baffa R, Palazzo J, et al. The human FHIT gene, spanning the chromosome 3p14.2 fragile site and renal carcinoma associated translocation breakpoint, is abnormal in digestive tract cancers. *Cell.* 1996 Feb; 84:587-97.
- [111] Stephens PJ, McBride DJ, Lin ML, Varela I, Pleasance ED, Simpson JT, et al. Complex landscapes of somatic rearrangement in human breast cancer genomes. *Nature.* 2009 Dec; 462(7276):1005-10.
- [112] Ding L, Ellis MJ, Li S, Larson DE, Chen K, Wallis JW, et al. Genome remodelling in a basal-like breast cancer metastasis and xenograft. *Nature.* 2010 Apr; 464(7291): 999-1005.
- [113] Lee EY, To H, Shew JY, Bookstein R, Scully P, Lee WH. Inactivation of the retinoblastoma susceptibility gene in human breast cancers. *Science.* 1988 Jul; 241(4862):218-21.
- [114] Walsh T, King MC. Ten genes for inherited breast cancer. *Cancer Cell.* 2007 Feb; 11(2):103-05.
- [115] Cooper GM. *Oncogenes*. 2nd ed Boston: Jones and Bartlett Publishers; 1995.
- [116] Armitage P, Doll R. The age distribution of cancer and a multi-stage theory of carcinogenesis. *Br. J. Cancer.* 1954 Mar; 8(1):1-12.
- [117] Bishop JM. Viral oncogenes. *Cell.* 1985 Aug; 42(1):23-38.
- [118] Krontiris TG, Cooper, GM. Transforming activity of human tumor DNAs. *Proc. Natl. Acad. Sci. USA.* 1981 Feb; 78(2):1181-84.

- [119] Land H, Parada LF, Weinberg RA. Tumorigenic conversion of primary embryo fibroblasts requires at least two cooperating oncogenes. *Nature*. 1983 Aug; 304(5927):596-02.
- [120] Hahn WC, Counter CM, Lundberg AS, Beijersbergen RL, Brooks MW, Weinberg RA. Creation of human tumour cells with defined genetic elements. *Nature*. 1999 Jul; 400(6743):464-68.
- [121] Felsher DW, Bishop JM. Reversible tumorigenesis by MYC in hematopoietic lineages. *Mol. Cell*. 1999 Aug; 4(2):199-207.
- [122] Rudolph B, Hueber A O, Evan GI. Reversible activation of c-Myc in thymocytes enhances positive selection and induces proliferation and apoptosis in vitro. *Oncogene*. 2000 Apr; 19(15):1891-900.
- [123] Pelengaris S, Khan M, Evan GI. Suppression of Myc-induced apoptosis in beta cells exposes multiple oncogenic properties of Myc and triggers carcinogenic progression. *Cell*. 2002 May; 109(3):321-34.
- [124] Beer S, Zetterberg A, Ihrie RA, McTaggart RA, Yang Q, Bradon N et al. Developmental context determines latency of MYC-induced tumorigenesis. *PLoS Biol*. 2004 Nov; 2(11):e332.
- [125] Boxer RB, Jang JW, Sintasath L, Chodosh LA. Lack of sustained regression of c-MYC-induced mammary adenocarcinomas following brief or prolonged MYC inactivation. *Cancer Cell*. 2004 Dec; 6(6):577-86.
- [126] Jonkers J, Berns A. Oncogene addiction: sometimes a temporary slavery. *Cancer Cell*. 2004 Dec; 6(6):535-38.
- [127] Fidler IJ. Critical determinants of metastasis. *Semin. Cancer Biol*. 2002 Apr; 12(2):89-96.
- [128] van 't Veer LJ, Dai H, van de Vijver MJ, He YD, Hart AA, Mao M, et al. Gene expression profiling predicts clinical outcome of breast cancer. *Nature*. 2002 Jan; 415(6871):530-36.
- [129] Ramaswamy S, Ross KN, Lander ES, Golub TR. A molecular signature of metastasis in primary solid tumors. *Nat. Genet*. 2003 Jan; 33(1):49-54.
- [130] Woelfle U, Cloos J, Sauter G, Riethdorf L, Jänicke F, van Diest P, et al. Molecular signature associated with bone marrow micrometastasis in human breast cancer. *Cancer Res*. 2003 Sep; 63(18):5679-84.
- [131] Goldman JM, Melo JV. Targeting the BCR-ABL tyrosine kinase in chronic myeloid leukemia. *N. Engl. J. Med*. 2001;344:1084-6.
- [132] Heinrich MC, Blanke CD, Druker BJ, Corless CL. Inhibition of KIT tyrosine kinase activity: a novel molecular approach to the treatment of KIT-positive malignancies. *J. Clin. Oncol*. 2002;20:1692-703.
- [133] Casali PG, Messina A, Stacchiotti S, et al. Imatinib mesylate in chordoma. *Cancer* 2004;101:2086-97.
- [134] Wang ZG, Delva L, Gaboli M, et al. Role of PML in cell growth and the retinoic acid pathway. *Science* 1998;279:1547-51.
- [135] Salomoni P, Pandolfi PP. The role of PML in tumor suppression. *Cell* 2002;108: 165-70.
- [136] Ambros V. The function of animal miRNAs. *Nature* 2004;431:350–355.
- [137] Sontheimer EJ, Carthew RW. Silence from within: endogenous siRNAs and miRNAs. *Cell* 2005;122:9–12.

- [138] Meister G, Tuschl T. Mechanisms of gene silencing by double-stranded RNA. *Nature* 2004;431:343–349.
- [139] Chapman EJ, Carrington JC. Specialization and evolution of endogenous small RNA pathways. *Nat. Rev. Genet.* 2007;8:884–896.
- [140] Calin GA, Dumitru CD, Shimizu M, Bichi R, Zupo S, Noch E, et al. Frequent deletions and down-regulation of micro-RNA genes miR15 and miR16 at 13q14 in chronic lymphocytic leukemia. *Proc. Natl. Acad. Sci. U S A.* 2002 Nov; 99(24):15524–29.
- [141] Calin GA, Sevignani C, Dumitru CD, et al. Human microRNA genes are frequently located at fragile sites and genomic regions involved in cancers. *Proc. Natl. Acad. Sci. U S A* 2004;101:2999–3004.
- [142] Borchert GM, Lanier W, Davidson BL. RNA polymerase III transcribes human microRNAs. *Nat. Struct. Mol. Biol.* 2006;13:1097–1101.
- [143] Francesco Fazi and Clara Nervi. MicroRNA: basic mechanisms and transcriptional regulatory networks for cell fate determination. *Cardiovascular Research* 2008;79: 553–561.
- [144] Calin GA, Ferracin M, Cimmino A, Di Leva G, Shimizu M, Wojcik SE, et al. A MicroRNA signature associated with prognosis and progression in chronic lymphocytic leukemia. *N. Engl. J. Med.* 2005 Oct; 353(17):1793–801.
- [145] Calin GA, Croce CM. MicroRNA signatures in human cancers. *Nat. Rev. Cancer*, 2006;6:857–66.
- [146] Mengfeng Li, Jun Li, Xiaofan Ding, Mian He, and Shi-Yuan Cheng. MicroRNA and Cancer. *The AAPS Journal* 2010; 12:309–317.
- [147] Volinia S, Calin GA, Liu CG, Ambs S, Cimmino A, Petrocca F, et al. A microRNA expression signature of human solid tumors defines cancer gene targets. *Proc. Natl. Acad. Sci. U S A.* 2006 Feb;103(7):2257–61.
- [148] Nakamura T, Canaani E, Croce CM. Oncogenic All1 fusion proteins target Drosha-mediated microRNA processing. *Proc. Natl. Acad. Sci. U S A.* 2007 Jun; 104(26): 10980–85.
- [149] Yanaihara N, Caplen N, Bowman E, Seike M, Kumamoto K, Yi M, et al. Unique microRNA molecular profiles in lung cancer diagnosis and prognosis. *Cancer Cell.* 2006 Mar; 9(3):189–98.
- [150] Nakamura T, Canaani E, Croce CM. Oncogenic All1 fusion proteins target Drosha-mediated microRNA processing. *Proc. Natl. Acad. Sci. USA* 2007;104: 10980–5.
- [151] Eis PS, Tam W, Sun L, Chadburn A, Li Z, Gomez MF, et al. Accumulation of miR-155 and BIC RNA in human B cell lymphomas. *Proc. Natl. Acad. Sci. USA.* 2005 Mar; 102(10):3627–32.
- [152] Costinean S, Zanesi N, Pekarsky Y, Tili E, Volinia S, Heerema N, et al. Pre-B cell proliferation and lymphoblastic leukemia/high grade lymphoma in E (mu)-miR155 transgenic mice. *Proc. Natl. Acad. Sci. USA.* 2006 May; 103(18):7024–29.
- [153] Johnson SM, Grosshans H, Shingara J, et al. RAS is regulated by the let-7 microRNA family. *Cell* 2005;120:635–47.
- [154] Cimmino A, Calin GA, Fabbri M, et al. miR-15 And miR-16 induce apoptosis by targeting BCL2. *Proc. Natl. Acad. Sci. USA* 2005;102:13944–9.
- [155] Meng F, Henson R, Lang M, et al. Involvement of human microRNA in growth and response to chemotherapy in human cholangiocarcinoma cell lines. *Gastroenterology* 2006;130:2113–29.

Chapter 103

**DICER AND MICRORNAS:
ONCOMIRS ARE THE NEXT FRONTIER
OF ONCOGENES AFFECTING CANCER
ETIOLOGY AND TUMOR PROGRESSION**

Ha T. Nguyen and Mandi M. Murph

American Cancer Society Research Scholar
Georgia Cancer Coalition Distinguished Cancer Scholar
University of Georgia College of Pharmacy
Department of Pharmaceutical and Biomedical Sciences
Athens, Georgia, US

ABSTRACT

Recent understanding of oncogene regulation has uncovered an emerging new field of molecular cancer biology in which tumor suppressors and mediators are controlled through RNA regulation. Among the most critical players are microRNA (miRNA), also referred to as 'oncomirs', and Dicer, the major enzyme responsible for cleaving double-stranded RNA and forming the RNA induced silencing complex. These components are aberrantly expressed in cancer and among some tumor types are hypothesized to be causative to the etiology of malignancy. For example, prostate and colorectal cancers express abundantly high levels of Dicer mRNA while lung, ovarian and endometrial cancers express low levels of Dicer, which is believed to correlate to poor cancer prognosis. In addition, mutations of the Dicer encoding gene (*DICER1*) occur in non-epithelial ovarian cancers and pediatric tumor pleuropulmonary blastoma. Paradoxically, Dicer is a haploinsufficient tumor suppressor; the loss of a single allele of *DICER* enhances tumor growth where the loss of the second allele results in halting tumor proliferation. MicroRNA regulates oncogenic signaling pathways as well as the expression of tumor suppressors and oncogenes, making it a major contributor to the overall status of the cell and its malignant potential. The let-7 miRNA family members are well-known as tumor suppressor genes, which target and silence the Ras oncogene. On the other hand, miRNAs may induce tumor growth; one example is miR-17-19 cluster negatively regulating two tumor suppressor genes, PTEN and Bim. In this chapter, the regulation of oncomirs will be discussed with focus on the

post-transcriptional control by miRNA biogenesis machinery, which consist Dicer as a major player.

INTRODUCTION

MicroRNAs (miRNAs) are noncoding RNAs 18-25 nucleotides in length, which regulate gene expression by targeting messenger RNAs (mRNAs). Through miRNA binding to mRNA's complementary base pair sequences, mRNA is then degraded, which ultimately silences gene expression in the cell. Alternatively, gene silencing may also be achieved by translation inhibition without a requirement for exact complementarity. Profiling miRNAs in human cancer showed that miRNA expression differs between normal and cancer tissues, and also between tumor types [1]. MicroRNAs are suggested to have roles in cancer development both as oncogenes as well as tumor suppressors [2, 3]; therefore they are referred as 'oncomirs'. To further complicate the issue, other miRNAs like miR-146 and miR-29 are context-dependent [4], which means they can function as either an oncogene or tumor suppressor, depending on the cell type and the specific regulatory pathways that are either functional or deficient in that system.

MICRORNA AND CANCER

MicroRNAs regulate mRNA levels by directing the RISC to the 3'-untranslated region (UTR) of target mRNAs by complementary sequences on miRNAs to achieve gene silencing [5, 6]. The RISC complex contains Ago, an RNase III endonuclease. Although human cells have four Ago proteins (Ago1-4), only Ago2 in the miRNA-RISC complex processes mRNA cleavage [7]. The extent of complementarity between miRNAs and target mRNAs also determines whether the mRNA will be cleaved or just inhibited from translation [6]. A single miRNA can regulate different target genes and one target mRNA can also be regulated by multiple miRNAs [3].

Since miRNAs regulate multiple fundamental biological processes, alterations in their normal functioning could result in contributing to the etiology of cancer. As such, research continues to elucidate the roles of miRNAs in tumorigenesis as both tumor suppressors and oncogenes. Analysis of tumor and normal tissues from different types of cancers showed altered expression levels of many miRNAs, either via overexpression or downregulation [1, 8-15]. To date many miRNAs have been identified for their crucial roles in cancer, among them the let-7 family and the miR-17-92 cluster are well studied as tumor suppressors and oncogenes.

MICRORNA-BASED THERAPIES IN CANCER

Increasing evidence demonstrates significant regulatory roles of miRNAs. Due to their aberrant expression and function in cancer, miRNAs and their antagomirs have become appealing therapeutic potentials against cancer. There are three principal directions or applications for miRNA therapies. The first application for miRNAs as therapeutics consists of

replacing the lost miRNA or using antagomirs as direct targets against malignancy, which means overexpressing antitumor miRNAs and/or repressing oncogenes. Combinations of appropriate miRNAs or antagomirs might increase specificity to the target and reduce the risk for acquired resistance. In order for this approach to be successful at the level of molecular mechanisms, an in-depth knowledge of all the biological targets and regulatory pathways, including the ensuing off-target effects would have to be elucidated for a given miRNA first.

The goal of the second application is to increase drug sensitization of cancer cells through modulation of related miRNAs. This direction is based on the results from multiple labs, which show that despite some miRNAs not displaying significant direct effects on cancer cell survival, their alterations increase cancer cell sensitivity to drug treatment or radio therapy [16-18]. With this approach, a new era in pharmacogenomics would emerge since it is highly likely there would be significant patient variability to drug resistance and miRNA expression. The third application of miRNA therapies is inhibiting specific aberrant functions of cancer cells in a context-dependent fashion (e.g., cancer invasion and migration) through the modulation of specific miRNAs, such as miR-34a, miR-34b, and miR-21 [19-21]. To sum up, miRNA therapies are attempts to block “bad miRNAs” and increase “good miRNAs” to achieve the desired therapeutic effects.

Since foreign oligonucleotides like synthetic small-interfering RNA (siRNA) molecules in general are subject to degradation by nucleases, many modifications have been developed to increase their stability. At the same time, increases in potency and reductions in toxicity have been improved from early modifications such as phosphorothioate antisense oligonucleotides, 2'-O-methyl or 2'-O-methoxy-ethyl anti-sense oligonucleotides [22-26]. Synthetic anti-sense oligonucleotides, which have complementary sequences to oncogenic miRNAs, are used to inactivate miRNAs at tumor sites. An improvement to these is the cholesterol conjugation of anti-sense oligonucleotides, which are very stable in vivo and improves therapeutic efficacy [27-29].

Among all the chemical modifications of antisense oligonucleotides, LNA is likely the most common. For example, it is used to create antagomirs for miR-34a and the miR-17-92 cluster in vivo and in vitro [30-33]. Locked nucleic acid (LNA) is defined as an oligonucleotide that contains one or more 2'-O,4'-methylene- β -D-ribofuranosyl monomer(s) [30]. Furthermore, LNA mediated silencing of miRNA-122 has reached clinical trials in non-human primates [34-36]. It will be intriguing to follow this progress and learn whether or not these modifications are sufficient for the application of miRNA therapy.

Since miRNAs are highly tissue specific, another challenge is to deploy these therapeutics at the appropriate site. In this regards, multiple efforts have been spent on developing carriers to deliver antitumor miRNAs to tumors and even a greater understanding of how miRNAs are transported in biological systems. Interestingly, studies have demonstrated that miRNAs are highly stable and abundant extracellular entities found in circulating plasma and protected from degradation by endogenous RNases, although subject to degradation by proteases [37, 38]. Thus, circulating miRNAs are protected through several mechanisms, including binding and forming a complex with Argonaute2 [38], secretion into microvesicles from the cell and high-density lipoproteins [39].

The scenario of the normal regulation of miRNAs in circulation may differ from lessons learned through the usage of foreign siRNAs, which have several challenges during the process to reach targets. First, foreign siRNAs are degraded in plasma by nucleases [40] (whereas endogenous miRNAs have natural protection). Second, siRNAs are cleared from circulation

rapidly through kidney filtration due to their small size [41]. Third, oligonucleotides' negative charge prevents them from passing through the cellular lipid bilayer, which is impermeable to ions and polar molecules. Fourth, even after cellular internalization, foreign oligonucleotides still need to escape endosomal entrapment and degradation in order to get loaded by the RISC complex and target complementary mRNAs.

To overcome barriers and facilitate therapeutic usage of miRNAs, several delivery systems have been introduced. Viral vector-based systems were successfully tested in murine models of lung and liver cancer [42, 43]. However, because safety and host immune responses are considerable concerns for viral vectors, lipid based delivery systems are a more promising option for clinical use. Using a liposome, which is a lipid bilayer vesicle, has been developed to carry nucleic acids for an unusually long time. The lipid cage allows the liposome to remain in circulation for a much longer time [44]. The particle itself is biodegradable to optimize safety administration to host body [41]. Moreover, the liposome lipid membrane can be cationized to improve cellular fusion and uptake [45]. Antibody incorporation into the liposomal membrane increases target site specificity, thus reducing undesirable off-target effects [46, 47]. Further techniques help facilitate liposomes to escape from endosomal engulfment, such as ion-pair formation, the "proton sponge" effect or de-assembly [41].

Other methods and rising trends in cancer therapy include using nanoparticles, either lipid-based or nonlipid-based, to deliver miRNAs or antagomir oligonucleotides. Interfering nanoparticles (iNOPs), a lipid nanoparticle with its surface modified with cationic lysines, has succeeded in delivering anti-miR-122 to mouse liver [48, 49]. Nanoparticles coated with cell-penetrating peptides were used to increase the efficiency of anti-miRNA-155 delivery to cancer cells [50]. Gold nanoparticles are another promising delivery system due to reduced toxicity, a favorable ability to penetrate cells, a longer retention in circulation and easiness in biosensing [51-55]. Gold nanoparticles conjugated with miRs or antagomirs have good transfection and delivery results in vitro [56-58].

MICRORNA BIOGENESIS AND DEGRADATION

MicroRNAs are transcribed by RNA polymerase II into large RNA precursor primary-miRNA (pri-miRNA) [59, 60]. Similar to mRNAs, pri-miRNAs have caps at the 5' end and are poly-adenylated at the 3' end [61]. Pri-miRNAs form stem-loop secondary structures and are processed in the nucleus by Drosha and DGCR8, reducing the structure from hundreds of nucleotides to approximately 70-nucleotide pre-miRNAs [62]. Pre-miRNAs are then exported to the cytoplasm by Exportin-5 (Exp-5) [63]. In the cytoplasm, pre-miRNAs are cleaved by DICER, with assistance from TARBP2, to generate a 22-nucleotide, double-stranded miRNA-miRNA* duplex, which is also referred to as the miR-5p/miR-3p duplex [64, 65]. Mature miRNAs and miRNAs* are then separated and miRNAs are loaded into the RNA-induced silencing complex (RISC) while miRNAs* are more often degraded [66, 67].

However, recent publications, including our own, have altered our understanding of the fate of miRNAs* and questioned whether they are always degraded. More work is needed to elucidate the roles and contributions of these so-called "passenger" strands because studies have shown significant effects of miRNAs* in cancer and also suggested they have important regulatory abilities on their own [68-70]. It is no longer the accepted dogma that miRNAs* are

simply non-functional byproducts of the miRNA-miRNA* duplex and have no purpose other than degradation. This is an exciting time in the field of miRNA research because there is still much to uncover in our understanding of these strands and which biological effects they regulate. In particular, are they involved when things go awry and lead to cancer?

LET-7 FAMILY

Let-7 in Cell Differentiation

Let-7, together with lin-4, was the first known miRNA [71-73]. Let-7 was first identified in the nematode *Caenorhabditis elegans* as a component involved in early development [71]. In *C. elegans*, let-7 is important for and highly expressed during the larval-to-adult transition. In this period, a type of *C. elegans* stem cell called seam cells switches from proliferating in the early larval stage to terminal differentiating [71]. In the absence of functioning let-7, seam cells fail to exit the cell cycle at this transition point, thus continuing to divide instead of differentiating. As a result, the let-7 mutant animals die by bursting through vulvas, and that is the origin of the name let-7, which means lethal-7 [71]. The sequence of let-7 and its temporal expression are conserved in a wide range of animal species, from as simple as *C. elegans* to as complicated as humans [74]. In zebra fish let-7 is also detected at 48 hours after fertilization, in annelids and mollusks, the time points are adult stages [74]. Later sequencing in mouse and humans extends let-7 to a family of identical mature miRNAs encoded by 13 genomic loci from let-7a to let-7i [75]. Among them, let-7a has its sequence identical across species; while the other members keep the seed sequence but vary in other nucleotides. Current hybridization-based techniques, such as microarray and northern blot, can hardly distinguish closely related miRNAs. Thus, it is a challenge to quantify each single let-7 family member in one sample. For this reason, in this book chapter we will refer to let-7 as a term to generally describe this whole let-7 family. Let-7 is reportedly absent in embryonic stem cells or pluripotent cells but up-regulated in differentiating cells, such as brain cells, or breast stem-cell progenitors at the differentiating phase [76-80]. Lower expression of let-7 is observed as one feature of certain types of stem cells while overexpression of let-7 seems to reduce dividing [80]. Let-7 is considered one factor involved in switching cells from the proliferation phase to differentiation phase [81].

LET-7 AND CANCER: RAS, HMGA2 AND THE CELL CYCLE

The early evidence of let-7's role in cancer was the observation that let-7 expression levels are reduced in lung cancer tissues and this event is associated with shorter postoperative survival among patients [90]. Furthermore, let-7 miRNAs are mapped to the frequently deleted chromosomal sites in lung cancer [91]. More importantly, the introduction of let-7 to a low-let-7 lung cancer cell line inhibited cell growth [90].

Based on this information, a later study showed that down regulating Ras is one important mechanism whereby let-7 ceases lung cancer cell growth [92]. Ras is an enzymatic protein at the cell membrane, which is activated through binding to GTP and then triggering a cascade of

downstream signaling events, including ERK/MAPK activation [82]. Ras signaling cascades favor cell proliferation and survival [83, 84]. Mutations in *RAS* genes (e.g., *HRAS*, *NRAS* and *KRAS*), which cause increased Ras expression and activity, are reported in a number of different cancers, including pancreatic and colorectal [85-87]. Similarly, overexpression and increased activity of Ras due to *RAS* gene mutations occurs in lung cancer [88, 89]. There are multiple let-7 complementary sites in the 3' UTRs of human *RAS* genes, allowing let-7 to bind and block the expression of Ras through Ras mRNA silencing [92]. The antitumor effect of let-7 was proved in vivo when the delivery of let-7 reduced tumor burden in K-Ras G12D, a Ras mutated, human lung cancer xenograft model [93]. Beside Ras, HMGA2 is another oncoprotein regulated by let-7 [94]. HMGA2 is a member of the high mobility group AT-hook (HMGA) family, a family of nonhistone chromatin proteins that act as architectural transcription factors [95]. HMGA2 is involved in the transcription of various genes and is essential for growth during embryonic development [95-97]. HMGA2 is reported to be up-regulated in tumors, such as lung cancers and liposarcomas [97-99]. Taken together, HMGA2 is suggested to be an oncogenic factor in differentiated tissues, which express negligible HMGA2 under normal conditions. Multiple studies have provided evidence of let-7 as a negative regulator of HMGA2 in cancers [94, 100, 101]. Sequencing and reporter assays confirm that the 3'UTR of HMGA2 contains let-7 complementary sites [94, 100]. In fact, removing the let-7 complementary sites in HMGA2 3'UTR causes HMGA2 overexpression and tumorigenesis [100]. Interestingly, a study on self-renewing-tumor-initiating breast cancer cells showed that increased let-7 levels had negative effects on both Ras and HMGA2, while Ras seemed to be involved in self renewal and HMGA2 was involved in differentiation [102]. However, the antitumor effects of let-7 though Ras and HMGA2 are not always equal. In non-small cell lung cancer, overexpression of let-7 caused a reduction in both Ras and HMGA2, but ectopic expression of Ras showed a reversal of effects to let-7 mediated tumor suppression more so than ectopic expression of HMGA2 [42]. It is predicted that one miRNA strand can target hundreds of mRNAs. Correspondingly, let-7 exerts antitumor effects through not only Ras and HMGA2 but also many other targets that are important to the regulation of the cell. For example, a large number of cell-cycle genes have been identified to have let-7 complementary sites [103]. Among them, *CDC25A*, which is an oncogene, *CDK6* and *cyclin D1* are confirmed to be directly regulated by let-7 [103-106]. Upsetting the delicate balance of the regulation of the cell cycle could wreck havoc upon normal cells and homeostasis. In some systems these alterations are involved in the etiology of cancer.

REGULATION OF LET-7

Lin-28 inhibits the processing of let-7 family members through binding to pri-let-7, then blocks the let-7 precursor from Drosha-DGCR8 cleavage in the nucleus [107, 108]. This binding and blocking process is mediated through the recognition of Lin-28 to several conserved nucleotides shared by let-7 family members, thus it seems to be selective for the let-7 family with little or no effect observed on other miRNAs [107-110]. Another step that Lin-28 uses to negatively regulate let-7 is inducing the uridylation and recruiting the uridylating enzyme TUT4 to pre-let-7 in the cytoplasm [111-114]. All these modifications block pre-let-7 from being processed by Dicer to form mature let-7 and ultimately cause it to be degraded in

the cytosol [113]. Lin-28 facilitates cellular transformation and is overexpressed in multiple human primary tumor types in concert with let-7 reduction [115]. Lin-28 is believed to be involved in let-7 regulation from the observation that Lin-28 expression is high in early developmental stages and low during differentiation, which is reciprocal to let-7 levels [116]. The same reciprocal pattern is also observed in cancer cells [117, 118]. Furthermore, Lin-28 is identified as one of the four genes that together can convert human adult fibroblasts to pluripotent stem cells (the others three names are Oct4, Nanog and Sox-2) [119]. Even though it is regulated by Lin-28, let-7 creates a regulatory loop through direct targeting of Lin-28 mRNA 3'UTR [120, 121]. These pieces of evidence suggest that let-7 can escape Lin-28 mediated down-regulation and amplify itself [121, 122].

In the process of transforming from somatic cells to pluripotent stem cells, Lin28 can be replaced by Myc [123, 124]. Myc is an oncogenic transcription factor, which induces tumorigenesis through transactivating multiple genes involved in proliferation and survival [125-128]. Interestingly, very similar to Lin-28, Myc also inhibits let-7 expression through binding to promoters or sequences upstream of genes encoding the let-7 family and directly repressing transcription [129]. On the other hand, Myc induces Lin-28B expression, which in turn further represses let-7 maturation as described above [130]. Again, similar to the case of Lin28, let-7 comes back to directly target and down-regulate Myc expression [131-135].

MiR-17-92 FAMILY AND ORGANIZATION OF miRNA CLUSTERS

The miR-17-92 family of miRNA includes six miRNAs (miR-17, miR-18a, miR-19a, miR-20a, miR-19b-1, and miR92a), which share the same seed sequence, and are encoded tightly in an intron on human chromosome 13 [136, 137]. This cluster has two paralogs: the miR-106b-25 located on human chromosome 7, and the miR-106a-363 located on chromosome X. The miR-106a-363 cluster consists of six miRNAs: miR-106, miR-18b, miR-20b, miR-19b-2, miR92a-2, and miR-363. The miR-106b-25 cluster consists of three miRNAs: miR-106b, miR-93, and miR-25 [137]. The miR-106b-25 cluster consists of three miRNAs: miR-106b, miR-93, and miR-25 [137]. Members of the microRNA-17-92 family clusters are grouped into 4 sub-families, in which they share the same seed: miR-17 (miR-17, miR-20a, miR-20b, miR-106a, miR-106b, and miR-93), miR-18 (miR-18a and miR-18b), miR-19 (miR-19a and miR-19b), and miR-92 (miR-25, miR-92a, and miR-363) [137, 138].

REGULATION OF THE miR-17-92 FAMILY AND ITS ROLES IN CANCER

The miR-17-92 family are potential oncogenes: studies showed that this cluster facilitates tumor development in a mouse B-cell lymphoma model [139]. The oncogenic effects of this family are confirmed in many other types of cancers, such as lung, colorectal, liver and thyroid cancer [140-144]. Studies showed that the miR-17-92 family induces cancer through multiple mechanisms. For example, they induce proliferation through repression of Bim, a proapoptotic protein, and PTEN, a tumor suppressor [145]. High expression of miR-17-92 inhibits hypoxia-induced apoptosis because the key transcription factor involved in this process, hypoxia-inducible factor 1 α (HIF-1 α), is a direct target [143, 146]. MiR-17-92, and miR106b down-

regulate CDKN1A (p21Waf1/Cip1), a well-known tumor suppressor, which stops cell proliferation by suspending the cell-cycle progression [147, 148]. Myc is the factor that transactivates the cluster through binding to the promoter regions [149]. The E2F family of transcription factors also up-regulate miR-17-92 expression through direct binding to its promoter [150, 151]. In an interesting twist, miR-17 and miR-20 come back to inhibit translation of E2F1, E2F2, and E2F3 [149-151].

Table 1. miRNAs associated with cancer and their known targets

miRNA	Cancer association	Function	Targets	References
miR-15a miR-16-1	B-cell lymphocytic leukemia	TS	BCL2, MCL1, CCND1, WNT3A	[152-154]
miR-122	Breast, liver cancer	TS	ADAM17, IGF1R	[155-157]
miR-143	Pancreatic, colorectal, prostate cancer	TS	GEF1, GEF2, Ras	[158-160]
miR-144	Colorectal cancer	TS	mTOR	[161]
miR-155	Leukemia	OC	HDAC4	[162]
Let-7 family	Lung cancer	TS	HMGA2, Ras	[93, 94]
miR-21	Colorectal and lung cancer	OC	PTEN	[21, 163]
miR-27a	Acute leukemia	TS	4-3-3 theta	[164]
miR-17-19 family	Lung cancer	OC	PTEN, Bim, CDKN1A (p21Waf1/Cip1), E2F	[145, 148, 150]
miR-106b-25	Breast cancer, Prostate cancer	OC	Smad7 p21	[165], [166]
miR-182	Ovarian cancer	OC	BRCA1, HMGA2, FOXO3	[167]
miR-200c	Ovarian cancer, breast cancer	TS	TUBB3 (class III beta-tubulin gene), TrkB, NFkB	[168], [169]
miR-23b	Prostate cancer	TS	Src kinase, Akt	[170]
miR-218	Cancer metastasis to bone Gastric cancer	OC/TS	Wnt inhibitors: SOST (Sclerostin), Dkkopf2 (DKK2), SFRP2 (Secreted frizzled-related protein2) Robo1 receptor	[171],[172]
miR-34c	Resistance to apoptosis induced by paclitaxel in lung cancer	OC	Bmf (Bcl-2-modifying factor), Myc	[173]
miR-302-367 cluster	Cervical cancer	TS	CyclinD1, Akt1, indirectly up-regulate p271, p21	[174]

OC- Oncogene, TS- Tumor Suppressor

CAUSES OF ALTERATIONS IN miRNA EXPRESSION

Chromosomal Instability

Large portions of miRNAs are located at cancer-associated genomic regions. These regions contain oncogenes, tumor-suppressor genes and fragile sites, which are sensitive to sister-

chromatid exchange, translocation, deletion and amplification. MicroRNA loci are also prone to have alterations in human cancers [175]. Furthermore, multiple miRNAs with roles in cancer are located in cancer fragile regions. For example, the cluster miR-17-92 is located at chromosome 13q31, a region amplified in B-cell lymphomas, malignant lymphomas and lung cancers [139, 140, 176].

Epigenetic Regulations

MicroRNAs are also subject to epigenetic regulations, processes which can be involved in tumorigenesis. Epigenetic alterations are changes in gene expression caused by factors other than DNA coding, including DNA methylation, histone modifications and miRNA regulation. Multiple genes encoding miRNAs are methylated in cancer progression, such as multiple myeloma and chronic lymphocytic leukemia [177, 178]. Histone modification is also reported to have an essential role in miRNA regulation in hepatocellular carcinoma [179]. Furthermore, simultaneous inhibition of DNA methylation and histone deacetylation with a combination of 5-aza-2'-deoxycytidine and 4-phenylbutyric acid (PBA) significantly changes miRNA expression in bladder cancer cells [180].

Abnormalities in miRNA Biogenesis Machineries

Since miRNAs regulate many crucial pathways to maintain normal biological functions, the regulation of miRNAs themselves is strictly controlled. After transcription, pri-miRNAs need to go through multiple steps in order to transform into mature and fully functional miRNAs, which can target and silence complementary mRNAs. MicroRNA microbiogenesis is a complex process involving multiple proteins, any alteration in such factors may lead to changes in miRNAs expression and furthermore, cancer.

Drosha

Drosha expression in cancer is controversial. In triple-negative breast cancer tissue, the expression of Drosha is reported to be significantly higher than in normal breast tissue [181]. The same pattern is also observed in ovarian serous carcinoma [182]. However, in another clinical study, Drosha mRNA and protein expression were shown to be reduced in epithelial ovarian cancer specimens [183] and nasopharyngeal carcinoma [184]. BRCA1 regulates miRNA biogenesis through interaction with Drosha [185]. The alterations of Drosha in cancer are usually coupled with changes in overall miRNA expression. However, there is not yet any report showing a correlation between pathological Drosha expression and any specific miRNA(s).

DGCR8

In order for Drosha to process pri-miRNA, it needs to form a complex with DGCR8, a double-stranded RNA binding protein [184, 186]. DGCR8 is encoded by a gene located in a region frequently deleted in DiGeorge syndrome, the most common genetic deletion syndrome in humans [187]. DGCR8 binds pri-miRNAs while Drosha cleaves them and the two proteins interact with each other and form a so-called microprocessor complex [188]. Interestingly, while DGCR8 is known to stabilize Drosha through protein-protein interactions, knockdown of Drosha increases both DGCR8 mRNA and protein levels in cells [188, 189]. This could be interpreted as a mechanism cells use to maintain certain levels of these important proteins in the miRNA biogenesis complex.

Exportin-5

After being processed by the microprocessor complex, pre-miRNAs are transported from the nucleus to the cytoplasm by Exportin-5, a nucleocytoplasmic transport factor. In order to transport pre-miRNAs, Exportin-5 creates a complex with RanGTP and migrates from the nucleus to the cytoplasm. In the cytoplasm, pre-miRNA is released when RanGTP is hydrolyzed to RanGDP, which will be transported back to the nucleus with Exportin-5 [190-193]. Exportin-5 is reported to have tumor suppressor features since cancer cells have a genetic defect in Exportin-5, leading to the accumulation of pre-miRNAs in the nucleus [194].

The Structure of Dicer

Dicer is an endonuclease III that cleaves pre-miRNA into its final products, miRNA and miRNA*. The protein structure of Dicer from N-terminal to C-terminal includes an ATPase/helicase domain, DUF283 domain, PAZ domain, RNaseIIIa and RNaseIIIb domain, and dsRBD domain (Figure 1a). The ATPase/helicase domain helps Dicer to differentiate RNA substrates through binding to their terminal loops. Deletion of the ATPase/helicase domain leads to equal enzyme activity on both pre-miRNA and dsRNA substrates, while wild-type Dicer has favorable activity toward pre-miRNA [195].

DUF283 is a function unknown domain [196]. The PAZ domain anchors the ds-RNA end for RNase cleavage; therefore, the distance between the PAZ domain and the active sites of the RNase III domains determines the size of Dicer products [195]. Besides that, the dsRBD domain in Dicer is required for dsRNA binding only when the PAZ domain is absent [195]. The RNase IIIb domain is required for miRNA cleavage while the RNase IIIa domain is preferred for miRNA* cleavage [197, 198] (Figure 1b).

Dicer in Early Development

Dicer is essential for development in the early stages [199, 200]. The deletion of the *DICER* gene can lead to deficient and abnormal sperm cell formation. [201]. Similarly, Dicer has

critical roles in meiosis of the female germline [202]. Dicer null oocytes appear to have non-efficient chromosome-microtubule engagement, triggering a spindle checkpoint and delayed arrest, which results in a failure in mitotic chromosome segregation [202]. *DICER* heterozygous mutant mice embryos are not viable [199]. Elimination of *DICER* during embryogenesis causes perinatal death with loss of skeletal muscle mass and abnormal myofiber morphology, thus arresting embryonic development before the body plan is configured [199, 203]. Heterozygous mutation of *DICER* leads to small and abnormal embryonic morphology [199]. More interestingly, *DICER* mutant embryos have reduced Oct4 expression, which is one of key components to maintain embryonic stem cell proliferation [119, 199]. Dicer is also important for the later stages of development. It plays crucial roles in the development of lung epithelial morphogenesis and the normal maintenance of the mature pancreas [204, 205]. Deletion of *DICER* in somatic cells such as ovarian granulosa cells, mesenchyme-derived cells of the oviducts and uterus resulted in female sterility and multiple reproductive defects [206].

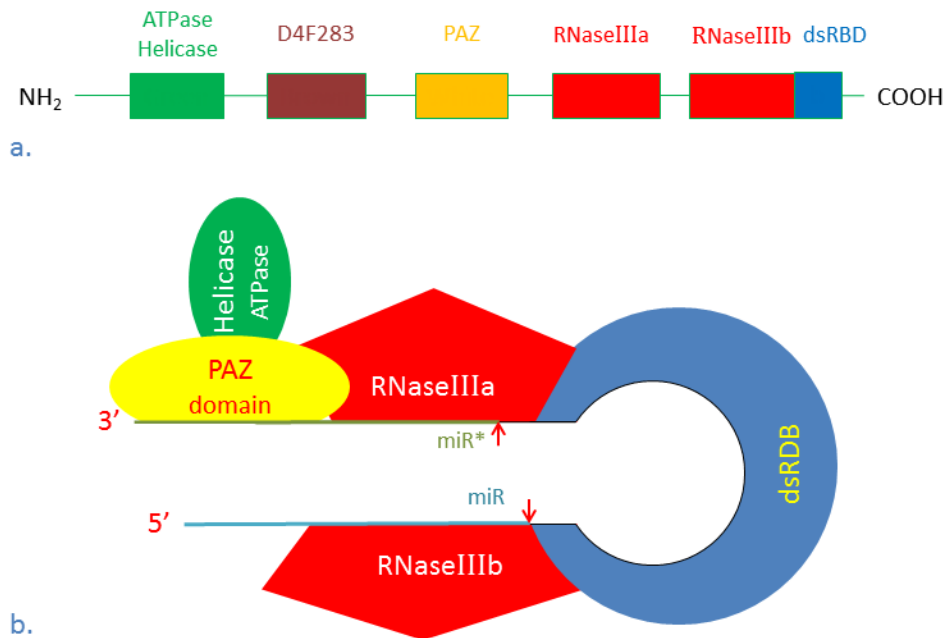


Figure 1. The schematic structure of the human Dicer protein.

The Role of Dicer in Cancer

Alterations of Dicer expression are observed in multiple genetic diseases. For example, “DICER1 syndrome” is a term used to describe familial pleuropulmonary blastoma, a rare malignant lung tumor that occurs in children under age of 6 [207]. Results from sequencing of DNA from a large amount tumors show that germ-line mutations in the *DICER* gene are the main causes of pleuropulmonary blastoma. Besides that, mutations of *DICER* may lead to different types of tumors, preferentially cystic nephroma and ovarian Sertoli-Leydig tumors [207]. Germline mutation of *DICER1*, the gene encoding Dicer in human, plays an important

role in Embryonal rhabdomyosarcoma, the most common childhood soft tissue sarcoma and its somatic mutation may have certain influence on the cause of the disease [208].

Dicer is a haploinsufficient tumor suppressor, which means the deletion of one *DICER* allele leads to tumorigenesis while complete loss of *DICER* suppresses tumors [209]. Mouse models of human cancers show that deletion of one *DICER* copy reduces survival compared to controls and *DICER* deleted animals. In a *DICER* conditional knockout mouse model of lung cancer, tumor progression showed selection against complete *DICER* recombination, which means growing tumors consisted mainly of Dicer haplodeficient cells instead of Dicer wholly deficient cells. On the other hand, forced complete deletion of Dicer led to a significant decrease in tumor burden compared to tumors where one *DICER* allele remained (Figure 2).

Surveys of human tumors also showed frequent loss of one allele of the gene encoding *DICER* in several tumor types, including breast, kidney, liver, lung, ovarian, pancreas and stomach [209]. The expression of Dicer is altered in many types of cancer. In epithelial ovarian cancer, low Dicer mRNA expression is recorded in the majority of specimens [183].

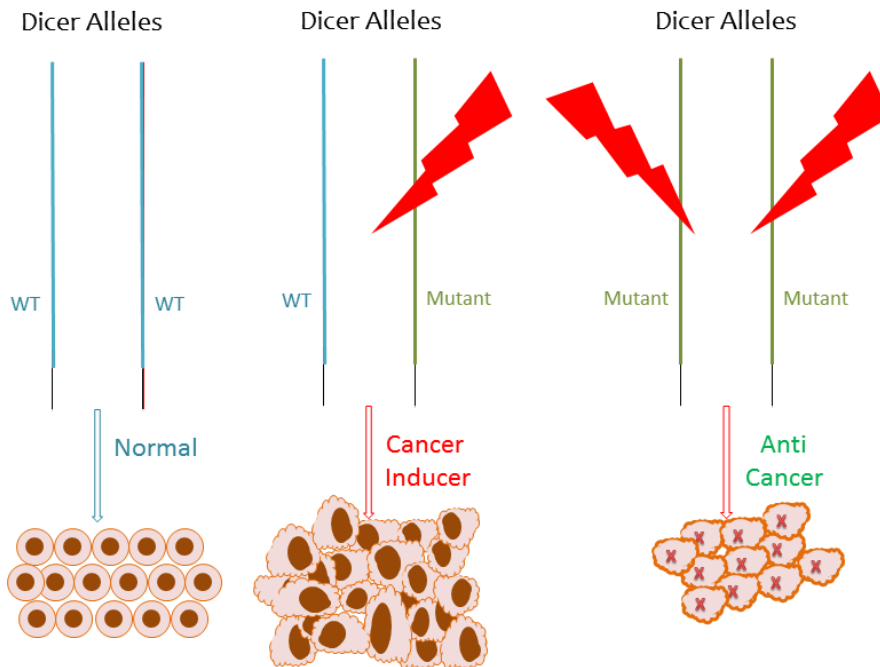


Figure 2. Haploinsufficiency of Dicer in cancer.

Reduction of Dicer is also correlated to advanced tumor stage, poor prognosis and lower survival [183]. Low Dicer reduces the cells' ability to process siRNAs [183]. DNA sequencing of *DICER* showed several mutations from tumor samples but there is not enough evidence to prove that *DICER* mutations are associated with the reduction of Dicer mRNA expression [183]. In endometrioid endometrial cancer, lower Dicer transcription is associated with disease recurrence and worse disease free survival [210]. The cause of low Dicer expression in this type of cancer was neither gene deletion nor DNA methylation and still remains unknown [210]. A study on chronic lymphocytic leukemia, the most common leukemia, showed that lower Dicer expression is associated with more aggressive cancer stages based on Binet

clarification and higher prognostic factors, such as CD38 and ZAP-70 [211]. Furthermore, a reduction of Dicer is also believed to be one indicator of shorter overall survival and lower disease free survival [211].

Dicer is over-expressed in many human breast cancer cells but the pattern is not consistent across all breast cancer subtypes [212]. High expression of Dicer increases breast cancer resistant protein, which effluxes tamoxifen from cells and increases resistance to tamoxifen treatment [212]. Dicer expression is reported to be higher in triple-negative breast cancer, another type of breast cancer with very poor prognosis due in part to its absence of drug targets (e.g., ER, PR and HER2 receptors) [181]. Interestingly, high concentrations of Dicer protein in triple-negative breast cancer samples are detected in the nuclear compartment while in normal breast tissue it is located mainly in the cytosol [181]. Thus, it is hypothesized that in triple-negative breast cancer, there are changes in the subcellular localization mechanisms of Dicer [181]. On the other hand, Dicer tends to be down-regulated in the non-luminal (ER negative) subgroup of breast cancer through correlations with high histological grade, lack of Bcl-2, high proliferation and expression of basal-like markers [213]. Basal-like breast cancer is associated with high grade, poor prognosis and younger patient age, and is identified by specific markers, such as EGFR, CKs, CAV1, CAV2 and nestin [214, 215]. The loss of Dicer is linked with breast cancer malignancy and a possible cause of miRNAs down-regulation in breast cancer [216, 217]. However, low Dicer expression does not affect the outcomes of breast cancer adjuvant anthracyclin-based therapy [213]. In lung cancer, Dicer is reduced in invasive adenocarcinoma but over-expressed in non-invasive precursor lesions (atypical adenomatous hyperplasia and bronchioloalveolar carcinoma) areas [218]. It is suggested that the up-regulation of Dicer can be an early sign of lung peripheral adenocarcinomas [218]. In another lung cancer study, reduced expression of Dicer was reported to be associated with poor prognosis and a shorter postoperative survival [219]. The cause of Dicer reduction was not known but it appears to be another mechanism than DNA methylation in the promoter region of *DICER* [219].

Table 2. Dicer alterations in cancers

Types of cancer	Regulation of Dicer	References
Ovarian	Down	[183]
Endometrial	Down	[210]
Breast	Down/Up	[181, 212, 213]
Leukemia	Down	[211]
Lung	Down	[218, 219]
Colorectal	Up	[220]
Prostate	Up	[224]
Melanoma	Up	[226]

In contrast, in certain types of cancers, such as colorectal cancer, high expression of Dicer is associated with poor overall survival and low disease free survival [220]. Since a decrease in overall miRNAs is believed to happen frequently in more aggressive tumors, dysfunction of alternative miRNA processing pathways might be an explanation for the occurrence of up-regulated Dicer in certain types of cancer [220-223]. In prostate adenocarcinoma, overall miRNAs expression is increased and Dicer levels increase as a correlation with disease stage, lymph node status and Gleason score [224]. Since it is believed that luminal cells tend to be the

origin of prostate cancer, the observance that Dicer is overexpressed in neoplastic luminal prostate cells suggested that high Dicer level might be an early sign of cancer development, probably facilitating the overall up-regulation of miRNAs [224, 225]. Also in cutaneous melanoma there is an up-regulation of Dicer mRNA in tumor cells, especially metastatic melanoma specimens, which suggests that Dicer overexpression may be a biomarker of cutaneous metastatic melanoma [226].

Regulation of Dicer

Currently it is still unknown exactly what or how Dicer is regulated. However, evidence suggests that the change in Dicer expression levels could be a consequence of feedback loops in miRNA processing machinery. MicroRNAs are an important factor that regulates DICER. For example, let-7 directly down-regulates Dicer in multiple normal and cancer cell lines, both at the mRNA and protein level [227]. Sequencing results showed that members of the let-7 family are complementary to the 3' UTR region of DICER [227]. In fact, Dicer regulation is considered an intermediate step for let-7 to regulate other miRNA expression [227]. This result requires further investigation on whether let-7 acts as tumor suppressor gene through Dicer down-regulation.

Another miRNA family that directly targets Dicer is the miR-103/107 family [228]. Similar to the let-7 family, this direct targeting attenuates miRNA biosynthesis [228]. In contrast to the let-7 family, the miR-103/107 family is a biomarker for worse prognosis in breast cancer [228]. Since they are oncogenic miRs, Dicer down-regulation is an intermediate step before these miRNAs are capable of exerting oncogenic effects. For example, miR-103/107 can trigger the events of metastasis or epithelial to mesenchymal transition (EMT) [228]. In others words, in spite of the complicated expression patter of Dicer in breast cancer, it is the miR-103/107 family, not DICER, that should be targeted for cancer therapy.

Exportin-5, the factor responsible for miRNA transport across the nuclear membrane, also regulates Dicer post-transcriptionally [229]. Inhibition of Exportin-5 leads to the accumulation of Dicer mRNA in the nucleus and ultimately the reduction of cytosolic DICER mRNA and functional DICER protein [229]. Furthermore, an increase in pre-miRNA saturates Exportin-5, thus preventing Dicer mRNA transport to the cytoplasm for translation [229]. This event is suggested to be a cross-regulation mechanism in miRNA biosynthesis with the purpose of maintaining miRNA homeostasis inside the cells [229].

Beyond miRNA processing machineries, several different factors are shown to regulate Dicer. One study of multiple cell lines with different Dicer expression levels showed that Dicer protein is repressed by reactive oxygen species, phorbol esters, the Ras oncogene, Type 1 interferon and double-stranded RNAs [230]. These data suggest that Dicer has a role in the cellular stress response and proposed that interferons are regulators of Dicer proteins [230]. Metformin, a medication for diabetes, increases Dicer mRNA and thus protein expression levels [231]. Indeed, metformin enhances the binding of transcription factor E2F3 and inhibits the transcriptional repressor E2F5 at the promoter of the Dicer gene [231]. Metformin exerts its anticancer effects by up-regulating Dicer and multiple miRNAs, many of which are involved in metabolism [231].

Interestingly, a recent study reported that Dicer has other functions beyond miRNA microbiogenesis. In fact, Dicer processes transcripts from RD2 Alu repeats into small RNAs

(28-65nt), which will target critical stem-cell RNAs, including Nanog mRNA [232]. These data suggest that our understanding of this multifaceted protein is incomplete; further research is needed to identify the complex nature of Dicer.

Argonautes

After being processed by Dicer, the miRNA:miRNA* duplex is loaded into the RISC (RNA-induced silencing complex) [233-235]. MicroRNA* is primarily released (although in some systems it may exert significant biological effects) while miRNA is used to target mRNA for cleavage [236, 237]. Argonaute proteins (AGO1-4) are essential components of the RISC complex. Even though all four AGOs can repress mRNA expression, only AGO2 plays an essential role in mRNA cleavage [238, 239]. AGO1 and AGO2 are shown to up-regulated in serous ovarian cancer [182]. Furthermore, promoting expression and activation of AGO2 has positive effects on cancer by enhancing multiple tumor-suppressive miRNAs [240].

REFERENCES

- [1] Lu J, Getz G, Miska EA, Alvarez-Saavedra E, Lamb J, Peck D, et al. MicroRNA expression profiles classify human cancers. *Nature*. 2005;435:834-8.
- [2] Zhang BH, Pan XP, Cobb GP, Anderson TA. microRNAs as oncogenes and tumor suppressors. *Developmental Biology*. 2007;302:1-12.
- [3] Esquela-Kerscher A, Slack FJ. Oncomirs - microRNAs with a role in cancer. *Nature Reviews Cancer*. 2006;6:259-69.
- [4] Kasinski AL, Slack FJ. Epigenetics and genetics. MicroRNAs en route to the clinic: progress in validating and targeting microRNAs for cancer therapy. *Nat. Rev. Cancer*. 2011;11:849-64.
- [5] Cummins JM, He Y, Leary RJ, Pagliarini R, Diaz LA, Jr., Sjoblom T, et al. The colorectal microRNAome. *Proc. Natl. Acad. Sci. USA*. 2006;103:3687-92.
- [6] Bartel DP. MicroRNAs: Genomics, biogenesis, mechanism, and function. *Cell*. 2004;116:281-97.
- [7] Filipowicz W, Bhattacharyya SN, Sonenberg N. Mechanisms of post-transcriptional regulation by microRNAs: are the answers in sight? *Nature Reviews Genetics*. 2008;9:102-14.
- [8] Lu J, Getz G, Miska EA, Alvarez-Saavedra E, Lamb J, Peck D, et al. MicroRNA expression profiles classify human cancers. *Nature*. 2005;435:834-8.
- [9] Volinia S, Calin GA, Liu CG, Ambs S, Cimmino A, Petrocca F, et al. A microRNA expression signature of human solid tumors defines cancer gene targets. *Proc. Natl. Acad. Sci. USA*. 2006;103:2257-61.
- [10] Garzon R, Pichiorri F, Palumbo T, Iuliano R, Cimmino A, Aqeilan R, et al. MicroRNA fingerprints during human megakaryocytopoiesis. *Proc. Natl. Acad. Sci. USA*. 2006;103:5078-83.

- [11] Calin GA, Liu CG, Sevignani C, Ferracin M, Felli N, Dumitru CD, et al. MicroRNA profiling reveals distinct signatures in B cell chronic lymphocytic leukemias. *Proc. Natl. Acad. Sci. USA*. 2004;101:11755-60.
- [12] Iorio MV, Ferracin M, Liu CG, Veronese A, Spizzo R, Sabbioni S, et al. MicroRNA gene expression deregulation in human breast cancer. *Cancer Res*. 2005;65:7065-70.
- [13] Ciafre SA, Galardi S, Mangiola A, Ferracin M, Liu CG, Sabatino G, et al. Extensive modulation of a set of microRNAs in primary glioblastoma. *Biochem. Biophys. Res. Commun*. 2005;334:1351-8.
- [14] Murakami Y, Yasuda T, Saigo K, Urashima T, Toyoda H, Okanoue T, et al. Comprehensive analysis of microRNA expression patterns in hepatocellular carcinoma and non-tumorous tissues. *Oncogene*. 2006;25:2537-45.
- [15] Roldo C, Missiaglia E, Hagan JP, Falconi M, Capelli P, Bersani S, et al. MicroRNA expression abnormalities in pancreatic endocrine and acinar tumors are associated with distinctive pathologic features and clinical behavior. *J. Clin. Oncol*. 2006;24:4677-84.
- [16] Xu K, Liang X, Cui D, Wu Y, Shi W, Liu J. miR-1915 inhibits Bcl-2 to modulate multidrug resistance by increasing drug-sensitivity in human colorectal carcinoma cells. *Mol. Carcinog*. 2013;52:70-8.
- [17] Pouliot LM, Chen YC, Bai J, Guha R, Martin SE, Gottesman MM, et al. Cisplatin sensitivity mediated by WEE1 and CHK1 is mediated by miR-155 and the miR-15 family. *Cancer Res*. 2012;72:5945-55.
- [18] Salim H, Akbar NS, Zong D, Vaculova AH, Lewensohn R, Moshfegh A, et al. miRNA-214 modulates radiotherapy response of non-small cell lung cancer cells through regulation of p38MAPK, apoptosis and senescence. *Br. J. Cancer*. 2012;107:1361-73.
- [19] Majid S, Dar AA, Saini S, Shahryari V, Arora S, Zaman MS, et al. miRNA-34b Inhibits Prostate Cancer through Demethylation, Active Chromatin Modifications, and AKT Pathways. *Clin. Cancer Res*. 2013;19:73-84.
- [20] Li L, Xie X, Luo J, Liu M, Xi S, Guo J, et al. Targeted expression of miR-34a using the T-VISA system suppresses breast cancer cell growth and invasion. *Mol. Ther*. 2012;20:2326-34.
- [21] Liu ZL, Wang H, Liu J, Wang ZX. MicroRNA-21 (miR-21) expression promotes growth, metastasis, and chemo- or radioresistance in non-small cell lung cancer cells by targeting PTEN. *Mol. Cell Biochem*. 2013;372:35-45.
- [22] Geary RS, Watanabe TA, Truong L, Freier S, Lesnik EA, Sioufi NB, et al. Pharmacokinetic properties of 2'-O-(2-methoxyethyl)-modified oligonucleotide analogs in rats. *J. Pharmacol. Exp. Ther*. 2001;296:890-7.
- [23] Henry SP, Geary RS, Yu R, Levin AA. Drug properties of second-generation antisense oligonucleotides: how do they measure up to their predecessors? *Curr. Opin Investig. Drugs*. 2001;2:1444-9.
- [24] Levin AA. A review of the issues in the pharmacokinetics and toxicology of phosphorothioate antisense oligonucleotides. *Biochim. Biophys. Acta*. 1999;1489:69-84.
- [25] Henry SP, Zuckerman JE, Rojko J, Hall WC, Harman RJ, Kitchen D, et al. Toxicological properties of several novel oligonucleotide analogs in mice. *Anticancer Drug Des*. 1997;12:1-14.
- [26] Henry SP, Monteith D, Levin AA. Antisense oligonucleotide inhibitors for the treatment of cancer: 2. Toxicological properties of phosphorothioate oligodeoxynucleotides. *Anticancer Drug Des*. 1997;12:395-408.

- [27] Bijsterbosch MK, Rump ET, De Vruh RL, Dorland R, van Veghel R, Tivel KL, et al. Modulation of plasma protein binding and in vivo liver cell uptake of phosphorothioate oligodeoxynucleotides by cholesterol conjugation. *Nucleic Acids Res.* 2000;28:2717-25.
- [28] Krutzfeldt J, Rajewsky N, Braich R, Rajeev KG, Tuschl T, Manoharan M, et al. Silencing of microRNAs in vivo with 'antagomirs'. *Nature.* 2005;438:685-9.
- [29] Yang M, Mattes J. Discovery, biology and therapeutic potential of RNA interference, microRNA and antagomirs. *Pharmacol. Ther.* 2008;117:94-104.
- [30] Petersen M, Wengel J. LNA: a versatile tool for therapeutics and genomics. *Trends Biotechnol.* 2003;21:74-81.
- [31] Chabot S, Orio J, Castanier R, Bellard E, Nielsen SJ, Golzio M, et al. LNA-based oligonucleotide electrotransfer for miRNA inhibition. *Mol. Ther.* 2012;20:1590-8.
- [32] Obad S, dos Santos CO, Petri A, Heidenblad M, Broom O, Ruse C, et al. Silencing of microRNA families by seed-targeting tiny LNAs. *Nat. Genet.* 2011;43:371-8.
- [33] Matsubara H, Takeuchi T, Nishikawa E, Yanagisawa K, Hayashita Y, Ebi H, et al. Apoptosis induction by antisense oligonucleotides against miR-17-5p and miR-20a in lung cancers overexpressing miR-17-92. *Oncogene.* 2007;26:6099-105.
- [34] Elmen J, Lindow M, Schutz S, Lawrence M, Petri A, Obad S, et al. LNA-mediated microRNA silencing in non-human primates. *Nature.* 2008;452:896-9.
- [35] Elmen J, Lindow M, Silaharoglu A, Bak M, Christensen M, Lind-Thomsen A, et al. Antagonism of microRNA-122 in mice by systemically administered LNA-antimiR leads to up-regulation of a large set of predicted target mRNAs in the liver. *Nucleic Acids Res.* 2008;36:1153-62.
- [36] Lanford RE, Hildebrandt-Eriksen ES, Petri A, Persson R, Lindow M, Munk ME, et al. Therapeutic silencing of microRNA-122 in primates with chronic hepatitis C virus infection. *Science.* 2010;327:198-201.
- [37] Mitchell PS, Parkin RK, Kroh EM, Fritz BR, Wyman SK, Pogosova-Agadjanyan EL, et al. Circulating microRNAs as stable blood-based markers for cancer detection. *Proc. Natl. Acad. Sci. USA.* 2008;105:10513-8.
- [38] Arroyo JD, Chevillet JR, Kroh EM, Ruf IK, Pritchard CC, Gibson DF, et al. Argonaute2 complexes carry a population of circulating microRNAs independent of vesicles in human plasma. *Proc. Natl. Acad. Sci. USA.* 2011;108:5003-8.
- [39] Vickers KC, Palmisano BT, Shoucri BM, Shamburek RD, Remaley AT. MicroRNAs are transported in plasma and delivered to recipient cells by high-density lipoproteins. *Nat. Cell Biol.* 2011;13:423-33.
- [40] Layzer JM, McCaffrey AP, Tanner AK, Huang Z, Kay MA, Sullenger BA. In vivo activity of nuclease-resistant siRNAs. *RNA.* 2004;10:766-71.
- [41] Tseng YC, Mozumdar S, Huang L. Lipid-based systemic delivery of siRNA. *Adv. Drug Deliv. Rev.* 2009;61:721-31.
- [42] Kumar MS, Erkeland SJ, Pester RE, Chen CY, Ebert MS, Sharp PA, et al. Suppression of non-small cell lung tumor development by the let-7 microRNA family. *Proc. Natl. Acad. Sci. USA.* 2008;105:3903-8.
- [43] Kota J, Chivukula RR, O'Donnell KA, Wentzel EA, Montgomery CL, Hwang HW, et al. Therapeutic microRNA delivery suppresses tumorigenesis in a murine liver cancer model. *Cell.* 2009;137:1005-17.
- [44] Klibanov AL, Maruyama K, Torchilin VP, Huang L. Amphipathic polyethyleneglycols effectively prolong the circulation time of liposomes. *FEBS Lett.* 1990;268:235-7.

- [45] Wu Y, Crawford M, Yu B, Mao Y, Nana-Sinkam SP, Lee LJ. MicroRNA delivery by cationic lipoplexes for lung cancer therapy. *Mol. Pharm.* 2011;8:1381-9.
- [46] Park JW, Kirpotin DB, Hong K, Shalaby R, Shao Y, Nielsen UB, et al. Tumor targeting using anti-her2 immunoliposomes. *J. Control Release.* 2001;74:95-113.
- [47] Chen Y, Zhu X, Zhang X, Liu B, Huang L. Nanoparticles modified with tumor-targeting scFv deliver siRNA and miRNA for cancer therapy. *Mol. Ther.* 2010;18:1650-6.
- [48] Baigude H, McCarroll J, Yang CS, Swain PM, Rana TM. Design and creation of new nanomaterials for therapeutic RNAi. *ACS Chem. Biol.* 2007;2:237-41.
- [49] Su J, Baigude H, McCarroll J, Rana TM. Silencing microRNA by interfering nanoparticles in mice. *Nucleic Acids Res.* 2011;39:e38.
- [50] Cheng CJ, Saltzman WM. Polymer nanoparticle-mediated delivery of microRNA inhibition and alternative splicing. *Mol. Pharm.* 2012;9:1481-8.
- [51] Daniel MC, Astruc D. Gold nanoparticles: assembly, supramolecular chemistry, quantum-size-related properties, and applications toward biology, catalysis, and nanotechnology. *Chem. Rev.* 2004;104:293-346.
- [52] Rosi NL, Giljohann DA, Thaxton CS, Lytton-Jean AK, Han MS, Mirkin CA. Oligonucleotide-modified gold nanoparticles for intracellular gene regulation. *Science.* 2006;312:1027-30.
- [53] Lee SH, Bae KH, Kim SH, Lee KR, Park TG. Amine-functionalized gold nanoparticles as non-cytotoxic and efficient intracellular siRNA delivery carriers. *International Journal of Pharmaceutics.* 2008;364:94-101.
- [54] Storhoff JJ, Elghanian R, Mucic RC, Mirkin CA, Letsinger RL. One-pot colorimetric differentiation of polynucleotides with single base imperfections using gold nanoparticle probes. *Journal of the American Chemical Society.* 1998;120:1959-64.
- [55] Prencipe G, Tabakman SM, Welsher K, Liu Z, Goodwin AP, Zhang L, et al. PEG branched polymer for functionalization of nanomaterials with ultralong blood circulation. *J. Am. Chem. Soc.* 2009;131:4783-7.
- [56] Ghosh R, Singh LC, Shohet JM, Gunaratne PH. A gold nanoparticle platform for the delivery of functional microRNAs into cancer cells. *Biomaterials.* 2013;34:807-16.
- [57] Crew E, Tessel MA, Rahman S, Razzak-Jaffar A, Mott D, Kamundi M, et al. MicroRNA conjugated gold nanoparticles and cell transfection. *Anal Chem.* 2012;84:26-9.
- [58] Kim JH, Yeom JH, Ko JJ, Han MS, Lee K, Na SY, et al. Effective delivery of anti-miRNA DNA oligonucleotides by functionalized gold nanoparticles. *J. Biotechnol.* 2011;155:287-92.
- [59] Lee Y, Kim M, Han JJ, Yeom KH, Lee S, Baek SH, et al. MicroRNA genes are transcribed by RNA polymerase II. *Embo Journal.* 2004;23:4051-60.
- [60] Lee O, Kim VN. Evidence that microRNA genes are transcribed by RNA polymerase II. *Cell Structure and Function.* 2004;29:68-.
- [61] Cai XZ, Hagedorn CH, Cullen BR. Human microRNAs are processed from capped, polyadenylated transcripts that can also function as mRNAs. *RNA-a Publication of the RNA Society.* 2004;10:1957-66.
- [62] Denli AM, Tops BBJ, Plasterk RHA, Ketting RF, Hannon GJ. Processing of primary microRNAs by the Microprocessor complex. *Nature.* 2004;432:231-5.
- [63] Bohnsack MT, Czaplinski K, Gorlich D. Exportin 5 is a RanGTP-dependent dsRNA-binding protein that mediates nuclear export of pre-miRNAs. *RNA-a Publication of the RNA Society.* 2004;10:185-91.

- [64] Hutvagner G, McLachlan J, Pasquinelli AE, Balint E, Tuschl T, Zamore PD. A cellular function for the RNA-interference enzyme DICER in the maturation of the let-7 small temporal RNA. *Science*. 2001;293:834-8.
- [65] Ketting RF, Fischer SEJ, Bernstein E, Sijen T, Hannon GJ, Plasterk RHA. DICER functions in RNA interference and in synthesis of small RNA involved in developmental timing in *C-elegans*. *Genes Dev*. 2001;15:2654-9.
- [66] Gregory RI, Chendrimada TP, Cooch N, Shiekhattar R. Human RISC couples microRNA biogenesis and posttranscriptional gene silencing. *Cell*. 2005;123:631-40.
- [67] Lin SL, Chang D, Ying SY. Asymmetry of intronic pre-miRNA structures in functional RISC assembly. *Gene*. 2005;356:32-8.
- [68] Sikand K, Slaibi JE, Singh R, Slane SD, Shukla GC. miR 488* inhibits androgen receptor expression in prostate carcinoma cells. *Int. J. Cancer*. 2011;129:810-9.
- [69] Eichner LJ, Perry MC, Dufour CR, Bertos N, Park M, St-Pierre J, et al. miR-378(*) mediates metabolic shift in breast cancer cells via the PGC-1beta/ERRgamma transcriptional pathway. *Cell Metab*. 2010;12:352-61.
- [70] Jia W, Eneh JO, Ratnaparkhe S, Altman MK, Murph MM. MicroRNA-30c-2* expressed in ovarian cancer cells suppresses growth factor-induced cellular proliferation and downregulates the oncogene BCL9. *Mol. Cancer Res*. 2011;9:1732-45.
- [71] Reinhart BJ, Slack FJ, Basson M, Pasquinelli AE, Bettinger JC, Rougvie AE, et al. The 21-nucleotide let-7 RNA regulates developmental timing in *Caenorhabditis elegans*. *Nature*. 2000;403:901-6.
- [72] Lee RC, Feinbaum RL, Ambros V. The *C. elegans* heterochronic gene *lin-4* encodes small RNAs with antisense complementarity to *lin-14*. *Cell*. 1993;75:843-54.
- [73] Wightman B, Ha I, Ruvkun G. Posttranscriptional regulation of the heterochronic gene *lin-14* by *lin-4* mediates temporal pattern formation in *C. elegans*. *Cell*. 1993;75:855-62.
- [74] Pasquinelli AE, Reinhart BJ, Slack F, Martindale MQ, Kuroda MI, Maller B, et al. Conservation of the sequence and temporal expression of let-7 heterochronic regulatory RNA. *Nature*. 2000;408:86-9.
- [75] Ruby JG, Jan C, Player C, Axtell MJ, Lee W, Nusbaum C, et al. Large-scale sequencing reveals 21U-RNAs and additional microRNAs and endogenous siRNAs in *C. elegans*. *Cell*. 2006;127:1193-207.
- [76] Schulman BR, Esquela-Kerscher A, Slack FJ. Reciprocal expression of *lin-41* and the microRNAs *let-7* and *mir-125* during mouse embryogenesis. *Dev. Dyn*. 2005;234:1046-54.
- [77] Wulczyn FG, Smirnova L, Rybak A, Brandt C, Kwidzinski E, Ninnemann O, et al. Post-transcriptional regulation of the let-7 microRNA during neural cell specification. *FASEB J*. 2007;21:415-26.
- [78] Suh MR, Lee Y, Kim JY, Kim SK, Moon SH, Lee JY, et al. Human embryonic stem cells express a unique set of microRNAs. *Dev. Biol*. 2004;270:488-98.
- [79] Sempere LF, Freemantle S, Pitha-Rowe I, Moss E, Dmitrovsky E, Ambros V. Expression profiling of mammalian microRNAs uncovers a subset of brain-expressed microRNAs with possible roles in murine and human neuronal differentiation. *Genome Biol*. 2004;5:R13.
- [80] Ibarra I, Erlich Y, Muthuswamy SK, Sachidanandam R, Hannon GJ. A role for microRNAs in maintenance of mouse mammary epithelial progenitor cells. *Genes Dev*. 2007;21:3238-43.

- [81] Bussing I, Slack FJ, Grosshans H. let-7 microRNAs in development, stem cells and cancer. *Trends Mol. Med.* 2008;14:400-9.
- [82] Shields JM, Pruitt K, McFall A, Shaub A, Der CJ. Understanding Ras: 'it ain't over 'til it's over'. *Trends Cell Biol.* 2000;10:147-54.
- [83] Bonni A, Brunet A, West AE, Datta SR, Takasu MA, Greenberg ME. Cell survival promoted by the Ras-MAPK signaling pathway by transcription-dependent and -independent mechanisms. *Science.* 1999;286:1358-62.
- [84] Pruitt K, Der CJ. Ras and Rho regulation of the cell cycle and oncogenesis. *Cancer Lett.* 2001;171:1-10.
- [85] Almoguera C, Shibata D, Forrester K, Martin J, Arnheim N, Perucho M. Most human carcinomas of the exocrine pancreas contain mutant c-K-ras genes. *Cell.* 1988;53:549-54.
- [86] Shibata D, Capella G, Perucho M. Mutational activation of the c-K-ras gene in human pancreatic carcinoma. *Baillieres Clin. Gastroenterol.* 1990;4:151-69.
- [87] Bos JL, Fearon ER, Hamilton SR, Verlaan-de Vries M, van Boom JH, van der Eb AJ, et al. Prevalence of ras gene mutations in human colorectal cancers. *Nature.* 1987;327:293-7.
- [88] Ahrendt SA, Decker PA, Alawi EA, Zhu Yr YR, Sanchez-Cespedes M, Yang SC, et al. Cigarette smoking is strongly associated with mutation of the K-ras gene in patients with primary adenocarcinoma of the lung. *Cancer.* 2001;92:1525-30.
- [89] Johnson L, Mercer K, Greenbaum D, Bronson RT, Crowley D, Tuveson DA, et al. Somatic activation of the K-ras oncogene causes early onset lung cancer in mice. *Nature.* 2001;410:1111-6.
- [90] Takamizawa J, Konishi H, Yanagisawa K, Tomida S, Osada H, Endoh H, et al. Reduced expression of the let-7 microRNAs in human lung cancers in association with shortened postoperative survival. *Cancer Res.* 2004;64:3753-6.
- [91] Calin GA, Sevignani C, Dan Dumitru C, Hyslop T, Noch E, Yendamuri S, et al. Human microRNA genes are frequently located at fragile sites and genomic regions involved in cancers. *Proc. Natl. Acad. Sci. USA.* 2004;101:2999-3004.
- [92] Johnson SM, Grosshans H, Shingara J, Byrom M, Jarvis R, Cheng A, et al. RAS is regulated by the let-7 microRNA family. *Cell.* 2005;120:635-47.
- [93] Trang P, Medina PP, Wiggins JF, Ruffino L, Kelnar K, Omotola M, et al. Regression of murine lung tumors by the let-7 microRNA. *Oncogene.* 2010;29:1580-7.
- [94] Lee YS, Dutta A. The tumor suppressor microRNA let-7 represses the HMGA2 oncogene. *Genes Dev.* 2007;21:1025-30.
- [95] Reeves R. Molecular biology of HMGA proteins: hubs of nuclear function. *Gene.* 2001;277:63-81.
- [96] Zhou X, Benson KF, Ashar HR, Chada K. Mutation responsible for the mouse pygmy phenotype in the developmentally regulated factor HMGI-C. *Nature.* 1995;376:771-4.
- [97] Rogalla P, Drechsler K, Frey G, Hennig Y, Helmke B, Bonk U, et al. HMGI-C expression patterns in human tissues. Implications for the genesis of frequent mesenchymal tumors. *Am. J. Pathol.* 1996;149:775-9.
- [98] Berner JM, Meza-Zepeda LA, Kools PF, Forus A, Schoenmakers EF, Van de Ven WJ, et al. HMGI-C, the gene for an architectural transcription factor, is amplified and rearranged in a subset of human sarcomas. *Oncogene.* 1997;14:2935-41.

- [99] Sarhadi VK, Wikman H, Salmenkivi K, Kuosma E, Sioris T, Salo J, et al. Increased expression of high mobility group A proteins in lung cancer. *J. Pathol.* 2006;209:206-12.
- [100] Mayr C, Hemann MT, Bartel DP. Disrupting the pairing between let-7 and Hmga2 enhances oncogenic transformation. *Science.* 2007;315:1576-9.
- [101] Hebert C, Norris K, Scheper MA, Nikitakis N, Sauk JJ. High mobility group A2 is a target for miRNA-98 in head and neck squamous cell carcinoma. *Mol Cancer.* 2007;6:5.
- [102] Yu F, Yao H, Zhu P, Zhang X, Pan Q, Gong C, et al. let-7 regulates self renewal and tumorigenicity of breast cancer cells. *Cell.* 2007;131:1109-23.
- [103] Johnson CD, Esquela-Kerscher A, Stefani G, Byrom M, Kelnar K, Ovcharenko D, et al. The let-7 microRNA represses cell proliferation pathways in human cells. *Cancer Res.* 2007;67:7713-22.
- [104] Galaktionov K, Lee AK, Eckstein J, Draetta G, Meckler J, Loda M, et al. CDC25 phosphatases as potential human oncogenes. *Science.* 1995;269:1575-7.
- [105] Huang JC, Babak T, Corson TW, Chua G, Khan S, Gallie BL, et al. Using expression profiling data to identify human microRNA targets. *Nat. Methods.* 2007;4:1045-9.
- [106] Schultz J, Lorenz P, Gross G, Ibrahim S, Kunz M. MicroRNA let-7b targets important cell cycle molecules in malignant melanoma cells and interferes with anchorage-independent growth. *Cell Res.* 2008;18:549-57.
- [107] Viswanathan SR, Daley GQ, Gregory RI. Selective blockade of microRNA processing by LIN28. *Science.* 2008;320:97-100.
- [108] Newman MA, Thomson JM, Hammond SM. Lin-28 interaction with the Let-7 precursor loop mediates regulated microRNA processing. *RNA.* 2008;14:1539-49.
- [109] Piskounova E, Viswanathan SR, Janas M, LaPierre RJ, Daley GQ, Sliz P, et al. Determinants of microRNA processing inhibition by the developmentally regulated RNA-binding protein LIN28. *J. Biol. Chem.* 2008;283:21310-4.
- [110] Nam Y, Chen C, Gregory RI, Chou JJ, Sliz P. Molecular basis for interaction of let-7 microRNAs with LIN28. *Cell.* 2011;147:1080-91.
- [111] Heo I, Joo C, Cho J, Ha M, Han J, Kim VN. LIN28 mediates the terminal uridylation of let-7 precursor MicroRNA. *Mol. Cell.* 2008;32:276-84.
- [112] Heo I, Joo C, Kim YK, Ha M, Yoon MJ, Cho J, et al. TUT4 in concert with LIN28 suppresses microRNA biogenesis through pre-microRNA uridylation. *Cell.* 2009;138:696-708.
- [113] Hagan JP, Piskounova E, Gregory RI. LIN28 recruits the TUTase Zcchc11 to inhibit let-7 maturation in mouse embryonic stem cells. *Nat Struct Mol. Biol.* 2009;16:1021-5.
- [114] Thornton JE, Chang HM, Piskounova E, Gregory RI. LIN28-mediated control of let-7 microRNA expression by alternative TUTases Zcchc11 (TUT4) and Zcchc6 (TUT7). *RNA.* 2012;18:1875-85.
- [115] Viswanathan SR, Powers JT, Einhorn W, Hoshida Y, Ng TL, Toffanin S, et al. LIN28 promotes transformation and is associated with advanced human malignancies. *Nat. Genet.* 2009;41:843-8.
- [116] Richards M, Tan SP, Tan JH, Chan WK, Bongso A. The transcriptome profile of human embryonic stem cells as defined by SAGE. *Stem. Cells.* 2004;22:51-64.
- [117] Pan L, Gong Z, Zhong Z, Dong Z, Liu Q, Le Y, et al. Lin-28 reactivation is required for let-7 repression and proliferation in human small cell lung cancer cells. *Mol. Cell Biochem.* 2011;355:257-63.

- [118] Sakurai M, Miki Y, Masuda M, Hata S, Shibahara Y, Hirakawa H, et al. LIN28: a regulator of tumor-suppressing activity of let-7 microRNA in human breast cancer. *J. Steroid Biochem. Mol. Biol.* 2012;131:101-6.
- [119] Yu J, Vodyanik MA, Smuga-Otto K, Antosiewicz-Bourget J, Frane JL, Tian S, et al. Induced pluripotent stem cell lines derived from human somatic cells. *Science*. 2007;318:1917-20.
- [120] Moss EG, Tang L. Conservation of the heterochronic regulator Lin-28, its developmental expression and microRNA complementary sites. *Dev. Biol.* 2003;258:432-42.
- [121] Rybak A, Fuchs H, Smirnova L, Brandt C, Pohl EE, Nitsch R, et al. A feedback loop comprising lin-28 and let-7 controls pre-let-7 maturation during neural stem-cell commitment. *Nat. Cell Biol.* 2008;10:987-93.
- [122] Viswanathan SR, Daley GQ. LIN28: A microRNA regulator with a macro role. *Cell*. 2010;140:445-9.
- [123] Takahashi K, Okita K, Nakagawa M, Yamanaka S. Induction of pluripotent stem cells from fibroblast cultures. *Nat. Protoc.* 2007;2:3081-9.
- [124] Park IH, Zhao R, West JA, Yabuuchi A, Huo H, Ince TA, et al. Reprogramming of human somatic cells to pluripotency with defined factors. *Nature*. 2008;451:141-6.
- [125] McCormick F. Signalling networks that cause cancer. *Trends Cell Biol.* 1999;9:M53-6.
- [126] Vogelstein B, Kinzler KW. The multistep nature of cancer. *Trends Genet.* 1993;9:138-41.
- [127] Vaux DL, Cory S, Adams JM. Bcl-2 gene promotes haemopoietic cell survival and cooperates with c-myc to immortalize pre-B cells. *Nature*. 1988;335:440-2.
- [128] Hartwell LH, Kastan MB. Cell cycle control and cancer. *Science*. 1994;266:1821-8.
- [129] Chang TC, Yu D, Lee YS, Wentzel EA, Arking DE, West KM, et al. Widespread microRNA repression by Myc contributes to tumorigenesis. *Nat. Genet.* 2008;40:43-50.
- [130] Chang TC, Zeitels LR, Hwang HW, Chivukula RR, Wentzel EA, Dews M, et al. Lin-28B transactivation is necessary for Myc-mediated let-7 repression and proliferation. *Proc. Natl. Acad. Sci. USA*. 2009;106:3384-9.
- [131] Sampson VB, Rong NH, Han J, Yang Q, Aris V, Soteropoulos P, et al. MicroRNA let-7a down-regulates MYC and reverts MYC-induced growth in Burkitt lymphoma cells. *Cancer Res.* 2007;67:9762-70.
- [132] Buechner J, Tomte E, Haug BH, Henriksen JR, Lokke C, Flaegstad T, et al. Tumour-suppressor microRNAs let-7 and mir-101 target the proto-oncogene MYCN and inhibit cell proliferation in MYCN-amplified neuroblastoma. *Br. J. Cancer*. 2011;105:296-303.
- [133] Koscianska E, Baev V, Skreka K, Oikonomaki K, Rusinov V, Tabler M, et al. Prediction and preliminary validation of oncogene regulation by miRNAs. *BMC Mol. Biol.* 2007;8:79.
- [134] Wong TS, Man OY, Tsang CM, Tsao SW, Tsang RK, Chan JY, et al. MicroRNA let-7 suppresses nasopharyngeal carcinoma cells proliferation through downregulating c-Myc expression. *J. Cancer Res. Clin. Oncol.* 2011;137:415-22.
- [135] Molenaar JJ, Domingo-Fernandez R, Ebus ME, Lindner S, Koster J, Drabek K, et al. LIN28B induces neuroblastoma and enhances MYCN levels via let-7 suppression. *Nat. Genet.* 2012;44:1199-206.
- [136] Ota A, Tagawa H, Karnan S, Tsuzuki S, Karpas A, Kira S, et al. Identification and characterization of a novel gene, C13orf25, as a target for 13q31-q32 amplification in malignant lymphoma. *Cancer Res.* 2004;64:3087-95.

- [137] Concepcion CP, Bonetti C, Ventura A. The microRNA-17-92 family of microRNA clusters in development and disease. *Cancer J.* 2012;18:262-7.
- [138] Mendell JT. miRiad roles for the miR-17-92 cluster in development and disease. *Cell.* 2008;133:217-22.
- [139] He L, Thomson JM, Hemann MT, Hernando-Monge E, Mu D, Goodson S, et al. A microRNA polycistron as a potential human oncogene. *Nature.* 2005;435:828-33.
- [140] Hayashita Y, Osada H, Tatematsu Y, Yamada H, Yanagisawa K, Tomida S, et al. A polycistronic microRNA cluster, miR-17-92, is overexpressed in human lung cancers and enhances cell proliferation. *Cancer Res.* 2005;65:9628-32.
- [141] Connolly E, Melegari M, Landgraf P, Tchaikovskaya T, Tennant BC, Slagle BL, et al. Elevated expression of the miR-17-92 polycistron and miR-21 in hepadnavirus-associated hepatocellular carcinoma contributes to the malignant phenotype. *Am. J. Pathol.* 2008;173:856-64.
- [142] Takakura S, Mitsutake N, Nakashima M, Namba H, Saenko VA, Rogounovitch TI, et al. Oncogenic role of miR-17-92 cluster in anaplastic thyroid cancer cells. *Cancer Sci.* 2008;99:1147-54.
- [143] Yan HL, Xue G, Mei Q, Wang YZ, Ding FX, Liu MF, et al. Repression of the miR-17-92 cluster by p53 has an important function in hypoxia-induced apoptosis. *EMBO J.* 2009;28:2719-32.
- [144] Lu Y, Thomson JM, Wong HY, Hammond SM, Hogan BL. Transgenic over-expression of the microRNA miR-17-92 cluster promotes proliferation and inhibits differentiation of lung epithelial progenitor cells. *Dev. Biol.* 2007;310:442-53.
- [145] Xiao C, Srinivasan L, Calado DP, Patterson HC, Zhang B, Wang J, et al. Lymphoproliferative disease and autoimmunity in mice with increased miR-17-92 expression in lymphocytes. *Nat. Immunol.* 2008;9:405-14.
- [146] Taguchi A, Yanagisawa K, Tanaka M, Cao K, Matsuyama Y, Goto H, et al. Identification of hypoxia-inducible factor-1 alpha as a novel target for miR-17-92 microRNA cluster. *Cancer Res.* 2008;68:5540-5.
- [147] Inomata M, Tagawa H, Guo YM, Kameoka Y, Takahashi N, Sawada K. MicroRNA-17-92 down-regulates expression of distinct targets in different B-cell lymphoma subtypes. *Blood.* 2009;113:396-402.
- [148] Ivanovska I, Ball AS, Diaz RL, Magnus JF, Kibukawa M, Schelter JM, et al. MicroRNAs in the miR-106b family regulate p21/CDKN1A and promote cell cycle progression. *Mol. Cell Biol.* 2008;28:2167-74.
- [149] O'Donnell KA, Wentzel EA, Zeller KI, Dang CV, Mendell JT. c-Myc-regulated microRNAs modulate E2F1 expression. *Nature.* 2005;435:839-43.
- [150] Sylvestre Y, De Guire V, Querido E, Mukhopadhyay UK, Bourdeau V, Major F, et al. An E2F/miR-20a autoregulatory feedback loop. *J. Biol. Chem.* 2007;282:2135-43.
- [151] Woods K, Thomson JM, Hammond SM. Direct regulation of an oncogenic micro-RNA cluster by E2F transcription factors. *J. Biol. Chem.* 2007;282:2130-4.
- [152] Bhattacharya R, Nicoloso M, Arvizo R, Wang E, Cortez A, Rossi S, et al. MiR-15a and MiR-16 control Bmi-1 expression in ovarian cancer. *Cancer Res.* 2009;69:9090-5.
- [153] Aqeilan RI, Calin GA, Croce CM. miR-15a and miR-16-1 in cancer: discovery, function and future perspectives. *Cell Death Differ.* 2010;17:215-20.

- [154] Cimmino A, Calin GA, Fabbri M, Iorio MV, Ferracin M, Shimizu M, et al. miR-15 and miR-16 induce apoptosis by targeting BCL2. *Proc. Natl. Acad. Sci. USA*. 2005;102:13944-9.
- [155] Wang B, Wang H, Yang Z. MiR-122 inhibits cell proliferation and tumorigenesis of breast cancer by targeting IGF1R. *PLoS One*. 2012;7:e47053.
- [156] Tsai WC, Hsu PW, Lai TC, Chau GY, Lin CW, Chen CM, et al. MicroRNA-122, a tumor suppressor microRNA that regulates intrahepatic metastasis of hepatocellular carcinoma. *Hepatology*. 2009;49:1571-82.
- [157] Kutay H, Bai S, Datta J, Motiwala T, Pogribny I, Frankel W, et al. Downregulation of miR-122 in the rodent and human hepatocellular carcinomas. *J Cell Biochem*. 2006;99:671-8.
- [158] Hu Y, Ou Y, Wu K, Chen Y, Sun W. miR-143 inhibits the metastasis of pancreatic cancer and an associated signaling pathway. *Tumour Biol*. 2012;33:1863-70.
- [159] Tavano F, di Mola FF, Piepoli A, Panza A, Copetti M, Burbaci FP, et al. Changes in miR-143 and miR-21 Expression and Clinicopathological Correlations in Pancreatic Cancers. *Pancreas*. 2012;41:1280-4.
- [160] Huang S, Guo W, Tang YB, Ren D, Zou XN, Peng XS. miR-143 and miR-145 inhibit stem cell characteristics of PC-3 prostate cancer cells. *Oncology Reports*. 2012;28:1831-7.
- [161] Iwaya T, Yokobori T, Nishida N, Kogo R, Sudo T, Tanaka F, et al. Downregulation of miR-144 is associated with colorectal cancer progression via activation of mTOR signaling pathway. *Carcinogenesis*. 2012;33:2391-7.
- [162] Sandhu SK, Volinia S, Costinean S, Galasso M, Neinast R, Santhanam R, et al. miR-155 targets histone deacetylase 4 (HDAC4) and impairs transcriptional activity of B-cell lymphoma 6 (BCL6) in the Emu-miR-155 transgenic mouse model. *Proc. Natl. Acad. Sci. USA*. 2012;109:20047-52.
- [163] Xiong B, Cheng Y, Ma L, Zhang C. MiR-21 regulates biological behavior through the PTEN/PI-3 K/Akt signaling pathway in human colorectal cancer cells. *Int. J. Oncol*. 2013;42:219-28.
- [164] Scheibner KA, Teaboldt B, Hauer MC, Chen X, Cherukuri S, Guo Y, et al. MiR-27a Functions as a Tumor Suppressor in Acute Leukemia by Regulating 14-3-3theta. *PLoS One*. 2012;7:e50895.
- [165] Smith AL, Iwanaga R, Drasin DJ, Micalizzi DS, Vartuli RL, Tan AC, et al. The miR-106b-25 cluster targets Smad7, activates TGF-beta signaling, and induces EMT and tumor initiating cell characteristics downstream of Six1 in human breast cancer. *Oncogene*. 2012;31:5162-71.
- [166] Li B, Shi XB, Nori D, Chao CK, Chen AM, Valicenti R, et al. Down-regulation of microRNA 106b is involved in p21-mediated cell cycle arrest in response to radiation in prostate cancer cells. *Prostate*. 2011;71:567-74.
- [167] McMillen BD, Aponte MM, Liu Z, Helenowski IB, Scholtens DM, Buttin BM, et al. Expression analysis of MIR182 and its associated target genes in advanced ovarian carcinoma. *Mod. Pathol*. 2012;25:1644-53.
- [168] Cittelty DM, Dimitrova I, Howe EN, Cochrane DR, Jean A, Spoelstra NS, et al. Restoration of miR-200c to Ovarian Cancer Reduces Tumor Burden and Increases Sensitivity to Paclitaxel. *Mol. Cancer Ther*. 2012;11:2556-65.

- [169] Howe EN, Cochrane DR, Cittelly DM, Richer JK. miR-200c targets a NF-kappaB up-regulated TrkB/NTF3 autocrine signaling loop to enhance anoikis sensitivity in triple negative breast cancer. *PLoS One*. 2012;7:e49987.
- [170] Majid S, Dar AA, Saini S, Arora S, Shahryari V, Zaman MS, et al. miR-23b Represses Proto-oncogene Src Kinase and Functions as Methylation-Silenced Tumor Suppressor with Diagnostic and Prognostic Significance in Prostate Cancer. *Cancer Res*. 2012;72:6435-46.
- [171] Hassan MQ, Maeda Y, Taipaleenmaki H, Zhang W, Jafferji M, Gordon JA, et al. miR-218 Directs a Wnt Signaling Circuit to Promote Differentiation of Osteoblasts and Osteomimicry of Metastatic Cancer Cells. *J. Biol. Chem*. 2012;287:42084-92.
- [172] Tie J, Pan Y, Zhao L, Wu K, Liu J, Sun S, et al. MiR-218 inhibits invasion and metastasis of gastric cancer by targeting the Robo1 receptor. *PLoS Genet*. 2010;6:e1000879.
- [173] Catuogno S, Cerchia L, Romano G, Pognonec P, Condorelli G, de Franciscis V. miR-34c may protect lung cancer cells from paclitaxel-induced apoptosis. *Oncogene*. 2012.
- [174] Cai N, Wang YD, Zheng PS. The microRNA-302-367 cluster suppresses the proliferation of cervical carcinoma cells through the novel target AKT1. *RNA*. 2013;19:85-95.
- [175] Zhang L, Huang J, Yang N, Greshock J, Megraw MS, Giannakakis A, et al. microRNAs exhibit high frequency genomic alterations in human cancer. *Proc. Natl. Acad. Sci. USA*. 2006;103:9136-41.
- [176] Tagawa H, Seto M. A microRNA cluster as a target of genomic amplification in malignant lymphoma. *Leukemia*. 2005;19:2013-6.
- [177] Wong KY, Huang X, Chim CS. DNA methylation of microRNA genes in multiple myeloma. *Carcinogenesis*. 2012;33:1629-38.
- [178] Baer C, Claus R, Frenzel LP, Zucknick M, Park YJ, Gu L, et al. Extensive promoter DNA hypermethylation and hypomethylation is associated with aberrant microRNA expression in chronic lymphocytic leukemia. *Cancer Res*. 2012;72:3775-85.
- [179] Wang Y, Toh HC, Chow P, Chung AY, Meyers DJ, Cole PA, et al. MicroRNA-224 is up-regulated in hepatocellular carcinoma through epigenetic mechanisms. *FASEB J*. 2012;26:3032-41.
- [180] Saito Y, Liang G, Egger G, Friedman JM, Chuang JC, Coetzee GA, et al. Specific activation of microRNA-127 with downregulation of the proto-oncogene BCL6 by chromatin-modifying drugs in human cancer cells. *Cancer Cell*. 2006;9:435-43.
- [181] Passon N, Gerometta A, Puppin C, Lavarone E, Puglisi F, Tell G, et al. Expression of DICER and Drosha in triple-negative breast cancer. *J. Clin. Pathol*. 2012;65:320-6.
- [182] Vaksman O, Hetland TE, Trope CG, Reich R, Davidson B. Argonaute, DICER, and Drosha are up-regulated along tumor progression in serous ovarian carcinoma. *Hum. Pathol*. 2012;43:2062-9.
- [183] Merritt WM, Lin YG, Han LY, Kamat AA, Spannuth WA, Schmandt R, et al. DICER, Drosha, and outcomes in patients with ovarian cancer. *N. Engl. J. Med*. 2008;359:2641-50.
- [184] Guo X, Liao Q, Chen P, Li X, Xiong W, Ma J, et al. The microRNA-processing enzymes: Drosha and DICER can predict prognosis of nasopharyngeal carcinoma. *J Cancer Res Clin Oncol*. 2012;138:49-56.
- [185] Kawai S, Amano A. BRCA1 regulates microRNA biogenesis via the DROSHA microprocessor complex. *J. Cell Biol*. 2012;197:201-8.

- [186] Gregory RI, Yan KP, Amuthan G, Chendrimada T, Doratotaj B, Cooch N, et al. The Microprocessor complex mediates the genesis of microRNAs. *Nature*. 2004;432:235-40.
- [187] Shiohama A, Sasaki T, Noda S, Minoshima S, Shimizu N. Molecular cloning and expression analysis of a novel gene DGCR8 located in the DiGeorge syndrome chromosomal region. *Biochem. Biophys. Res. Commun.* 2003;304:184-90.
- [188] Han J, Pedersen JS, Kwon SC, Belair CD, Kim YK, Yeom KH, et al. Posttranscriptional crossregulation between Drosha and DGCR8. *Cell*. 2009;136:75-84.
- [189] Triboulet R, Chang HM, Lapierre RJ, Gregory RI. Post-transcriptional control of DGCR8 expression by the Microprocessor. *RNA*. 2009;15:1005-11.
- [190] Zeng Y, Cullen BR. Structural requirements for pre-microRNA binding and nuclear export by Exportin 5. *Nucleic Acids Research*. 2004;32:4776-85.
- [191] Yi R, Qin Y, Macara IG, Cullen BR. Exportin-5 mediates the nuclear export of pre-microRNAs and short hairpin RNAs. *Genes Dev*. 2003;17:3011-6.
- [192] Lund E, Guttinger S, Calado A, Dahlberg JE, Kutay U. Nuclear export of microRNA precursors. *Science*. 2004;303:95-8.
- [193] Wang X, Xu X, Ma Z, Huo Y, Xiao Z, Li Y, et al. Dynamic mechanisms for pre-miRNA binding and export by Exportin-5. *RNA*. 2011;17:1511-28.
- [194] Melo SA, Moutinho C, Roperio S, Calin GA, Rossi S, Spizzo R, et al. A genetic defect in exportin-5 traps precursor microRNAs in the nucleus of cancer cells. *Cancer Cell*. 2010;18:303-15.
- [195] Ma E, Zhou K, Kidwell MA, Doudna JA. Coordinated activities of human DICER domains in regulatory RNA processing. *J. Mol. Biol.* 2012;422:466-76.
- [196] Nicholson AW. Dissecting human DICER: some assembly required. *J. Mol. Biol.* 2012;422:464-5.
- [197] Takeshita D, Zenno S, Lee WC, Nagata K, Saigo K, Tanokura M. Homodimeric structure and double-stranded RNA cleavage activity of the c-terminal RNase III domain of human DICER. *Journal of Molecular Biology*. 2007;374:106-20.
- [198] Zhang H, Kolb FA, Jaskiewicz L, Westhof E, Filipowicz W. Single processing center models for human DICER and bacterial RNase III. *Cell*. 2004;118:57-68.
- [199] Bernstein E, Kim SY, Carmell MA, Murchison EP, Alcorn H, Li MZ, et al. DICER is essential for mouse development. *Nat. Genet.* 2003;35:215-7.
- [200] Wienholds E, Koudijs MJ, van Eeden FJ, Cuppen E, Plasterk RH. The microRNA-producing enzyme DICER1 is essential for zebrafish development. *Nat. Genet.* 2003;35:217-8.
- [201] Korhonen HM, Meikar O, Yadav RP, Papaioannou MD, Romero Y, Da Ros M, et al. DICER is required for haploid male germ cell differentiation in mice. *PLoS One*. 2011;6:e24821.
- [202] Murchison EP, Stein P, Xuan Z, Pan H, Zhang MQ, Schultz RM, et al. Critical roles for DICER in the female germline. *Genes Dev*. 2007;21:682-93.
- [203] O'Rourke JR, Georges SA, Seay HR, Tapscott SJ, McManus MT, Goldhamer DJ, et al. Essential role for DICER during skeletal muscle development. *Developmental Biology*. 2007;311:359-68.
- [204] Harris KS, Zhang Z, McManus MT, Harfe BD, Sun X. DICER function is essential for lung epithelium morphogenesis. *Proc. Natl. Acad. Sci. USA*. 2006;103:2208-13.
- [205] Morita S, Hara A, Kojima I, Horii T, Kimura M, Kitamura T, et al. DICER is required for maintaining adult pancreas. *PLoS One*. 2009;4:e4212.

- [206] Nagaraja AK, Andreu-Vieyra C, Franco HL, Ma L, Chen R, Han DY, et al. Deletion of DICER in somatic cells of the female reproductive tract causes sterility. *Mol. Endocrinol.* 2008;22:2336-52.
- [207] Slade I, Bacchelli C, Davies H, Murray A, Abbaszadeh F, Hanks S, et al. DICER1 syndrome: clarifying the diagnosis, clinical features and management implications of a pleiotropic tumour predisposition syndrome. *Journal of Medical Genetics.* 2011;48:273-8.
- [208] Doros L, Yang J, Dehner L, Rossi CT, Skiver K, Jarzembowski JA, et al. DICER1 mutations in embryonal rhabdomyosarcomas from children with and without familial PPB-tumor predisposition syndrome. *Pediatr Blood Cancer.* 2012;59:558-60.
- [209] Kumar MS, Pester RE, Chen CY, Lane K, Chin C, Lu J, et al. DICER1 functions as a haploinsufficient tumor suppressor. *Genes Dev.* 2009;23:2700-4.
- [210] Zigelboim I, Reinhart AJ, Gao F, Schmidt AP, Mutch DG, Thaker PH, et al. DICER1 expression and outcomes in endometrioid endometrial adenocarcinoma. *Cancer.* 2011;117:1446-53.
- [211] Zhu DX, Fan L, Lu RN, Fang C, Shen WY, Zou ZJ, et al. Downregulated DICER expression predicts poor prognosis in chronic lymphocytic leukemia. *Cancer Sci.* 2012;103:875-81.
- [212] Selever J, Gu G, Lewis MT, Beyer A, Herynk MH, Covington KR, et al. DICER-mediated upregulation of BCRP confers tamoxifen resistance in human breast cancer cells. *Clin. Cancer Res.* 2011;17:6510-21.
- [213] Dedes KJ, Natrajan R, Lambros MB, Geyer FC, Lopez-Garcia MA, Savage K, et al. Down-regulation of the miRNA master regulators Drosha and DICER is associated with specific subgroups of breast cancer. *Eur. J. Cancer.* 2011;47:138-50.
- [214] Jennifer R. Choo TON. Biomarkers for Basal-like Breast Cancer. *Cancers.* 2010;2:1040-65.
- [215] Cheang MC, Voduc D, Bajdik C, Leung S, McKinney S, Chia SK, et al. Basal-like breast cancer defined by five biomarkers has superior prognostic value than triple-negative phenotype. *Clin. Cancer Res.* 2008;14:1368-76.
- [216] Khoshnaw SM, Rakha EA, Abdel-Fatah TM, Nolan CC, Hodi Z, Macmillan DR, et al. Loss of DICER expression is associated with breast cancer progression and recurrence. *Breast Cancer Res. Treat.* 2012;135:403-13.
- [217] Yan M, Huang HY, Wang T, Wan Y, Cui SD, Liu ZZ, et al. Dysregulated expression of DICER and drosha in breast cancer. *Pathol. Oncol. Res.* 2012;18:343-8.
- [218] Chiosea S, Jelezcova E, Chandran U, Luo J, Mantha G, Sobol RW, et al. Overexpression of DICER in precursor lesions of lung adenocarcinoma. *Cancer Res.* 2007;67:2345-50.
- [219] Karube Y, Tanaka H, Osada H, Tomida S, Tatematsu Y, Yanagisawa K, et al. Reduced expression of DICER associated with poor prognosis in lung cancer patients. *Cancer Sci.* 2005;96:111-5.
- [220] Faber C, Horst D, Hlubek F, Kirchner T. Overexpression of DICER predicts poor survival in colorectal cancer. *Eur. J. Cancer.* 2011;47:1414-9.
- [221] Cheloufi S, Dos Santos CO, Chong MM, Hannon GJ. A DICER-independent miRNA biogenesis pathway that requires Ago catalysis. *Nature.* 2010;465:584-9.
- [222] Cifuentes D, Xue H, Taylor DW, Patnode H, Mishima Y, Cheloufi S, et al. A novel miRNA processing pathway independent of DICER requires Argonaute2 catalytic activity. *Science.* 2010;328:1694-8.

- [223] Kumar MS, Lu J, Mercer KL, Golub TR, Jacks T. Impaired microRNA processing enhances cellular transformation and tumorigenesis. *Nat. Genet.* 2007;39:673-7.
- [224] Chiosea S, Jelezcova E, Chandran U, Acquafondata M, McHale T, Sobol RW, et al. Up-regulation of DICER, a component of the MicroRNA machinery, in prostate adenocarcinoma. *Am. J. Pathol.* 2006;169:1812-20.
- [225] Garber K. A tale of two cells: discovering the origin of prostate cancer. *J. Natl. Cancer Inst.* 2010;102:1528-9, 35.
- [226] Ma Z, Swede H, Cassarino D, Fleming E, Fire A, Dadras SS. Up-regulated DICER expression in patients with cutaneous melanoma. *PLoS One.* 2011;6:e20494.
- [227] Tokumaru S, Suzuki M, Yamada H, Nagino M, Takahashi T. let-7 regulates DICER expression and constitutes a negative feedback loop. *Carcinogenesis.* 2008;29:2073-7.
- [228] Martello G, Rosato A, Ferrari F, Manfrin A, Cordenonsi M, Dupont S, et al. A MicroRNA targeting DICER for metastasis control. *Cell.* 2010;141:1195-207.
- [229] Bennasser Y, Chable-Bessia C, Triboulet R, Gibbings D, Gwizdek C, Dargemont C, et al. Competition for XPO5 binding between DICER mRNA, pre-miRNA and viral RNA regulates human DICER levels. *Nat. Struct. Mol. Biol.* 2011;18:323-7.
- [230] Wiesen JL, Tomasi TB. DICER is regulated by cellular stresses and interferons. *Mol. Immunol.* 2009;46:1222-8.
- [231] Blandino G, Valerio M, Cioce M, Mori F, Casadei L, Pulito C, et al. Metformin elicits anticancer effects through the sequential modulation of DICER and c-MYC. *Nat. Commun.* 2012;3:865.
- [232] Hu Q, Tanasa B, Trabucchi M, Li W, Zhang J, Ohgi KA, et al. DICER- and AGO3-dependent generation of retinoic acid-induced DR2 Alu RNAs regulates human stem cell proliferation. *Nat. Struct. Mol. Biol.* 2012;19:1168-75.
- [233] Hammond SM, Bernstein E, Beach D, Hannon GJ. An RNA-directed nuclease mediates post-transcriptional gene silencing in *Drosophila* cells. *Nature.* 2000;404:293-6.
- [234] Nykanen A, Haley B, Zamore PD. ATP requirements and small interfering RNA structure in the RNA interference pathway. *Cell.* 2001;107:309-21.
- [235] Martinez J, Patkaniowska A, Urlaub H, Luhrmann R, Tuschl T. Single-stranded antisense siRNAs guide target RNA cleavage in RNAi. *Cell.* 2002;110:563-74.
- [236] Schwarz DS, Hutvagner G, Haley B, Zamore PD. Evidence that siRNAs function as guides, not primers, in the *Drosophila* and human RNAi pathways. *Mol. Cell.* 2002;10:537-48.
- [237] Schwarz DS, Hutvagner G, Du T, Xu Z, Aronin N, Zamore PD. Asymmetry in the assembly of the RNAi enzyme complex. *Cell.* 2003;115:199-208.
- [238] Chu Y, Yue X, Younger ST, Janowski BA, Corey DR. Involvement of argonaute proteins in gene silencing and activation by RNAs complementary to a non-coding transcript at the progesterone receptor promoter. *Nucleic Acids Research.* 2010;38:7736-48.
- [239] Janas MM, Wang B, Harris AS, Aguiar M, Shaffer JM, Subrahmanyam YV, et al. Alternative RISC assembly: binding and repression of microRNA-mRNA duplexes by human Ago proteins. *RNA.* 2012;18:2041-55.
- [240] Hagiwara K, Kosaka N, Yoshioka Y, Takahashi RU, Takeshita F, Ochiya T. Stilbene derivatives promote Ago2-dependent tumour-suppressive microRNA activity. *Sci. Rep.* 2012;2:314.

Chapter 104

THE ROLE OF THE EPIDERMAL GROWTH FACTOR RECEPTOR AS A THERAPEUTIC TARGET IN GLIOBLASTOMA AND OTHER MALIGNANCIES

Andrej Pala^{*}, Christian Rainer Wirtz and Marc-Eric Halatsch[†]

Department of Neurosurgery, University of Ulm School of Medicine, Ulm, Germany

ABSTRACT

Glioblastoma multiforme (GB) is the most common primary brain tumor among adults. Rapid tumor progression and diffuse invasion of brain tissue results in a poor prognosis despite advances in the understanding of the tumor's molecular biology. Recently, targeted therapy has been introduced as a potential therapeutic option in several types of cancer. Dysregulation of the epidermal growth factor receptor (EGFR) has been identified in a number of different malignant tumor entities, such as GB. Despite promising reports from preclinical studies, clinical trials yielded no significant improvement of outcomes of patients with GB who were treated with anti-EGFR-targeted agents. Molecular mechanisms underlying resistance to this treatment approach have become a focus of scientific efforts in the last years. The variations and complexity of EGFR signaling pathways demand a composite therapeutic strategy to quench alternative signaling routes which otherwise might enable cancer cell survival and subsequently tumor progression.

INTRODUCTION

In the last years, molecularly targeted therapy has rapidly evolved. Within this concept, the epidermal growth factor receptor (EGFR) became a common therapeutic target in many cancers [1-3]. The EGFR is a 170 kDa single-chain transmembraneous glycoprotein transmitting signals for various pathways, thereby playing an important role in cellular proliferation [1,2].

^{*} Corresponding Author's Email: andrej.pala@gmail.com.

[†] Corresponding Author's Email: marc-eric.halatsch@uniklinik-ulm.de.

Glioblastoma (GB) is the most common primary brain tumor among adults. Despite intense research efforts, the prognosis of patients with GB remains poor with current standard therapy consisting of tumor resection, adjuvant irradiation and temozolomide [2-4]. The EGFR is the most frequently altered oncogene in GB [3]. While the clinical use of small-molecule EGFR tyrosine kinase (TK) inhibitors has fallen short of expectations in patients with GB [4], a better understanding of molecular signaling pathway interactions may allow for the development of new rational therapeutic strategies to overcome intrinsic and/or acquired resistance of GB cells towards EGFR inhibition.

THE EGFR AND ITS ROLE IN TUMORIGENESIS

The human epidermal growth factor receptor (HER) family consists of four structurally related receptors, *i.e.*, HER1/EGFR, HER2/neu, HER3 and HER4. They are composed of internal, transmembraneous and external domains. The EGFR is activated by various ligands, most importantly epidermal growth factor (EGF) and transforming growth factor (TGF)- α [1,2]. Dimerization is a crucial step in HER signaling pathway activation. Both homo- and heterodimerization are possible, allowing variable receptor combinations [5]. Formation of dimers results in the activation of EGFR's internal TK domain with subsequent phosphorylation of tyrosine residues which, in turn, promote the activation of cytoplasmic downstream signaling routes. The phosphorylated tyrosine residues bind to signaling molecules with SRC homology 2 (SH2) which promote further intracellular signaling. The signalling routes involve activation of RAS-RAF-mitogen-activated kinase (MAPK), phosphatidylinositol-3-kinase (PI3K), phospholipase C γ (PLC γ) or Janus-activated kinase (JAK)-2. These molecular signaling mechanisms regulate various cellular processes involving proliferation and differentiation (Figure 1) [1-5].

Growth factor receptor binding protein 2 is a docking protein that is able to bind to activated EGFR-TKs, thereby creating complexes with the guanine nucleotide factor son of sevenless (SOS) through a SH3 domain. Activated SOS mediates the exchange of guanosine diphosphate (GDP) for guanosine triphosphate (GTP) in RAS that induces the activation of the serine/threonine protein kinase RAF, which, in turn, phosphorylates another serine/threonineprotein kinase, MEK. MEK activates MAPK that targets nuclear transcription factors, such as the E-Twenty-Six (ETS) family factors FOS, JUN and SLUG (SNAI-2) and hereby regulates cellular proliferation or differentiation [2,6]. The other signalling pathway mediated by EGFR is activated through PI3K, which forms phosphatidylinositol-3,4,5-triphosphate (PIP3) that further (through activation of AKT also known as PKB (protein kinase B)) indirectly phosphorylates the mammalian target of rapamycin (mTOR), which is playing a crucial role in cellular proliferation [7]. In addition, AKT is involved in the regulation of apoptosis through inactivation of proapoptotic proteins, such as BCL-2-associated death protein (BAD) or caspase 9. The next well-known signaling cascade promoted through EGFR-TK activation is the PLC γ pathway. Phospholipase C γ produces diacylglycerol (DAG) and inositoltriphosphate (IP3) through hydrolysis of phosphatidylinositol-4,5-bisphosphate (PIP2). Inositoltriphosphate binds to the IP3 receptor that triggers the influx of calcium ions from the endoplasmic reticulum and thereby triggers the activity of various proteins within the cytoplasm (Figure 1) [6,8].

These diverse signaling routes, mediated through EGFR activation, possess various messengers, alteration of which may contribute to the dysregulation of precisely coordinated signaling networks. Moreover, the plasticity of tumor cells may result in preference and potentiation of parallel signaling pathways and hereby contribute to the development of resistance to targeted therapy. Thus, simultaneous targeting of more signaling routes could lead to better inhibition and control of tumor cell proliferation.

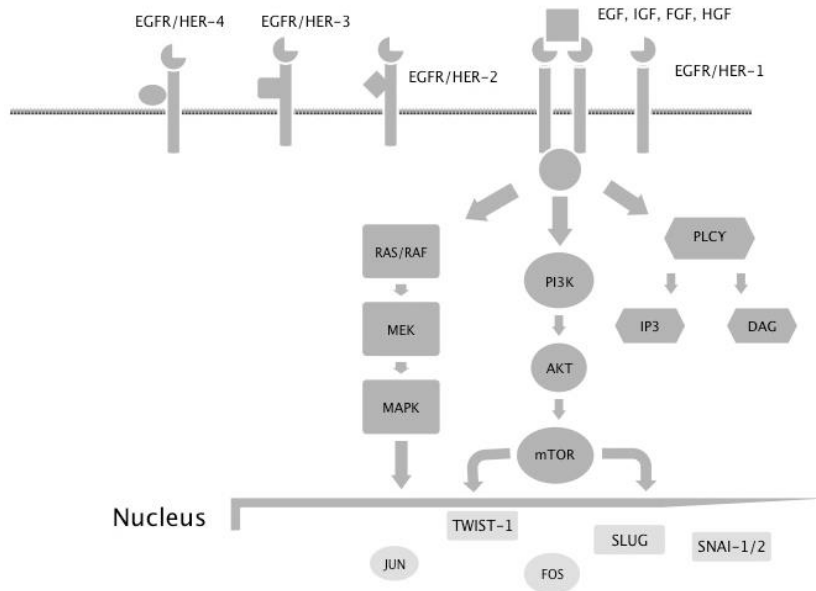


Figure 1. EGFR-mediated signaling cascades. DAG – diacylglycerol, EGF – epidermal growth factor, FGF – fibroblast growth factor, HGF – hepatocyte growth factor, IGF – insulin-like growth factor, IP3 – inositoltriphosphate, MAPK – mitogen-activated protein kinase, MEK – mitogen-activated protein kinase kinase, mTOR – mammalian target of rapamycin, PI3K – phosphatidylinositol-3-kinase.

Amplification and overexpression of the EGFR gene is frequently present in GB and other malignancies such as breast cancer or non-small cell lung cancer (NSCLC), contributing to unregulated cell proliferation [3,4,9]. Other mechanisms that lead to functional alteration of the EGFR are autocrine or paracrine overproduction of ligands and deletion-mutation of the receptor. The most common mutation, with an incidence of up to 58% in GBM, is EGFR variant (v)III which results from in-frame deletion of 801 base pairs in the DNA sequence encoding the extracellular receptor domain [3,4,9]. This mutation yields a constitutively activated receptor without any control mechanisms. In addition, mutations in the carboxyl terminal domain, referred to as EGFRvIV, have been described. Their activity exhibits significant tumorigenic potential [10].

The EGFR-TK inhibitors gefitinib and erlotinib are small molecules that compete with adenosine triphosphate (ATP) for intracellular TK domains. Hereby they prevent TK phosphorylation and inhibit downstream signaling that engages intracellular MAPK and PI3-K/AKT pathways. Gefitinib has been used for the treatment of NSCLC [11]. Even if the initial results were disappointing, further analysis enabled specification and selection of patients who harbour mutations of the intracellular EGFR domain that confer significant treatment

responses. Thus, gefitinib became an option in advanced and metastatic NSCLC. Analysis of structural receptor changes in both NSCLC and GB has revealed different types of alterations. Mutations found in NSCLC that confer response to erlotinib affect almost exclusively the intracellular EGFR domain whereas mutations in the external domain are mostly present in GB. Accumulating evidence suggests that the type of mutation that is characteristic for GB does not confer sensitivity to erlotinib, in contrast to EGFR mutations present in NSCLC [12,13]. Erlotinib binds to the active conformation of the TK domain that is less effective in GB. The absence of ATP-binding domain mutations in GB may be the explanation for failure of erlotinib against this disease. In contrast, lapatinib, a dual HER1/2 inhibitor, interacts with the inactive conformation of the TK domain, inducing cell death in EGFR-mutated GB cells. Almost complete inhibition of EGFR is required to achieve cell death induction [12,13]. The concentration of lapatinib that effectively leads to EGFR inhibition in GB cells *in vivo* remains unclear, and this may contribute to actually disappointing clinical results that show no clinical benefit of this approach [12-15]. It has been reported that low doses of dual EGFR and HER2 inhibitors result in better overall survival in xenograft models. In contrast, high-dose treatments may evoke rapid activation of alternative signalling routes and hereby enhance the development of resistance and tumor invasion [14,15].

The EGFRvIII and EGFRvIV contribute significantly to the activation of distinct signaling routes. The variability and complexity of this signaling enhance the adaptability of tumor cells [10]. The deletion-mutants EGFRvIII and EGFRvIV are stabilized by heat shock protein (HSP) 90 [16]. This is a common intracellular protein that guides folding and intracellular distribution of many other proteins, such as AKT or RAF-1. Heat shock protein 90 interacts with other co-chaperones and hereby forms and modulates protein complex conformation. Its action is attached to ATP hydrolysis, which ultimately results in conformational changes of other proteins. Heat shock protein 90 together with co-chaperone p50^{cdc37} build complexes with nascent EGFR and HER2/neu. Mature EGFR dissolves from Hsp90-p50^{cdc37} [16,17]. HER2/neu requires association with Hsp90-p50^{cdc37}. This molecular dependence might be another possible target that could bring about new therapeutic options [17].

Monoclonal antibodies such as cetuximab have been investigated as possible inhibitors of the EGFR signaling pathway [18,19]. Cetuximab binds to EGFR with high affinity and blocks further signal transduction within the EGFR-mediated signaling route. This effect contributes to the inhibition of cell cycle progression and angiogenesis. Moreover, cetuximab may induce apoptosis through overexpression of BAX and reduction of BCL-2 activity. It has been reported that in combination with other chemotherapeutic drugs, cetuximab increases the frequency of apoptosis *in vitro* and *in vivo*. However, the insufficient penetration of the blood-brain barrier may impede a relevant therapeutic effect [18,19].

The scientific efforts of the last few years were preferably concerned with the kinase-related activity of EGFR. However, TK-independent EGFR activation may play a significant role within cellular signaling pathways as well and may constitute an important reason for clinical failure of EGFR-TK inhibition in GB. Kinase-independent activity of EGFR is involved in cellular energy metabolism by maintaining intracellular glucose levels that prevent cells from undergoing cell death [20]. This observation is supported by the fact that the inhibition of EGFR phosphorylation results in diminished glucose uptake. However, the level of intracellular glucose does not reach the low concentration found in cells with entirely inactive EGFR. It has been reported that EGFR stabilizes the sodium-glucose linked transporter

(SGLT)-1 and hereby maintains the cellular uptake of glucose that is necessary for cell survival [20]. The SGLT-1 is a transmembraneous protein that enables glucose transport across the cell membrane. The inhibition of EGFR-TK activity provides deceleration of cell proliferation, but in the majority of cases does not result in cell death [20].

EPITHELIAL-TO-MESENCHYMAL TRANSITION

Epithelial-to-mesenchymal transition (EMT) may serve as another possible explanation for cell survival despite the inhibition of EGFR. Generally, EMT is a sequence of molecular signaling events that ultimately lead to increased invasiveness and dissemination of tumor cells [21,22]. These molecular changes play a crucial role in many physiological processes such as wound healing or organ development. Under pathological conditions, EMT is considered an important factor responsible for cancer progression. During EMT, cells lose their epithelial characteristics and acquire mesenchymal features such as the expression of fibronectin and vimentin [4,9,21,22]. These steps contribute to the acquisition of the biologically aggressive mesenchymal phenotype. Numerous signaling pathways have been implicated in EMT during cancer progression. For example, E-cadherin is a cell surface protein that mediates cell-to-cell contact. This main epithelial surface protein is involved in the regulation of critical processes in the development of epithelium [21,22]. Thus, minimal alterations in the regulatory network may significantly influence the overall physiology of the epithelial cell.

E-cadherin belongs to a large superfamily of cell-to-cell, calcium-dependent, homophilic adhesion molecules that are specified according to the tissue in which they are preferably formed. In general, cadherins build dimers that bind through their extracellular domains to cadherin dimers of neighbouring cells. The intracellular region is anchored to the plasma membrane and binds to the cytoskeleton through catenins. In particular, β -catenin and p120 catenin mediate the intracellular E-cadherin contact with the cytoskeleton, either directly or through other proteins, such as vinculin or β -actinin [9,21,22].

Epithelial cells are contact-inhibited through cell cycle arrest after reaching monolayer formation. Dysregulation of E-cadherin results in the dissolution of tight junctions, so that neighbouring cells lose their connections [9,21-23]. In addition, the loss of apico-basal polarity together with reorganisation of cytoskeletal components potentiates the cellular transformation from a cuboidal to a spindle-like shape. The anti-migratory and anti-proliferative effects of E-cadherin decrease. During organ development or wound healing, these processes are clearly coordinated and well-defined. Under pathological conditions, dysregulation of the signaling network occurs, facilitating the progression of cell transformation and behavioural modulation [21-23].

E-cadherin expression can be altered in different manners. Production of growth factors by the microenvironment may result in E-cadherin downregulation. Furthermore, overexpression of EGFR has a direct influence on E-cadherin function. Intracellular E-cadherin complexes can become direct targets of TKs. Phosphorylated complexes destabilize, and the strength of cell-to-cell adhesion weakens. Moreover, phosphorylation of E-cadherin complexes may be regulated through cytoplasmic kinases, such as SRC [21-23].

E-cadherin activates different signaling pathways, such as MAPK or PI3-K, after formation of cell-to-cell contact [9,21-23]. Additionally, E-cadherin can activate EGFR in the absence of

ligand and hereby trigger its signaling pathways [24]. This mechanism may significantly enhance cell proliferation or survival.

Alterations of several transcription factors and oncogenes contribute to the complex multilevel molecular transformation involved in EMT. The key transcriptional factors are SNAI-1/2, zinc finger E-box (ZEB)-1/2, TWIST and WNT/ β -catenin [21-24]. Their activity promotes loss of epithelial characteristics within cells and progressive acquisition of the mesenchymal phenotype. TWIST-1 is a helix-loop-helix protein that binds to DNA as a heterodimer, activates the mesenchymal marker N-cadherin and inhibits the expression of E-cadherin [4]. Specific inhibition of TWIST expression leads to reduction of glioma invasion *in vitro* [4,25]. Moreover, TWIST is involved in the regulation of p53-regulated apoptosis. The prevention of growth arrest further promotes the invasive phenotype of tumor cells, especially in GB [4,25]. This mechanism may lead to increased resistance to different therapeutic agents and ultimately to failure of targeted therapy.

SNAI-1 represents a transcriptional repressor that targets the expression of E-cadherin and belongs to a family composed of three members. SNAI-1 contains a SNAG domain which plays a crucial role for SNAI-1 activity. The repression of E-cadherin potentiates cellular invasion and migration [9,21-23]. Additionally, SNAI-1 induces the activation of matrix metalloprotease (MMP)-2. Matrix metalloproteases belong to a large family of zinc-dependent endoproteases that dissolve extracellular matrix proteins. This effect contributes to the increased invasiveness and aggressiveness of cells with the mesenchymal phenotype [9, 21-23].

Signal transducer and activator of transcription (STAT) signaling routes play an important role in EGFR-mediated signal transmission. Dysregulation of this pathway is associated with many human malignancies. Signal transducer and activator of transcription (STAT) proteins are located in the cytoplasm of all dividing cells. The STAT family contains seven protein members (STAT-1-4, STAT-5a, STAT-5b, STAT-6) which interact with different cytokines and various downstream signaling molecules as well as with EGFR autophosphorylation sites [26]. Cytokine receptors, such as interleukin (IL)-6, possess no intrinsic TK activity. They initiate Janus activated kinase (JAK) family members that mediate signaling to STAT molecules through phosphorylation of specific tyrosine residues. This process results in homo- or heterodimerisation of STAT molecules via SH2 domains. These complexes translocate into the nucleus, bind to specific DNA sequences and influence STAT genes that are associated with cell proliferation, differentiation, motility and apoptosis. Signal transducer and activator of transcription dimers can also be formed directly through phosphorylation by EGFR-TKs. Additionally, the activation via cytoplasmic SRC or Abelson murine leukemia viral oncogene homolog (ABL)-1 kinases is possible as well [26,27].

Signal transducer and activator of transcription-3 expression is typically associated with GB, and STAT-3 is activated by phosphorylation of tyrosine residue Y705 or by serine phosphorylation at the domain S727. The extent of STAT-3 activation correlates with glioma grade [27]. Activated STAT-3 translocates into the nucleus and regulates the expression of various genes, such as PIM-1, c-MYC, cyclin D2 or cyclin A, that are involved in cell cycle progression, oncogenesis, and apoptosis. Increased expression of TWIST results in the activation of EMT signaling. In addition, STAT-3 is able to interact with EGFRvIII and contributes to the malignant transformation of astrocytes in gliomas [27,28]. Signal transducer and activator of transcription-3 and EGFR-EGFRvIII promote together the activation of the pro-inflammatory gene cyclooxygenase (COX)-2 at the nuclear level [28]. Cyclooxygenase-2

is overexpressed in various malignancies such as NSCLC. Its activity may promote resistance to apoptosis and through the activation of angiogenesis confer invasiveness of tumour cells. Prostaglandin E-2 (PGE-2) as the main effector of COX-2 reduces the expression of E-cadherin via ZEB-1 and SNAI-1. Signal transducer and activator of transcription-3 contributes to the regulation of different signaling pathways and has an important influence on EMT and the cell cycle. It may represent an important strategic factor that could become an additional molecular target in the therapy of GB [27,28].

Signal transducer and activator of transcription-5 is activated by numerous cytokines, such as IL-2, IL-3, IL-5, erythropoietin, granulocyte macrophage colony stimulating factor (GM-CSF) or growth hormone. Upon activation and translocation into the nucleus, STAT-5 dimers regulate genes that are involved in apoptosis, e.g., the BCL genes. Constitutive STAT-5 activation has been implicated in hematopoietic malignancies and various solid tumors, such as breast or prostate cancer [29].

The general regulatory network may be modified through EMT epigenetically. For example, methylation of histones or DNA may significantly influence activation of genes associated with cell cycle regulation or apoptosis. CDH-1 represents the gene that encodes E-cadherin. It has been reported that SNAI-1, which is an important EMT-associated repressor of E-cadherin, activates DNA-methylating enzymes and simultaneously histone demethylases [30].

Transforming growth factor (TGF)- β is a protein that takes part in the regulation of cell proliferation and differentiation. Transforming growth factor- β contributes to EMT induction and activates many routes that result in the mesenchymal transformation of cells, e.g., SNAI, ZEB and TWIST which are commonly associated with tumor invasion. Additionally, TGF- β is associated with apoptosis induction via the SMAD pathway. Transforming growth factor- β enhances signaling in the MAPK and PI3K-AKT pathways and hereby mediates the expression of SNAI-1 and SNAI-2 [31]. Subsequently, the expression of fibronectin, vimentin and other proteins associated with the mesenchymal phenotype is increased. In contrast, the expression of E-cadherin decreases. SNAI proteins influence the activation of RHO and RAC. This action reduces adhesiveness and stimulates motility of cells. RHO is a small signaling G protein that belongs to the RAS superfamily of proteins that are involved in the regulation of actin [31].

EXTRACELLULAR MATRIX AND MICROENVIRONMENT

The extracellular matrix (ECM) is separated from epithelial cells by the basement membrane. As a result of malignant transformation and tumor dissemination, the integrity of the basement membrane destabilises. After elimination of this barrier, different signaling molecules produced by the ECM interact with epithelial cells and hereby activate cell responses at the nuclear level that are normally inhibited through the basement membrane. Besides the changes within cell-to-cell adhesion, alterations of cell-to-matrix adhesions determine the invasive features of cells [21-24,31]. Integrins build a large protein family involved in cell-to-matrix interactions [32,33]. They form heterodimer receptors for ECM proteins such as collagen, laminin or fibronectin [31,32]. Intracellular components of integrins cooperate with different TKs such as EGFR-TK. This signaling leads ultimately to actin polymerisation and phosphorylation of focal adhesion kinase (FAK), and binding residues for PI3-K, SRC or PLC γ

form [31,32,34]. Within this context, cell migration is induced by ECM via integrins. Dysregulated EGFR signaling subsequently dephosphorylates and inactivates FAK [34]. This process increases the migratory potential of solitary tumor cells. They detach from the ECM, disseminate and create new metastasis. On the other hand, integrin-mediated alignment may be an important step for the inverted process, called mesenchymal-to-epithelial transition (MET) [21-23]. Tumor cells acquire some adhesive properties again and form new solid masses. It has been reported that EGFRvIII interacts with integrins and hereby promotes metastatic progression [35].

MECHANISMS OF RESISTANCE TO CURRENT THERAPY OF GB

Within current GB studies, cancer stem-like cells (CSC) have received much attention. They exhibit a high tumorigenic potential in combination with relatively low proliferative activity. Cancer stem-like cells express the CD133 (prominin-1) gene which is a characteristic feature of all normal neural stem cells [36,37]. Additionally, CSCs demonstrate stem cell properties, such as colony formation or multilineage potency. Glioblastoma typically presents as a diffuse tumor that invades normal brain tissue and frequently recurs or progresses after radiation therapy. Cancer stem-like cells possess high resistance to radiation and chemotherapy [36,37]. A great extent of activation of DNA repair mechanisms represents a relevant factor that may contribute to this phenomenon. Moreover, the high self-renewal activity of cells expressing CD133 stimulates initiation, growth, invasion and recurrence of GB. Moreover, CSCs overexpress angiogenic vascular endothelial growth factor under hypoxic conditions [36-38]. A hypoxic microenvironment plays a crucial role in the recruitment and growth of stromal cells that consecutively support tumor progression.

Compensatory activation of HER2/neu and HER3 may result in increased resistance of CSCs against the inhibition of EGFR signaling [36,37]. Based on this observation, lapatinib as a dual EGFR/HER2 inhibitor yielded significantly better antiproliferative responses compared to cetuximab and other anti-EGFR-targeted agents. In the absence of an exogenous EGF stimulus, MAPK and AKT signaling remained constant through the activation of alternative signaling mediated by HER2/neu and HER3. In contrast, lapatinib decreases AKT and MAPK downstream signalling more effectively. This example provides a possible explanation for GB resistance to anti-EGFR-targeted monotherapy [36]. The multiple targeting of HER family members is an option which could potentially overcome resistance to anti-EGFR therapy and improve the clinical outcome of patients with GB.

Normal stem cells play a crucial role in differentiation, development and organogenesis. Transcription factors, such as NOTCH, ID1, SOX2 or OCT4 have been implicated in the regulation of balance between stem cell proliferation and differentiation. Dysregulation of this balance contributes to tumorigenesis. Myeloid Elf-1-like factor (MEF) is a member of the EST family of proteins. Myeloid Elf-1-like factor downregulates p53 through overexpression of MDM2 [36-38]. This process reduces activation of INK4a and promotes phosphorylation of the retinoblastoma protein. SOX2 is a high mobility group (HMG) box transcription factor that contributes to the self-renewal of stem cells. Moreover, its activity has been described in different malignancies and in GB. SOX2 is a direct target of MEF and through this action may potentiate stem-like features of tumor cells [37].

The WNT signaling route is related to EMT as well. WNT is associated with β -catenin. The transmembraneous Frizzled receptor is activated via WNT protein and targets Dishevelled, which inhibits glycogen synthase kinase (GSK)-3 β and thereby reduces the phosphorylation and degradation of β -catenin. Disinhibited β -catenin translocates into the nucleus and binds to T-cell factor (TCF) and to lymphoid enhancer factor (LEF)-1 [38]. This signaling activates various genes that are associated with EMT, such as SNAI-2 or JUN. In addition, the expression of proteins with mesenchymal characteristics increases, resulting subsequently in the development of a more aggressive cellular phenotype. The WNT/ β -catenin pathway induces cell migration which is important not only for embryonic development but for tumor progression as well. Additionally, this signaling route may lead to the acquisition of stem-like features in GB cells, which, in turn, increases the resistance to current standard therapy and may play a crucial role in tumor recurrence. Overexpression of the signaling regulator Frizzled 4 potentiates signalling within the WNT route and consecutively may influence the development and maintenance of CSCs [39].

CONCLUSION

The complexity and structural density of the molecular pathways involved in EMT result in fast and dynamic changes in the pathophysiology of GB. The inhibition of one downstream pathway may potentiate alternative signaling routes and challenge or induce adaptation mechanisms of tumor cells. The combination of various inhibitory mechanisms could lead to a better control of dysregulated signaling pathways in the future. Additionally, this approach might be effective in preventing the development of tumor cell resistance. The definition of mutational variations and their systematic conquering at different levels together with the identification of patients who would probably benefit from targeted therapy is urgently needed. The pharmacokinetics of anti-EGFR-targeted agents remains another challenge to be addressed. Future stepwise improvements in the areas outlined above hold the potential for progressively improving the prognosis of patients with GB.

REFERENCES

- [1] Normanno N., De Luca A., Bianco C., *et al.* Epidermal growth factor receptor signaling in cancer. *Gene* 2006, 366, 2-16.
- [2] Hynes N.E., Lane H.A. ERBB receptors and cancer: the complexity of targeted inhibitors. *Nat Rev Cancer* 2005, 5, 341-354.
- [3] Loew S., Schmidt U., Unterberg A., *et al.* The epidermal growth factor receptor as a therapeutic target in glioblastoma multiforme and other malignant neoplasms. *Anticancer Agents Med Chem* 2009, 9, 703-715.
- [4] Mikheeva S.A., Mikheev A.M., Petit A., *et al.* Twist 1 promotes invasion through mesenchymal change in human glioblastoma. *Mol Cancer* 2010, 9, 194-212.
- [5] Schlessinger J. Ligand-induced, receptor-mediated dimerization and activation of EGF receptor. *Cell* 2002, 110, 669-672.

- [6] Citri A., Yarden Y. EGF-ERBB signalling: towards the systems level. *Nat Rev Mol Cell Biol* 2006, 7, 505-516.
- [7] Sekulic A., Hudson C.C., Homme J.L., *et al.* A direct linkage between the phosphoinositide-3-kinase AKT signalling pathway and the mammalian target of rapamycin in mitogen-stimulated and transformed cells. *Cancer Res* 2000, 60(13), 3504-3513.
- [8] Berridge M.J. Inositol triphosphate and calcium signalling. *Nature* 1993, 361, 315-325.
- [9] Pala A., Karpel-Massler G., Kast R.E., *et al.* Epidermal to mesenchymal transition and failure of EGFR-targeted therapy in glioblastoma. *Cancers* 2012, 4, 523-530.
- [10] Pines G., Huang P.H., Zwang Y., *et al.* EGFRvIV: a previously uncharacterized oncogenic mutant reveals a kinase autoinhibitory mechanism. *Oncogene* 2010, 29, 5850-5860.
- [11] Suda K., Tomizawa K., Fujii M., *et al.* Epithelial to mesenchymal transition in an epidermal growth factor receptor-mutant lung cancer cell line with acquired resistance to erlotinib. *J Thorac Oncol* 2011, 6, 1152-1161.
- [12] Vivanco I., Robins H.I., Rohle D. Differential sensitivity of glioma- versus lung cancer-specific EGFR mutations to EGFR kinase inhibitors. *Cancer Discov* 2012, 2, 458-471.
- [13] Barkovich K.J., Hariono S., Garske A.L. Kinetics of inhibitor cycling underlies therapeutic disparities between EGFR-driven lung and brain cancers. *Cancer Discov* 2012, 2, 450-457.
- [14] Thiessen B., Stewart C., Tsao M., *et al.* A phase I/II trial of GW572016 (lapatinib) in recurrent glioblastoma multiforme: clinical outcomes, pharmacokinetics and molecular correlation. *Cancer Chemother Pharmacol* 2010, 65, 353-361.
- [15] Wang L., Liu Q., Zhao H., *et al.* Differential effects of low- and high-dose GW2974, a dual epidermal growth factor receptor and HER2 kinase inhibitor, on glioblastoma multiforme invasion. *J Neurosci Res* 2013, 91, 128-137.
- [16] Hernandez M.P., Sullivan W.P., Toft D.O. The assembly and intermolecular properties of Hsp70-Hsp90 molecular chaperone complex. *J Biol Chem* 2002, 277, 38294-38304.
- [17] Xu W., Mimnaugh E.G., Kim J.S., *et al.* Hsp90, not Grp94, regulates the intracellular trafficking and stability of nascent ERB2. *Cell Stress Chaperones* 2002, 7, 21-26.
- [18] Lu Y., Li X., Liang K., *et al.* Epidermal growth factor receptor (EGFR) ubiquitination as a mechanism of acquired resistance escaping treatment by the anti-EGFR monoclonal antibody cetuximab. *Mol Cancer Res* 2007, 67, 8240-8247.
- [19] Li S., Schmitz K.R., Jeffrey P.D., *et al.* Structural basis for inhibition of the epidermal growth factor receptor by cetuximab. *Cancer Cell* 2005, 7, 301-311.
- [20] Weihua Z., Tsan R., Huang W.-C., *et al.* Survival of cancer cells is maintained by EGFR independent of its kinase activity. *Cancer Cell* 2008, 13(5), 385-393.
- [21] Guarino M., Rubino B., Ballabio G. The role of epithelial-mesenchymal transition in cancer pathology. *Pathology* 2007, 39, 305-318.
- [22] Wever O.D., Pauwels P., Craene B.D., *et al.* Molecular and pathological signatures of epithelial-mesenchymal transitions at cancer invasion front. *Histochem Cell Biol* 2008, 130, 481-494.
- [23] Thiery J.P., Sleeman J.P. Complex networks orchestrate epithelial-mesenchymal transition. *Nat Rev Mol Cell Biol* 2006, 7, 131-142.

- [24] Qian X., Karpova T., Sheppard A.M., *et al.* E-cadherin mediated adhesion inhibits ligand-dependent activation of diverse receptor tyrosine kinases. *Embo J* 2004, 23, 1739-1748.
- [25] Elias M.C., Tozer K.R., Silber J.R., *et al.* TWIST is expressed in human gliomas and promotes invasion. *Neoplasia* 2005, 9, 824-837.
- [26] Yu H., Jove R. The STATs of cancer - new molecular targets come of age. *Nat Rev Cancer* 2004, 4(2), 97-105.
- [27] Lo H.W., Cao X., Zhu H., *et al.* Constitutively activated STAT-3 frequently coexpresses with epidermal growth factor receptor in high-grade gliomas and targeting STAT 3 sensitizes them to Iressa and alkylators. *Clin Cancer Res* 2008, 14, 6042-6054.
- [28] Lo H.W., Cao X., Zhu H., *et al.* COX-2 is a novel transcriptional target of the nuclear EGFR-STAT3 and EGFRvIII-STAT3 signaling axis. *Mol Cancer Res* 2010, 8, 232-245.
- [29] Moriggl R., Sexl V., Kenner L., *et al.* STAT5 tetramer formation is associated with leukemogenesis. *Cancer Cell* 2005, 7(1), 87-99.
- [30] Micalizzi D.S., Farabaugh S.M., Ford H.L. Epithelial-mesenchymal transition in cancer: parallels between normal development and tumor progression. *J Mammary Gland Biol Neoplasia* 2010, 15, 117-134.
- [31] Xu J., Lamouille S., Derynck R. TGF- β -induced epithelial to mesenchymal transition. *Cell Res* 2009, 19, 156-172.
- [32] Guo W., Giancotti F.G. Integrin signalling during tumor progression. *Nat Rev Mol Cell Biol* 2004, 5, 16-26.
- [33] Hynes R.O., Lively J.C., McCarthy J.H., *et al.* The diverse roles of integrins and their ligands in angiogenesis. *Cold Spring Harb Symp Quant Biol* 2002, 67, 143-153.
- [34] Sieg D.J., Hauck C.R., Illic D., *et al.* FAK integrates growth factor and integrin signals to promote cell migration. *Nat Cell Biol* 2000, 2, 249-256.
- [35] Ning Y., Zeineldin R., Liu Y., *et al.* Down-regulation of integrin $\alpha 2$ surface expression by mutant epidermal growth factor receptor (EGFRvIII) induces aberrant cell spreading and focal adhesion formation. *Cancer Res* 2005, 65, 9280-9286.
- [36] Clark P.A., Lida M., Treisman D.M., *et al.* Activation of multiple ERBB family receptors mediates glioblastoma cancer stem-like cell resistance to EGFR-targeted inhibition. *Neoplasia* 2012, 4(5), 420-428.
- [37] Bazzoli E., Pulvirenti T., Oberstadt M.C., *et al.* MEF promotes stemness in the pathogenesis of gliomas. *Cell Stem Cell* 2012, 11, 836-844.
- [38] Bao S., Wu Q., Sathornsumetee S., *et al.* Stem cell-like glioma cells promote tumor angiogenesis through vascular endothelial growth factor. *Cancer Res* 2006, 66, 7843-7848.
- [39] Jin X., Jeon H.-Y., Joo K.M., *et al.* Frizzled 4 regulates stemness and invasiveness of migrating glioma cells established by serial intracranial transplantation. *Cancer Res* 2011, 71(8), 3066-3075.

Chapter 105

TUMOR SUPPRESSORS INVOLVED IN DNA REPAIR AND CARCINOPREVENTION

Yasuko Kitagishi and Satoru Matsuda*

Department of Environmental Health Science,
Nara Women's University Kita-Uoya Nishimachi Nara, Japan

ABSTRACT

Cells possess a machinery to maintain the genomic integrity in response to the DNA damage. Mutations in several genes whose product improve the effects of DNA damage are known to predispose to develop a cancer. Under the genotoxic conditions, cells do not progress into S or M phase by activating DNA damage checkpoint, which acts as a process to transmit information from the damaged DNA lesions to cell cycle regulators. Tumor development may be accelerated by disruption of the balance between cell proliferation and the cell cycle regulation, which is maintained through regulations of various signal transduction pathways. It has been demonstrated that the signal transduction pathways are built on complicated networks between oncogenes and tumor suppressor genes such as p53 and its downstream factors. When the tumor suppressors lose their function, the cell may progress to cancer in combination with other genetic changes. Tumor suppressors regulate diverse cellular activities including DNA damage repair, cell proliferation, cell differentiation, cell migration, and programmed cell death. A better understanding of the cellular response to DNA damage will not only inform our knowledge of carcinogenesis but also provide better therapeutic opportunities.

Keywords: DNA repair, p53, BRCA1, PTEN, PI3K, AKT, ROS, cell signaling

* Corresponding Author's Email: smatsuda@cc.nara-wu.ac.jp (Tel: +81 742 20 3451; Fax: +81 742 20 3451).

ABBREVIATIONS

ATF2:	activating transcription factor 2
ATM:	ataxia telangiectasia-mutated
DSBs:	DNA double strand breaks
ERK:	Extracellular Signal-regulated Kinase
MAPK:	mitogen activated kinase
NF- κ B:	nuclear factor kappaB
mTOR:	mammalian target of rapamycin
PI3K:	phosphoinositide-3 kinase
PIP2:	phosphatidylinositol 4,5- bisphosphate
PIP3:	phosphatidylinositol 3,4,5-triphosphate
PTEN:	phosphatase and tensin homologue deleted on chromosome 10
PPAR:	peroxisome proliferator-activated receptor
PTP:	protein tyrosine phosphatase
ROS:	Reactive oxygen species
SAPK:	stress-activated protein kinase
TGF:	transforming growth factor

1. INTRODUCTION

Increased level of the oxidative stress results in cellular damage. To deal with DNA damages, cells are equipped with the multiple DNA repair mechanisms, which provoke a process to inhibit cell cycle progression and to induce DNA repair [1, 2]. One mechanism by which the oxidative stresses are thought to exert the effects may be through the reversible regulation of target molecules including several kinases, PI3K, and PTEN [3]. In addition, the main DNA damage recognition molecule is ataxia telangiectasia-mutated (ATM), which is a checkpoint kinase that phosphorylates a number of proteins including p53 and BRCA1 in response to the DNA damage (Figure 1). The p53 protein is a key transcription factor that regulates several signaling pathways involved in the cellular response to genome stress and DNA damage. Through the stress-induced activation, p53 triggers the expression of target genes that protect the genetic integrity of cells [4, 5]. A number of studies have demonstrated an antioxidant role for tumor suppressor proteins, activating the expression of some antioxidant genes in response to the oxidative stress. The tumor suppressors regulate diverse cellular activities including DNA damage repair, cell cycle arrest, cell proliferation, cell differentiation, migration, and apoptosis [6]. Normal cells show an exquisite balance among these various mechanisms of DNA repair. Mutations in the ATM have been associated with increased risk of developing a cancer. In addition, it is well known that mutations in the p53 and BRCA1 tumor suppressor genes account for a certain amount of cancers.

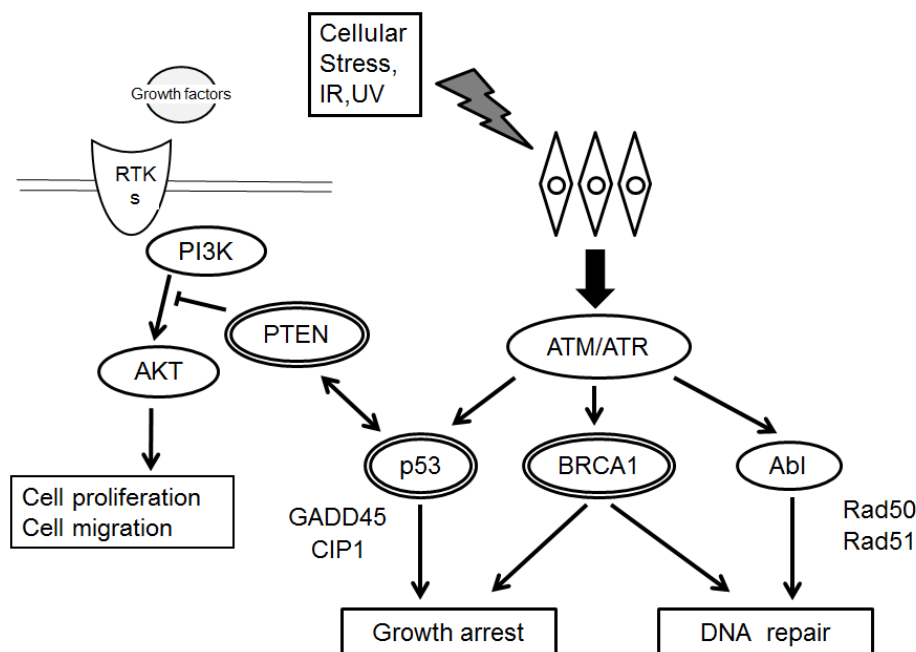


Figure 1. Schematic representation of the Cell proliferation, DNA repair and Growth arrest signaling pathways. Examples of the molecule known to act on the regulatory pathways are shown.

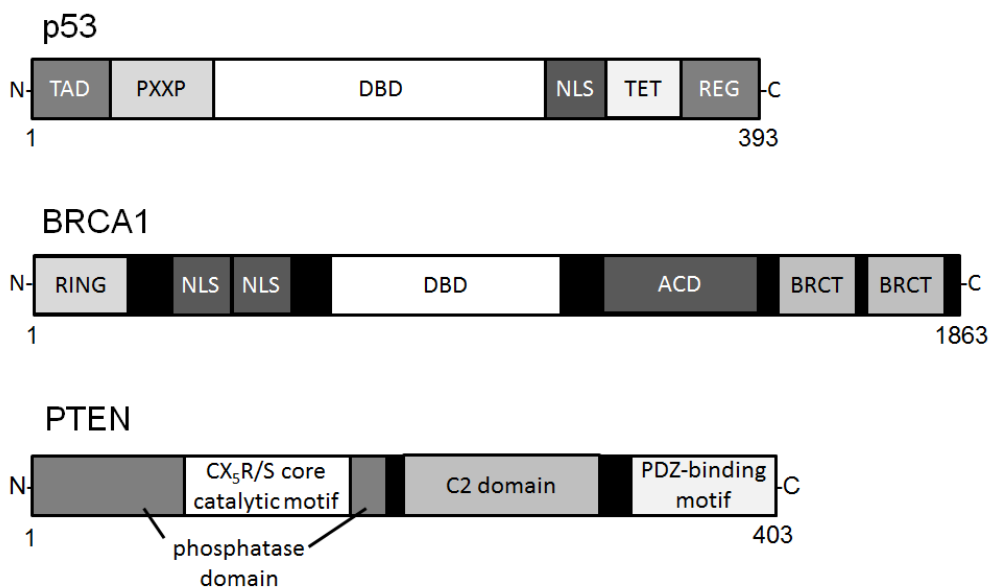


Figure 2. Schematic diagram indicating the domain structures of the p53, BRCA1, and PTEN proteins. The functionally important sites including the sites are also shown. Note that the sizes of protein are modified for clarity. TA= transactivation domain; PxxP= proline rich region; RING= (Really Interesting New Gene) finger domain, NLS= Nuclear Localization Signal, BRCT= BRCA1 C Terminus; C2 domain= a protein structural domain involved in targeting proteins to cell membranes; PDZ= a common structural domain in signaling proteins (PSD95, Dlg, ZO-1, etc.).

Genomic instability is often linked to DNA repair deficiencies. Standard DNA repair pathways available in mammalian cells include homologous repair, nonhomologous end joining, single strand annealing and so on. Those are different pathways that repair DNA double strand breaks (DSBs) [7]. The DNA repair is essential for the survival of both normal and cancer cells. An elaborate set of signaling pathways detect the DSBs and mediate either survival on the DNA repair or apoptotic cell death [8, 9]. The DNA damaging agents for cancer therapies are potent inducers of cell death triggered by the apoptosis. Recent advances in basic science have led to a better understanding of the molecular events important in the pathogenesis of cancer. In the present review, we summarize the function of prominent DNA repair molecules and the tumor suppressor gene products, p53, BRCA1 and PTEN (Figure 2), at a viewpoint of carcinogenic DNA damage and cellular response in cancer.

2. RELATIONSHIP BETWEEN DNA REPAIR AND CARCINOGENESIS

The DNA repair system is a highly conserved DNA editing process that maintains genomic fidelity through the recognition and repair of the damaged nucleotides. Genetic defects in DNA damage response genes and/or down-regulation of the DNA repair mechanism promote genomic instability, which can lead to carcinogenesis [10]. Cells are then equipped with multiple DNA repair mechanisms to the maintenance of genomic stability. The main DNA damage recognition molecule is ATM [11], which is a checkpoint kinase that phosphorylates a number of proteins in response to DNA damage including p53 and BRCA1 (**Figure 1**). Schematic structures of these important molecules are shown in Figure 2. An additional consequence of defective DNA repair is cellular hypersensitivity to DNA damaging agents [12]. Inhibition of DNA repair pathway seems to block the mechanisms that are required for survival in the presence of oncogenic mutations. Epigenetic mechanisms such as histone modifications and DNA methylation have been evaluated with a view for enhancing the cancer therapy via the regulation of the expression of genes involved in DNA repair [13]. Through the stresses-induced activation, p53 triggers the expression of the target genes that protect the genetic integrity of cells. The p53 gene is frequently mutated in multiple cancer tissues, suggesting that p53 plays a critical role in preventing cancers. Mutant p53 can be classified as a loss-of-function or gain-of-function protein depending on the mutation-type [14]. Wild-type p53 is inactive under normal physiological conditions and is activated in response to various types of DNA damaging stresses. The p53 activation may lead to regression of existing neoplastic lesions and therefore may be important in developing cancer prevention [15]. Failure of the DNA repair functions leads to p53-mediated induction of apoptotic cell death [16]. In this way, the p53 has been known to play a central role in maintaining a stable genome through its role in DNA repair, and apoptosis.

BRCA1 also fulfills the criteria for a tumor suppressor gene whose function is required to block cancer development. The mutation is associated with increased genomic instability in cells, which accelerates the mutation rate of other critical genes. Studies have established functional roles for BRCA1 in DNA damage signaling, DNA repair processes, and cell cycle checkpoints [17]. Consistent with these functional roles, cells deficient for *BRCA1* exhibit severe genomic instability and chromosomal aberrations. *BRCA1* cDNA encodes for 1863 amino acids protein with an amino terminal zinc ring finger motif and two putative nuclear

localization signals (Figure 2). The amino-terminal domain possesses E3 ubiquitin ligase activity [18] and the carboxyl-terminal domain is involved in binding to specific phosphoproteins [19]. BRCA1 becomes hyperphosphorylated after exposure to the DNA damaging agents, and the specific function of BRCA1 seems to be regulated by the phosphorylations [20]. Hence the DNA repair levels may be a novel therapeutic modality in cancer. Either survival or apoptosis, which is determined by the balance between DNA damage and DNA repair levels, may raise the major problems in cancer therapy.

3. SIGNAL TRANSDUCTION OF P53 AND BRCA1 IN DNA REPAIR PATHWAY

The p53 is inactive under normal physiological conditions and activated in response to various types of cellular stresses including DNA damage, which is also induced and activated in the nucleus by a stress such as hypoxia and oxidative stress. In addition, p53 undergoes post-translational modifications in response to the stresses [21]. The p53 protein is involved in a lot of signaling pathways of cell growth regulation, and multiple mechanisms have been revealed to accomplish the regulation of p53 activity, which determines the selectivity of p53 for specific transcriptional targets, resulting in control of the p53 activity. A number of molecules capable of activating p53 have been developed. Convincing evidence exists for the 53BP1 affecting the outcome of DNA double strand break repair [22, 23]. Among a number of transcriptional targets of the p53, the p21WAF1 has been shown to play an important role in both p53-dependent and independent pathways [24]. The p21 WAF1 inhibits cell cycle progression through interaction with the cyclin and CDK complexes. Studies have shown that p53 is mutated or deleted in nearly half of all human cancers, suggesting that p53 plays a critical role in preventing cancers. During neoplastic progression, the p53 is often mutated and fails to perform its normal functions. The p53 activation by something cellular regulator including a gain of function-mutation may lead to regression of an early neoplastic lesion, and therefore may be important in developing cancer-prevention.

Mutations in the tumor suppressor gene BRCA1 confer an increased risk for the development of breast and ovarian cancers [25]. In particular, BRCA1 hereditary breast cancer is a type of cancer with defects in a DNA repair pathway. Mutation of a single allele of the cancer susceptibility gene BRCA1 is associated with increased genomic instability in human breast epithelial cells [26], which accelerates the mutation rate of other critical genes. Several functions of BRCA1 may contribute to its tumor suppressor activity including roles in the DNA repair. Although BRCA1 gene mutations are rare in sporadic breast and/or ovarian cancers, BRCA1 protein expression is frequently reduced in the sporadic cases. The BRCA1 has the important role in concert with BRCA2, Rad50 and Rad51 [27], in order to activate the checkpoints. For example, BRCA1 is colocalized with Rad51, a DNA recombinase related to the bacterial RecA protein. The BRCA1 protein becomes hyper-phosphorylated after exposure to the DNA damaging agents, and the function of BRCA1 seems to be regulated by the phosphorylation in response to DNA damage. Pharmacological inhibition of poly-ADP-ribose polymerase induces cell death in tumors with mutations in certain DNA repair pathways, when combined with DNA damaging chemotherapies. Then, poly-ADP-ribose polymerase inhibitors have been investigated for the treatment of patients with BRCA 1 mutation, as a strategy to

potentiate the DNA damaging effects of chemotherapy and irradiation [28, 29]. The role of BRCA1 in cell cycle control has been understood by its ability to interact with various cyclins and cyclin-dependent kinases. The BRCA1 activates the CDK inhibitor p21 and the p53 tumor suppressor protein, which regulates several genes that control cell cycle checkpoints.

4. PROTEIN INTERACTION AND FUNCTIONAL INTERPLAY BETWEEN PTEN AND p53

PTEN gene is the frequently mutated and deleted tumor suppressor in human cancer. Human genomic *PTEN* locus consists of nine exons on chromosome 10q23.3, encoding a 5.5 kb mRNA that specifies a 403 amino-acid open reading frame [30, 31] (Figure 2), which is ubiquitously expressed throughout early embryogenesis [32]. The translation product is a 53 kDa protein with homology to tensin and protein tyrosine phosphatases. The PTEN inactivation is implicated in the carcinogenesis of several cancers [33], which causes an increase in cellular PIP3 levels, by which activated PI3K/AKT signaling causes increased expression of several genes for cell survival. PTEN and p53 is known to interact and regulate each other (Figure 1) at the transcription as well as protein level, which could be at the important control machinery for switching between survival and death. This cross talk is frequently a combination of reciprocally antagonistic pathways, which often involves another tumor suppressor gene MDM2. The cross talk may serve as an added regulatory effect on the expression of key genes involved in cancer. It has also been revealed that PTEN regulates p53 stability and in turn regulates its own transcriptional activity. The PTEN and p53 complex enhances p53 DNA binding and transcriptional activity [34]. An important p53 function is to act as a transcription factor by binding to the specific DNA consensus sequence in responsive genes, which may increase the synthesis of p21^{waf1} that is an important protein involved in cell cycle arrest [35]. In addition, one of transcriptional targets of p53 is also the PTEN. One way by which p53 inhibits production of PIP3 indirectly is by inducing the expression of PTEN [36]. The p53 and AKT influence the process of apoptosis in opposite ways. The AKT promotes cell survival by suppressing pro-apoptotic proteins such as Bad through phosphorylation [37]. There are also cross talks between p53 and AKT involving gene transcription as well as posttranslational protein modifications. One way by which p53 inhibits PIP3 production indirectly is by repressing the catalytic subunit of PI3K. A subsequent p53-induced expression of PTEN causes the p53-PTEN interaction, which then suppresses the cell survival machinery of AKT pathway. PTEN is required for the maintenance of p53 acetylation, which is also required for target gene transcription [38].

Growth factor-activated AKT signaling promotes progression of cell cycles by acting on downstream factors involved in controlling the G1/S and/or G2/M transitions. Several studies have also implicated AKT in modulating DNA damage responses and genome stability [39]. AKT therefore modifies downstream signaling in complex ways. In addition, PTEN also plays a critical role in DNA damage repair and DNA damage response through its interaction with p53 pathways in an AKT-independent manner [40]. Furthermore, nuclear PTEN is sufficient to reduce tumor progression in a p53 dependent manner. It has also been suggested that nuclear PTEN play a unique role to protect cells upon oxidative damage and to regulate carcinogenesis [41]. One aspect of the PTEN tumor suppressor signaling is achieved through stabilization of

the p53 protein. PTEN has been shown to physically interact with p53 and prevent its degradation. The instability of PTEN correlated with its missense mutations has been shown to involve protein interactions. PTEN may be regulated by ubiquitin-mediated proteasomal degradation, a common mechanism to control protein levels.

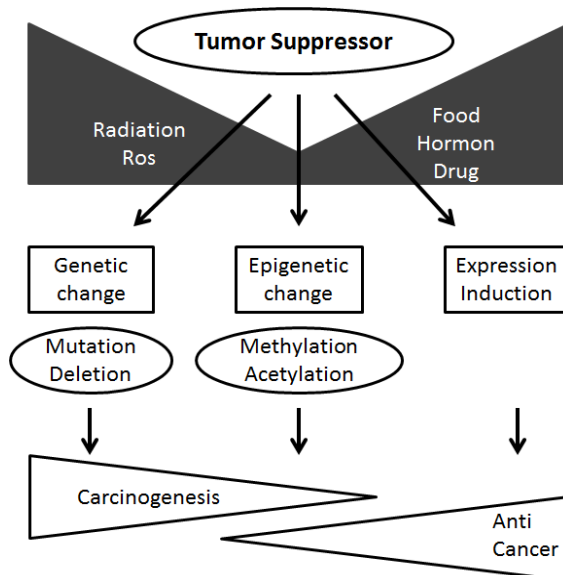


Figure 3. Implication of tumor suppressor gene modulations in cancer. Expression of tumor suppressor genes is regulated by genetic, epigenetic, and transcriptional changes, which may result in the DNA repair activity in a cell. The DNA repair downregulation can contribute to genomic instability, which promotes malignant transformation of cells.

5. PERSPECTIVE

Tumor suppressor molecules protect a cell from cancers. When the tumor suppressor genes lose their function by genetic changes such as mutations or deletions, the cell may progress to cancer in combination with other genetic changes (Figure 3). The phenotype of the cancerous cell may also arise from epigenetic events that may alter the gene expression. Epigenetic silencing of the tumor suppressor genes is a well-established oncogenic process [42]. So, the epigenetic modifications are of particular importance for the imprinted genes which are generally located in clusters [43, 44]. They are differentially marked by DNA methylation, histone acetylation, deacetylation, and histone methylation [45]. However, the molecular mechanisms by which they regulate developmental transitions have not yet been well defined. Epigenetic differences accumulate with age and different environments create different patterns of cellular modifications. Transient nutritional, biophysical, or biochemical stimuli occurring at specific stages may have influences on gene expression by interacting with epigenetic mechanisms and altering chromatin compaction and transcription factor accessibility [46] (Figure 3). Actually, the mRNA and protein expression level of the tumor suppressor genes is increased following treatment with drug, hormone, or food. For example, rosemary extract represses *PTEN* expression in K562 leukemic culture cells [47]. It has been paid more attention

to the DNA repair, because specific cancer prevention may be possible via the mechanisms. The cancer cell genome is aberrant as a consequence of incomplete DNA repair. As many anticancer drugs further reduce the integrity of DNA, they may be able to cause more mutations and another cancer, if the lesions are not repaired. However, cancer cells, in which its DNA repair is down-regulated, have been shown to exhibit increased sensitivity to DNA damaging chemotherapy. Understanding of the cellular aberrations of cancer cells has allowed the development of therapies to target biological pathways. Several studies have evaluated the role of DNA repair enzyme inhibitors for treatment of cancer [48, 49]. Further investigations will be required to identify other additional mechanisms associated with the therapeutic sensitivity. Also, future studies should be conducted to determine whether the combination of DNA damaging agents and DNA repair modulator has potential for the treatment against cancer.

ACKNOWLEDGMENTS

This work was supported by grants-in-aid from the Ministry of Education, Culture, Sports, Science and Technology in Japan. In addition, this work was supported in part by the grant from Nakagawa Masashichi Shoten Co., Ltd.

Competing Interests Statement

The authors declare that they have no competing financial interests.

REFERENCES

- [1] Frame FM, Maitland NJ (2011) Cancer stem cells, models of study and implications of therapy resistance mechanisms. *Adv Exp Med Biol.* 720: 105-118.
- [2] Athar M, Elmets CA, Kopelovich L (2011) Pharmacological activation of p53 in cancer cells. *Curr Pharm Des.* 17: 631-639.
- [3] Li ZY, Yang Y, Ming M, and Liu B, "Mitochondrial ROS generation for regulation of autophagic pathways in cancer," *Biochem Biophys Res Commun*, vol. 414, pp. 5-8, 2011.
- [4] Bykov VJ, Lambert JM, Hainaut P, Wiman KG (2009) Mutant p53 rescue and modulation of p53 redox state. *Cell Cycle.* 8: 2509-2517.
- [5] Liu B, Chen Y, St Clair DK (2008) ROS and p53: a versatile partnership. *Free Radic Biol Med.* 44: 1529-1535.
- [6] Asai T, Liu Y, Bae N, and Nimer SD, "The p53 tumor suppressor protein regulates hematopoietic stem cell fate," *J Cell Physiol*, vol. 226, pp. 2215-2221, 2011.
- [7] Gostissa M, Alt FW, Chiarle R (2011) Mechanisms that promote and suppress chromosomal translocations in lymphocytes. *Annu Rev Immunol.* 29: 319-350.
- [8] Al-Ejeh F, Kumar R, Wiegman A, Lakhani SR, Brown MP, Khanna KK (2010) Harnessing the complexity of DNA-damage response pathways to improve cancer treatment outcomes. *Oncogene.* 29: 6085-6098.

- [9] Krystof V, Uldrijan S (2010) Cyclin-dependent kinase inhibitors as anticancer drugs. *Curr Drug Targets*. 11: 291-302.
- [10] Kelly GL and Strasser A, "The essential role of evasion from cell death in cancer," *Adv Cancer Res*, vol. 111, pp. 39-96, 2011.
- [11] Bhatti S, Kozlov S, Farooqi AA, Naqi A, Lavin M, and Khanna KK, "ATM protein kinase: the linchpin of cellular defenses to stress," *Cell Mol Life Sci*, vol. 68, pp. 2977-3006, 2011.
- [12] Okumura N, Yoshida H, Kitagishi Y, Nishimura Y, Iseki S, and Matsuda S, "Against Lung Cancer Cells: To Be, or Not to Be, That Is the Problem," *Lung Cancer Int* doi:10.1155/2012/659365, in press, 2012.
- [13] Gigeck CO, Chen ES, Calcagno DQ, Wisnieski F, Burbano RR, and Smith MA, "Epigenetic mechanisms in gastric cancer," *Epigenomics*, vol. 4, pp. 279-294, 2012.
- [14] Martinez-Rivera M and Siddik ZH, "Resistance and gain-of-resistance phenotypes in cancers harboring wild-type p53," *Biochem Pharmacol*, vol. 83, pp. 1049-1062, 2012.
- [15] Stegh AH, "Targeting the p53 signaling pathway in cancer therapy - the promises, challenges and perils," *Expert Opin Ther Targets*, vol. 16, pp. 67-83, 2012.
- [16] Zawacka-Pankau J, Krachulec J, Grulkowski I, Bielawski KP, and Selivanova G, "The p53-mediated cytotoxicity of photodynamic therapy of cancer: recent advances," *Toxicol Appl Pharmacol*, vol. 232, no. 3, pp. 487-497, 2008.
- [17] Roy R, Chun J, and Powell SN, "BRCA1 and BRCA2: different roles in a common pathway of genome protection," *Nat Rev Cancer*, vol. 12, pp. 68-78, 2011.
- [18] Ohta T, Sato K, and Wu W, "The BRCA1 ubiquitin ligase and homologous recombination repair," *FEBS Lett*, vol. 585, pp. 2836-2844, 2011.
- [19] Leung CC and Glover JN, "BRCT domains: easy as one, two, three," *Cell Cycle*, vol. 10, pp. 2461-2470, 2011.
- [20] Ouchi T, "BRCA1 phosphorylation: biological consequences," *Cancer Biol Ther*, vol. 5, pp. 470-475, 2006.
- [21] Kim DH, Kundu JK, Surh YJ (2011) Redox modulation of p53: mechanisms and functional significance. *Mol Carcinog*. 50: 222-234.
- [22] Aly A, Ganesan S (2011) BRCA1, PARP, and 53BP1: conditional synthetic lethality and synthetic viability. *J Mol Cell Biol*. 3: 66-74.
- [23] Kobayashi J, Iwabuchi K, Miyagawa K, Sonoda E, Suzuki K, Takata M, Tauchi H (2008) Current topics in DNA double-strand break repair. *J Radiat Res*. 49: 93-103.
- [24] Lo PK, Lee JS, Sukumar S (2012) The p53-p21WAF1 checkpoint pathway plays a protective role in preventing DNA rereplication induced by abrogation of FOXF1 function. *Cell Signal*. 24: 316-324.
- [25] Chionh F, Mitchell G, Lindeman GJ, Friedlander M, Scott CL (2011) The role of poly adenosine diphosphate ribose polymerase inhibitors in breast and ovarian cancer: current status and future directions. *Asia Pac J Clin Oncol*. 7: 197-211.
- [26] Konishi H, Mohseni M, Tamaki A, Garay JP, Croessmann S, Karnan S, Ota A, Wong HY, Konishi Y, Karakas B, Tahir K, Abukhdeir AM, Gustin JP, Cidado J, Wang GM, Cosgrove D, Cochran R, Jelovac D, Higgins MJ, Arena S, Hawkins L, Lauring J, Gross AL, Heaphy CM, Hosokawa Y, Gabrielson E, Meeker AK, Visvanathan K, Argani P, Bachman KE, Park BH (2011) Mutation of a single allele of the cancer susceptibility gene BRCA1 leads to genomic instability in human breast epithelial cells. *Proc Natl Acad Sci U S A*. 108: 17773-17778.

- [27] Rebbeck TR, Mitra N, Domchek SM, Wan F, Chuai S, Friebel TM, Panossian S, Spurdle A, Chenevix-Trench G; kConFab, Singer CF, Pfeiler G, Neuhausen SL, Lynch HT, Garber JE, Weitzel JN, Isaacs C, Couch F, Narod SA, Rubinstein WS, Tomlinson GE, Ganz PA, Olopade OI, Tung N, Blum JL, Greenberg R, Nathanson KL, Daly MB (2009) Modification of ovarian cancer risk by BRCA1/2-interacting genes in a multicenter cohort of BRCA1/2 mutation carriers. *Cancer Res.* 69: 5801-5810.
- [28] Orlando L, Schiavone P, Fedele P, Calvani N, Nacci A, Cinefra M, D'Amico M, Mazzoni E, Marino A, Sponziello F, Morelli F, Lombardi L, Silvestris N, Cinieri S (2012) Poly (ADP-ribose) polymerase (PARP): rationale, preclinical and clinical evidences of its inhibition as breast cancer treatment. *Expert Opin Ther Targets.* 16:S83-S89.
- [29] Banerjee S, Kaye S (2011) PARP inhibitors in BRCA gene-mutated ovarian cancer and beyond. *Curr Oncol Rep.* 13: 442-449.
- [30] Okumura N, Yoshida H, Kitagishi Y, Murakami M, Nishimura Y, and Matsuda S, "PI3K/AKT/PTEN Signaling as a Molecular Target in Leukemia Angiogenesis," *Adv Hematol*, vol. 2012, pp. 843085, 2012.
- [31] Okumura N, Yoshida H, Kitagishi Y, Nishimura Y, and Matsuda S, "Alternative splicings on p53, BRCA1 and PTEN genes involved in breast cancer," *Biochem Biophys Res Commun*, vol. 413, pp. 395-399, 2011.
- [32] Croushore JA, Blasiole B, Riddle RC et al., "Ptena and ptenb genes play distinct roles in zebrafish embryogenesis," *Dev Dyn*, vol. 234, pp. 911-921, 2005.
- [33] Barbieri SS, Ruggiero L, Tremoli E, and Weksler BB, "Suppressing PTEN activity by tobacco smoke plus interleukin-1beta modulates dissociation of VE-cadherin/beta-catenin complexes in endothelium," *Arterioscler Thromb Vasc Biol*, vol. 28, pp. 732-738, 2008.
- [34] Freeman DJ, Li AG, Wei G et al., "PTEN tumor suppressor regulates p53 protein levels and activity through phosphatase-dependent and -independent mechanisms," *Cancer Cell*, vol. 3, no. 2, pp. 117-130, 2003.
- [35] Lo PK, Lee JS, and Sukumar S, "The p53-p21WAF1 checkpoint pathway plays a protective role in preventing DNA rereplication induced by abrogation of FOXF1 function," *Cell Signal*, vol. 24, no. 1, pp. 316-324, 2012.
- [36] Puszyński K, Hat B, and Lipniacki T, "Oscillations and bistability in the stochastic model of p53 regulation," *J Theor Biol*, vol. 254, no. 2, pp. 452-465, 2008.
- [37] Sen P, Mukherjee S, Ray D, and Raha S, "Involvement of the Akt/PKB signaling pathway with disease processes," *Mol Cell Biochem*, vol. 253, no. 1-2, pp. 241-246, 2003.
- [38] Li AG, Piluso LG, Cai X, Wei G, Sellers WR, and Liu X, "Mechanistic insights into maintenance of high p53 acetylation by PTEN," *Mol Cell*, vol. 23, no. 4, pp. 575-587, 2006.
- [39] Quevedo C, Kaplan DR, and Derry WB, "AKT-1 regulates DNA-damage-induced germline apoptosis in *C. elegans*," *Curr Biol*, vol. 17, no. 3, pp. 286-292, 2007.
- [40] Ming M, and He YY, "PTEN in DNA damage repair," *Cancer Lett*, vol. 319, no. 2, pp. 125-129, 2012.
- [41] Bonavida B, and Baritaki S, "The novel role of Yin Yang 1 in the regulation of epithelial to mesenchymal transition in cancer via the dysregulated NF- κ B/Snail/YY1/RKIP/PTEN Circuitry," *Crit Rev Oncog*, vol. 16, no. 3-4, pp. 211-226, 2011.
- [42] Ren J, Singh BN, Huang Q et al., "DNA hypermethylation as a chemotherapy target," *Cell Signal*, vol. 23, no. 7, pp. 1082-1093, 2011.

- [43] Ribarska T, Bastian KM, Koch A, and Schulz WA, "Specific changes in the expression of imprinted genes in prostate cancer--implications for cancer progression and epigenetic regulation," *Asian J Androl*, vol. 14, no. 3, pp. 436-450, 2012.
- [44] Abramowitz LK and Bartolomei MS, "Genomic imprinting: recognition and marking of imprinted loci," *Curr Opin Genet Dev*, vol. 22, no. 2, pp. 72-78, 2012.
- [45] Issa JP, "Epigenetic changes in the myelodysplastic syndrome," *Hematol Oncol Clin North Am*, vol. 24, no. 2, pp. 317-330, 2010.
- [46] Martino DJ and Prescott SL, "Silent mysteries: epigenetic paradigms could hold the key to conquering the epidemic of allergy and immune disease," *Allergy*, vol. 65, no. 1, pp. 7-15, 2010.
- [47] Yoshida H, Okumura N, Kitagishi Y, Nishimura Y, and Matsuda S, "Ethanol extract of Rosemary repressed PTEN expression in K562 culture cells," *Int J appl Boil pharm Technol*, vol. 2, pp. 316-322, 2011.

Chapter 106

MOLECULAR PATHOGENESIS OF NEUROFIBROMATOSIS TYPE 1

***Ming-Jen Lee¹, Alton Etheridge²,
David J. Galas^{2,3} and Kai Wang²***

¹Department of Neurology and Medical Genetic,
National Taiwan University School of Medicine, National
Taiwan University, Taipei, Taiwan, Republic of China

²Institute for Systems Biology, Seattle, WA, US

³Luxembourg Centre for Systems Biomedicine,
Esch-sur-Alzette, Luxembourg

ABSTRACT

Neurofibromatosis type 1 (NF1) is one of the most common inherited neurological disorders. Clinically, NF1 is characterized by café-au-lait spots, skinfold freckling, cutaneous and subcutaneous neurofibroma, plexiform neurofibroma, Lisch nodules, bony defects and frequently by a positive family history in first degree relatives. The causative gene encodes a large protein, neurofibromin, a negative regulator of the RAS signaling pathway. In addition to neurofibromatosis, several other disorders such as Noonan syndrome, Costello syndrome, Legius syndrome and cardio-facio-cutaneous (CFC) syndrome also result from a dysfunctional RAS pathway; therefore, all of these diseases have been grouped as RASopathies. While the molecular genetics of NF1 are well studied, the biological processes underlying the development of NF1-associated pathologies have yet to be fully understood. Using a systems biology approach to integrate various global, molecular profiling data can provide insights into the function and interactions of key regulatory molecules associated with NF1. These include transcription factors and microRNAs. This may ultimately lead to the discovery of new diagnostic markers and the identification of therapeutic targets.

INTRODUCTION

Neurofibromatosis (NF) is the most common hereditary neurological disorder. It was first described by German pathologist Friedrich Daniel von Recklinghausen [1]. Neurofibromatosis type 1 and 2 (NF1 and NF2) are the two major forms of autosomal dominant neuro-cutaneous syndromes. They show no special preference for race and gender. NF1 is caused by mutation(s) of the neurofibromatosis type 1 gene (*NF1*), which encodes a multi-domain protein, neurofibromin. Clinically, NF1 has extremely variable manifestations which include multiple café-au-lait spots (CALS), cutaneous neurofibromas, plexiform neurofibromas, iris hamartomas (Lisch nodules), bony defects and neoplasms of the nervous system. Although tumors in NF1 are usually benign in nature, transformation of plexiform neurofibroma to form malignant peripheral nerve sheath tumors, occurs in approximately 10% of NF1 patients.

EPIDEMIOLOGY OF NF1

In 1981, Samuelsson reported the first population-based prevalence study of NF1 with a prevalence of 1/4600 in the Gothenburg region, Sweden [2]. Huson *et al.* surveyed South East Wales, UK and reported a prevalence of 1/4150 [3]. The highest estimated prevalence of NF1, 1/2190, was reported in 1989 in Dunedin, New Zealand [4]. Several studies with populations from northeast Italy and northern Finland also gave an estimated occurrence rate between 1/2983 and 1/6711 [5]. Accordingly the average prevalence of NF1 has been estimated to be about 1/3500 [6-8] with a birth incidence varying from 1/2558 to 1/4292. There is no evidence of ethnicity differences in the frequency of NF1 based on the results of different prevalence studies [9].

CLINICAL FEATURES OF NF1

For the majority of patients, the diagnosis of NF1 is based exclusively on the clinical features. The NF1 diagnosis requires the presence of at least two of the major clinical criteria: six or more CALS, axillary or inguinal freckling, two or more cutaneous neurofibromas, one plexiform neurofibroma, characteristic bony defects (pseudarthrosis, sphenoid wing hypoplasia, and scoliosis), optic glioma, two or more iris Lisch nodules, or a positive family history with a first-degree relative with NF1 [10]. Most of the clinical features associated with NF1 increase in frequency with age among affected children [3, 6, 11-15]. Many infants and young children who are *NF1* gene mutation carriers do not fulfill the clinical diagnostic criteria for NF1 [3, 14, 16-18], but the disease is apparent in almost all affected individuals by 8 years of age and in all affected individuals by age 20 [11, 14].

CALS, pseudarthrosis, and externally visible plexiform neurofibromas can be identified in early infancy, while freckling, optic gliomas, and severe scoliosis occur by the first decade of life. Upon progression, cutaneous neurofibromas and iris Lisch nodules usually appear in teenage to young adult years in NF1 patients [19]. Therefore, NF1 is a progressive condition with variable complications and may worsen with time.

Skin Manifestations

CALS are well circumscribed, evenly pigmented light to dark brown macules averaging 2–5cm in diameter in adults, but they may vary from 2mm to more than 20cm [20]. The frequency of having more than six CALS, which is considered as a cut-off for the NF1 diagnostic criteria, is rare in normal individuals. The CALS must be at least 0.5 cm in diameter in prepubertal individuals and 1.5 cm in postpubertal patients. In addition to NF1, other syndromes associated with one or multiple CALS include McCune-Albright syndrome, Legius syndrome, Noonan syndrome, other cardio-facial-cutaneous syndromes (RASopathies), and constitutional mismatch repair deficiency syndrome [20]. A significant increase in melanocyte and mast cell density in CALS indicates the possible involvement of mast cells and melanocytes in CALS formation [21].

Even though the etiopathogenesis of these skin macules is still elusive, the local fluctuations of keratinocyte-derived membrane-bound stem cell factor (SCF) and mast cell growth factors (MCF) could be important for the development of CALS [22, 23]. Diffuse freckling is another common clinical phenotype in NF1 and clustering hyperpigmentation on the axilla and inguinal area is very common in NF1 patients. Whether the freckling is also associated with SCF and MCF has yet to be determined.

Neurofibroma, a benign tumor derived from cutaneous or subcutaneous nerve sheath is common among NF1 patients. The tumor is comprised of Schwann cells, fibroblasts, perineural cells, mast cells, nerve axons and blood vessels. Some neurofibromas erupt with a broad base of skin involvement, but others are pedunculated lesions. Neurofibroma may be discrete, homogeneous, and well circumscribed, or diffuse, heterogeneous, and infiltrative (plexiform).

Plexiform neurofibromas account for substantial morbidity associated with NF1, including disfigurement, functional impairment, and can be life threatening. Facial dysmorphisms with visual loss are not uncommon in large facial plexiform neurofibromas [24, 25]. In a retrospective survey, six patients with hemifacial hypertrophy were identified to be caused by plexiform neurofibroma with a delayed diagnosis [26]. Histologically, plexiform neurofibromas are similar to cutaneous neurofibromas, but have more extracellular matrix and blood supply. They can arise from the dorsal spinal roots, nervous plexi, large nerve trunks, or sympathetic chains. Malignant transformation into malignant peripheral nerve sheath tumors (MPNST) occurs in approximately 10% of NF1 patients. The diagnosis is often delayed since these cancers typically arise in pre-existing tumors, i.e., neurofibromas. Surgical resection is the mainstay of treatment, and the adjuvant radiation therapy and/or chemotherapy are not well defined, although adjuvant therapy may be of benefit in some patients [27]. The prognosis of patients with incompletely resected, recurrent or metastatic MPNST is dismal.

Skeletal and Osseous Abnormalities

Abnormal skeletal development occurs in roughly 10–20% of NF1 patients [28, 29]. Skeletal manifestations in NF1 patients include bony dysplasia, bony erosion, demineralizing osteoporosis, non-ossifying fibromas, and scoliosis [30]. The most incapacitating problem is congenital pseudarthroses of long bones.

Five percent of NF1 patients develop pseudarthrosis, usually involving the distal one third of the tibia and fibula [31]. Orthopedic surgery to correct NF1 skeletal defects often fails due to pseudarthrosis [28].

Low bone mass, consistent with osteoporosis, has been reported in both young and adult NF1 patients [32]. Fractures of deformed bones occur in 5–10% of young NF1 children, especially boys [31]. Both the demineralization and non-ossifying fibromas can lead to fracture especially in the femur, tibia, and humerus.

Multiple forms of scoliosis occur in at least 10% of NF1 patients. In our NF1 clinic, 25% (17/68) of the NF1 patients were diagnosed with spinal scoliosis [33]. Images from an NF1 patient with both scoliosis and ankylosing spondylitis are shown in Figure 1, implying both developmental and inflammatory processes are involved in the development of skeletal problems in NF1 patients (Figure 1A and 1B). Rapid progressive abnormal curvature of the spine (kyphoscoliosis) needs to be corrected surgically. Fortunately, juvenile scoliosis is often self-limited.

A mouse model lacking a functional *Nf1* allele develops misalignment between vertebral bodies by 1 month of age, scoliosis by 3 months of age and vertebral fusion with severe loss of bone density by 6 months [34]. Blocking RAS/ERK activation by lovastatin during embryonic development reduces the cortical porosity, which suggests the involvement of neurofibromin in RAS/ERK pathway activation. Therefore, the abnormal activity of RAS/ ERK pathway maybe critical for the vertebral and tibia lesions in NF1 patients, and proteins associated with the RAS/ERK pathway may represent good therapeutic targets for NF1 [34].

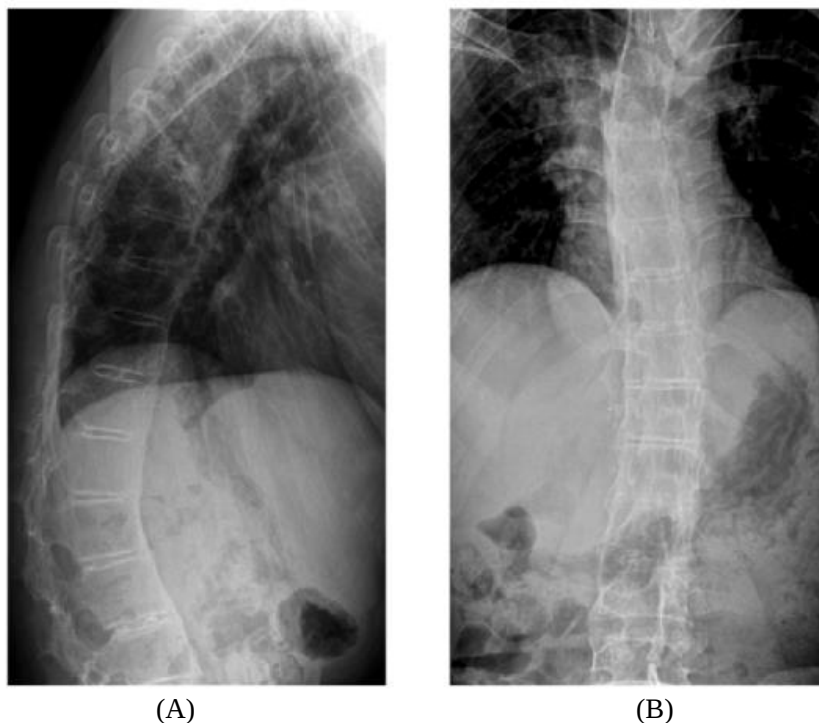


Figure 1. Plain X-ray images of the thoracolumbar spine viewing through lateral (A) and anterior-posterior (B). The patient suffered from scoliosis and ankylosing spondylitis, resulting in vertebral fusion (bamboo spine).

Vascular Lesions

The majority of functional defects in NF1 involve neuroectodermal tissues including nervous tissue, skin, skeletal system, tooth and blood vessels. Structural abnormalities such as overgrowth after intimal injury, dysplasia of the vascular smooth muscle layer and adventitia in both large and small vessels can be identified in NF1 patients [35]. Common vascular complications of NF1 include arterial stenosis, aneurysm, and arteriovenous fistulas involving the abdominal aorta and its branches. We have reported a fatal brain ischemia caused by multifocal stenosis in intracranial arteries with complete occlusion of the left middle cerebral artery in a NF1 patient [36]. Recently, Gao *et al.* also reported a case with coexistence of a left renal artery stenosis and aneurysm detected by the color duplex ultrasound [37].

Psychiatric Disorders and Cognitive Dysfunction

As compared to the general population, psychiatric disorders occur more frequently in patients with NF1 (33% of the patients) [38]. The most common psychiatric problem is dysthymia (21%). There is also a high prevalence of depression, anxiety, and personality disorders in NF1 patients. The impaired quality of life (QoL) associated with NF1 might contribute to the development of psychiatric disorders. Using the questionnaire SF-36 and Skindex, Page *et al.* screened 176 American NF1 patients and showed that the more visible NF1 associated skin disfigurement, the greater the impact on the QoL [39].

Learning disabilities and attention deficit have been reported in as many as 60% of NF1 patients, which may lead to poor performance in school [19]. However, the frequency of major cognitive impairment in our cohort is low [33] and most of our NF1 patients' intelligence falls within normal limits.

Rare Manifestations

In addition to the common clinical features associated with NF1, there are also some rare clinical manifestations such as segmental neurofibromatosis (NF type V) which involves only parts of the body. Most segmental neurofibromatosis patients do not have a positive family history and the disorder has been speculated to be derived from postzygotic *NF1* gene mutation(s). The disease is usually confined to one sector of the body and patients may not have other NF1 stigmata, such as axillary freckling, neurofibroma or CALS [40]. Segmental NF does not progress to the common NF1 phenotype and malignancy is also rare among segmental NF1 patients [40].

RASopathies

A few diseases, such as NF1, Noonan, Costello, Legius and cardio-facio-cutaneous (CFC) syndromes share some clinical features including mental subnormalities, facial dysmorphism, cardiomyopathy, short-stature, dysplasia, bony defects, and cutaneous freckling and

pigmentation. All of these disorders result from aberrations in the RAS-MAPK pathway and they have been grouped together as RASopathies. Relative macrocephaly, ventriculomegaly and Chiari malformation can also occur in patients with RASopathy [41]. Dinçer *et al.* reported that 7 out of 54 NF1 patients suffered from hydrocephalus. Although the types of obstruction vary among patients, the presence of hamartomas was a common finding [42]. Figure 2A and 2B show the obstructive hydrocephalus with large ventricles in one of our NF1 patients. An abnormal flap bridges the fourth ventricle between the vermis and area postrema.

Dysplasia and Developmental Defects

Dysplasia is an abnormal growth of a tissue during development or dysregulated wound healing, while neoplasia is an uncontrolled proliferation of cells. Cortical dysplasia with seizure has been reported in some NF1 patients.

Animal models with neurofibromin haploinsufficiency also resulted in abnormal astrocyte proliferation in different brain regions [43]. In some cases, cortical dysplasia is correlated with ganglioglioma and dysembryoplastic neuroepithelial tumors in NF1 patients (Figure 3A and 3B) [44, 45]. NF1 patients also show wide-spread vasculature dysplasia in endothelial cells and pericytes, especially in the brain, gastrointestinal tract and kidney [36] [46-48]. The common clinical symptoms associated with NF1 including short stature, scoliosis, and joint pseudarthrosis are also dysplastic features of osteoblasts.

Even though some neurofibromas in NF1 patients appear to arise from dysplastic responses to injury, a significant number of them are present from birth, suggesting embryonic or fetal growth abnormalities [49].

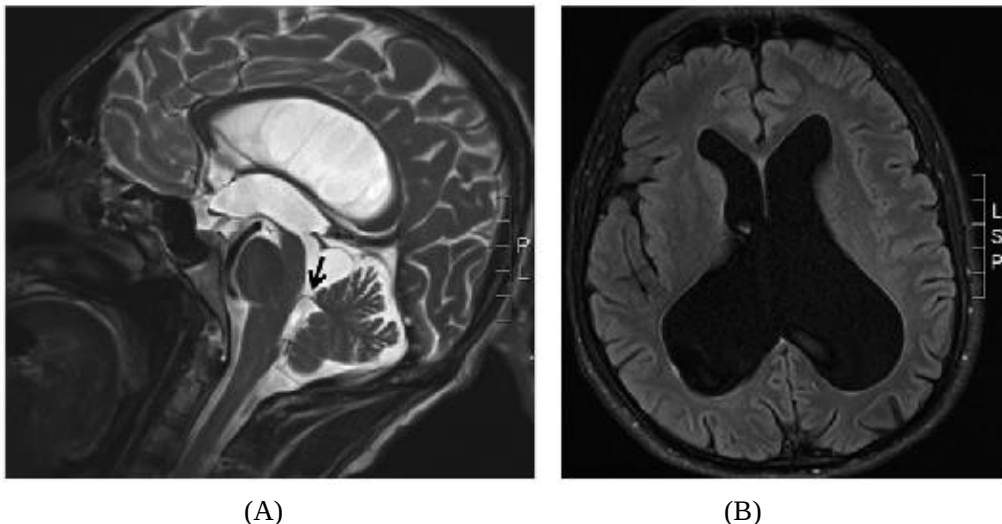


Figure 2. Sagittal T2WI (A) and axial T1WI (B) MRI images of an NF1 patient demonstrates the prominent enlarged ventricles and an abnormal flap bridge (arrow) in the fourth ventricle between the vermis and area postrema. There is no tonsillar herniation, Chiari malformation, progressive forehead bossing or posterior fossa crowding.

Histopathologically, neurofibroma is different from the proliferation of a single cell type frequently seen in neoplasm. Instead, neurofibromas are a mixture of multiple cell types including Schwann cells, fibroblasts, endothelial cells, pericytes, mast cells, occasionally lymphocytes and perineurial cells with extensive extracellular matrix, intercellular collagen and mucopolysaccharide-rich substances [50, 51].

In an animal model, the grafted plexiform neurofibroma will not grow unless neurofibromin-deficient Schwann cells are explanted with *NF1*^{+/-} mast cells [52]. In addition, a few patients report intense itching during the growth of neurofibroma which may indicate the involvement of mast cells [53]. Based on these findings, an initial trial of imatinib mesylate (a receptor tyrosine kinase and cKit inhibitor for mast cells) on one patient with a large mediastinum plexiform neurofibroma, had positive results [53]. Imatinib mesylate has also shown efficacy in treating plexiform neurofibroma in a xenograft model [54]. Studying the interactions among different cell types in neurofibromas will continue to shed light on the mechanisms of tumor formation and growth [55].

MOLECULAR GENETICS OF NF1

Genetic Diagnosis

The human *NF1* gene has 60 exons and encodes a 12-kb transcript covering 350 kb at chromosome 17q11.2 [56, 57]. Mutations of the *NF1* gene consist of small deletions, small insertions, nonsense and missense mutations, intronic mutations, and deletions of the majority of the *NF1* gene or the entire *NF1* gene. Most of the mutations are minor changes while only approximately 5% are total gene deletions [58]. Many germline mutations result in frameshift or nonsense mutations. The mutation rate at the *NF1* gene is estimated to be $\sim 1 \times 10^{-4}$ /gamete/generation making it one of the most frequently mutated genes in human [3, 8]. Approximately 50% of all mutations arise *de novo* and do not appear to be clustered within the gene.

Mutations in the *NF1* gene have been observed in several different types of malignancies including neurofibroma, glioma, MPNST, non-lymphocytic leukemia, and pheochromocytoma.

To date, 1348 publically available mutations of the *NF1* gene have been collected in The Human Mutation Database (HGMD, <http://www.hgmd.cf.ac.uk/ac/gene.php?gene=NF1>) which may explain, in part, the wide spectrum of clinical presentation of NF1. There are probably additional mutations associated with the *NF1* gene. However, the identification of new mutations is hindered by the absence of mutation hotspots and the size of the gene. For the vast majority of patients, the identification of the specific type of NF1 mutation does not currently aid in prognostication, disease management, or treatment selection.

Several methods are used to screen for *NF1* mutations including single-strand conformational polymorphism (SSCP) [59], denaturing gradient gel electrophoresis [60], denaturing high-performance liquid chromatography [33], protein truncation test [61] and fluorescence *in situ* hybridization [62]. Some of these tests must be combined with DNA sequence analysis to document changes leading to abnormal profiles. Loss of heterozygosity of the flanking and intragenic microsatellite markers and the quantification of *NF1* copy number, have been used in detecting large deletions [33].

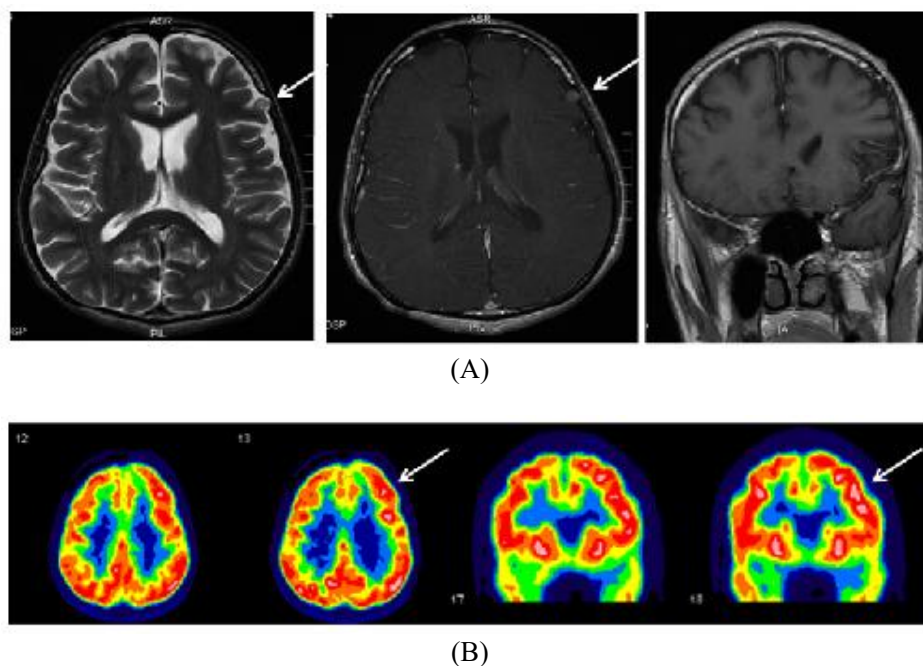


Figure 3. Brain MRI (A) and PET (B) images of an NF1 patient with intractable seizure. The patient demonstrates a suspected dysembryoplastic neuroepithelial tumor (DPNT) in the left frontal region (arrow). The ^{18}F -glucose PET images (B) show a high intensity at the tumor site indicating a correlation between the tumor and epilepsy.

Because screening for NF1 mutations is laborious and expensive, a robust, sensitive and cost-effective alternative for the heteroduplex detection or large-scale sequencing would be preferred by genetic laboratories. DNA sequencing-by-hybridization (SBH) was initially proposed by Drmanac and Crkvenjakov [63]. This approach is an indirect sequencing method in which probes designed to read all possible sequences in any DNA sample hybridizing to DNA template molecules.

A further advance of SBH technology is combinatorial SBH (cSBH), which utilizes a DNA ligation step for two short probes, in which one is attached to a solid support (a glass array slide) and the other is free in solution and labeled with a fluorophore. When both array-bound and solution-phase labeled probes hybridize to the target DNA at contiguous complementary positions, they are covalently linked by DNA ligase, creating one long labeled probe attached to the array surface.

The combinatorial process generates all possible probes that are complementary to the target. Using the cSBH method, Schirinz *et al.* screened 30 NF1 patients and identified 25 mutations with high accuracy and readability [64].

Although genetic testing for NF1 has been available since the mid-1990s, technical difficulties in sequencing the entire *NF1* region have prevented it from becoming a routine practice in the clinic [65]. The recent development and maturation of next generation sequencing technology provides a new tool to address the complexity of mutations associated with NF1 [66].

Tumors Associated with NF1

Neurofibromas are comprised of multiple cell types. The most prevalent cell type in the tumor is Schwann cells, which have been proposed to be the tumorigenic cell population [67-69]. In addition to haploinsufficiency resulting from germline mutations in the *NF1* gene, complete inactivation of both copies of the *NF1* gene through additional *de novo* somatic mutations have been reported in NF1 related tumors [70-72]. Based on Schwann cell cultures derived from neurofibromas (38 tumors from 9 NF1 patients), Maertens and his colleagues detected 29 somatic mutations (29/38) in the neurofibromin gene, and the majority of the mutations are small alterations (26/29) of the gene [73, 74]. Only 3 in 29 mutations are loss of heterozygosity (LOH). A similar phenomenon was reported in tumors of patients with von Hippel-Lindau syndrome and in patients with retinoblastoma [75-77]. Given the high occurrence of somatic mutations, reduced DNA repair efficiency could be a trigger in neurofibroma development. The frequent changes in the genome of cancer cells led to the hypothesis of genetic instability as a prerequisite factor for cancer progression. MPNST in NF1 patients usually originates from plexiform neurofibroma. To elucidate the transition, Kobayashi *et al.* karyotyped 10 MPNSTs from nine patients and found more chromosomal aberrations as a consequence of chromosomal instability in MPNST [78]. However, these results must be interpreted with caution due to inadequate detection techniques, the limited number of samples and the mixed cell population that were used in the studies.

Disease Modifier Genes in NF1

The high inter- and intra-familial variability of NF1 symptoms suggests the possible involvement of modifier genes and/or environmental factors in NF1.

Identifying these modifier genes or environmental factors could provide new insights into the underlying mechanism of the disease and offer more effective therapeutic or disease management approaches. Atypical NF1-like symptoms may be the result of mutations in genes other than *NF1* [79], and these mutations may also affect the symptoms and severity of *NF1* gene-associated neurofibromatosis. Clinical statistics show that patients who are homozygous or compound heterozygous with *MSH6* (mutS homolog 6) mutations, a mismatch repair gene, will develop NF1 like symptoms [80, 81]. The expression level of *MSH6* gene also correlates with the number of café-au-lait spots, which is one of the key clinical phenotypes of NF1. In addition to *MSH6*, individuals with *SPRED1* (sprouty-related, EVH1 domain containing 1) mutations also display NF1-like disorders. Easton and his colleague showed monozygotic twins have the highest *NF1* gene genotype and neurofibromatosis phenotype correlation followed by first-degree relatives and more distant relatives. The high degree of correlation between monozygotic twins suggests a strong genetic component in NF1 pathology, but the low correlation between distant relatives suggests additional genetic factors other than the *NF1* locus may play a role in the pathology of neurofibromatosis [82]. The NF-France Network has phenotyped and genotyped 750 NF1 patients and affected relatives from 275 families. NF1-related clinical features, including five quantitative traits including number and size of café-au-lait spots, and number of cutaneous, subcutaneous and plexiform neurofibromas were scored.

Heritability estimates of those quantitative traits ranged only from 0.26 to 0.62 [83] which suggest the contribution of additional genetic factors to the NF1 clinical symptoms.

However, a separate study of 175 NF1 individuals, including six monozygotic twin pairs, showed high degree of concordance in clinical features between twins and first degree relatives [83]. Two additional studies also demonstrated strong phenotype and *NF1* genotype correlation [82, 84]. In addition, patients with deletions in the *NF1* gene exhibit severe phenotypes with early-onset neurofibroma [85]. Upadhyaya and her colleague also demonstrated that NF1 patients from five unrelated multi-generation families having a single amino acid deletion ('AAT' p.990delM) in exon 17 do not developed cutaneous neurofibromas [86]. These studies revealed a strong phenotype-genotype correlation which argues against the existence of symptom-specific modifier genes for NF1. More recently, a study using a Gene-Chip containing over 900K single nucleotide polymorphism to genotype 300 patients with extreme tumor burden failed to identify a single common polymorphism associated with tumor burden [87].

Molecular Pathogenesis of NF1

NF1 is expressed ubiquitously in different tissues and cell types [88, 89]. Results from immunoprecipitation and Western blot analysis suggest neurological tissues and associated cells including neurons, oligodendrocytes, dorsal root ganglia, and nonmyelinating Schwann cells have the highest concentration of neurofibromin. In rat and human skin, neurofibromin is also expressed in keratinocytes and menalocytes [90].

The *NF1* encoded protein, neurofibromin, contains a GTPase activating protein (GAP) family domain [91]. It has been shown that loss of neurofibromin function is associated with increased levels of activated RAS-GTP in malignant Schwann cells, which suggests the involvement of neurofibromin in modulating cellular RAS-GDP and RAS-GTP levels [92, 93]. In addition, neurofibromin degradation is necessary for maximal RAS activation by growth factors, and neurofibromin expression is required to attenuate the RAS-mediated signaling process [94].

Activation of PKC rapidly induces neurofibromin degradation. McGillicuddy *et al.* demonstrated that only the protein kinase C (PKC) inhibitor bisindolylmaleimide I (Bis I), but not the inhibitors of PI3K or MEK, blocked neurofibromin degradation in response to multiple growth factors in NIH 3T3 cells [95]. These results suggest the involvement of neurofibromin in the growth factor receptor-PKC-RAS signal pathway [95] and NF1 associated clinical phenotypes may be the results of aberrant activation of the RAS pathway [92, 96].

Recently, Yang *et al.* demonstrated that transforming growth factor- β (TGF β) secreted by *Nf1*^{+/-} mast cells can increase the proliferation, migration and collagen synthesis of the *Nf1*^{+/-} fibroblasts [97]. This TGF β -induced change in fibroblasts was mediated through RAS-dependent activation of c-abl [97].

The same group further demonstrated the need of proper microenvironment for the growth of neurofibroma from NF1 deficient Schwann cells in a xenograft animal model. After transplanting human *NF1*^{-/-} Schwann cells in bone marrow, the growth of neurofibroma from nerve roots can be identified in the recipient mice with a defective *Nf1* allele in bone marrow [53], while those in the wild type did not develop any neoplasm.

Based on the understanding of the RAS-c-abl and c-Kit signaling pathway, these investigators demonstrated the inhibition of neurofibroma growth in animal model by treating the animal with imatinib mesylate which is an inhibitor for c-Kit tyrosine kinase. Imatinib has

also been successfully tested on a patient with an enlarged plexiform neurofibroma on the neck and mediastinum [53].

Neurofibromin may have additional non-RAS related functions, including regulating intracellular cAMP level [98]. The phenotype associated with *NF1*-deficiency in *Drosophila* can be rescued by over expression of cAMP dependent activated protein kinase A but not by manipulating RAS signaling [99]. Tong *et al.* also showed that the cAMP-related phenotypes of these flies could be rescued by the expression of a human *NF1* transgene [100].

In mouse *Nf1*^{-/-} astrocyte culture, loss of neurofibromin was found to have a positive influence on cAMP generation and activation of cAMP regulatory targets [101]. In yeast, it was shown that the C-terminal to the GAP related domain (GRD) region of the neurofibromin homologue, Ira1 and Ira2 proteins (RasGAP protein), interact with the kelch Gb proteins Gpb1/2 to stabilize the complex leading to unchecked cAMP–protein kinase A signaling. Thus, the destabilization of the kelch protein–neurofibromin complex may facilitate tumor initiation [102].

Dasgupta *et al.*, using a proteomics-based approach, demonstrated that loss of neurofibromin in astrocytes results in hyperactivation of RAS-dependent and phosphatidylinositol 3-kinase (PI3)-dependent rapamycin (mTOR) signaling pathway [103]. In both *Nf1* mutant mouse optic nerve glioma and in human NF1-associated pilocytic astrocytoma tumors, high levels of ribosomal S6 activation were found to be involved in hyperactivation of the mTOR pathways. Inhibition of the mTOR pathway results in the improvement of *Nf1*^{-/-} astrocyte growth *in vitro* [103]. AKT signaling in fibroblasts of the *Nf1* null mouse was found to be aberrantly activated leading to S6 kinase hyper-phosphorylation [104]. Wortmannin, a PI3 kinase inhibitor, blocks inappropriate activation of mTOR, which depends on the RAS/PI3 kinase effector pathway. The primary target for AKT in this pathway is proposed to be tuberlin, and phosphorylation of tuberlin has been shown to inactivate the TSC1/TSC2 complex, which results in subsequent activation of mTOR [105, 106]. Finally, Johansson *et al.* reported that treatment of subcutaneous sporadic MPNST cell xenografts with mTOR inhibitor, RAD001 (Everolimus) significantly delayed tumor growth and decreased vessel permeability within xenografts [107]. The preclinical results support the consideration of RAD001 in treating NF1 associated and sporadic MPNST.

USING SYSTEMS APPROACHES TO STUDY NF1

One of the aims of systems biology is to decipher biological systems at the network level. The systems approach uses global experimental measurements, such as data from the genome, transcriptome, proteome and metabolome, to build testable hypotheses on molecular interactions. Ideally, identifying networks perturbed in diseases will allow us to identify key molecular events underlying the disease, as well as provide new approaches for restoring disease-perturbed networks. We have previously adapted the systems approach to study complex diseases including human interstitial lung diseases, B cell chronic lymphocytic leukemia, chronic obstructive pulmonary disease and a prion disease model [108-111].

In each of these cases, new molecular processes involved in the diseases and key regulatory factors that could provide new therapeutic intervention were identified.

While previous work has elucidated some of the genes likely to be involved in NF1 pathogenesis and has identified some NF1 biomarkers, a thorough systems approach to identify the changes in the molecular networks underlying NF1 remains to be done. The most comprehensive global molecular profiling with NF1 samples to date, including 20 MPNSTs, 37 neurofibroma, 27 Schwannomas, and 13 synovial sarcomas by Subramanian *et al.* revealed inactivation of p53 and down-regulation of miR-34a in most of the MPNST samples [112]. Down-regulation of the tumor suppressor p53 is commonly observed in cancers, and miR-34a has been previously shown to be directly regulated by p53 [113]. Furthermore, after overexpression of miR-34a in the colon cancer cell line HCT116, expression levels of over 1200 genes were changed [113]. However, the potential roles of p53 and miR-34a dysregulation in MPNST remain unclear.

To expand the network of genes possibly affected by the down-regulation of p53 and miR-34a in MPNST, the results of miR-34a overexpression [113] were used to construct a gene network, taking into account gene expression patterns, protein-protein interactions, transcription factor and miRNA regulation [114]. One of the more interesting changes observed from this analysis is the identification of E2F transcription family members that are upregulated after miR-34 overexpression. E2F transcription factors are involved in cell cycle regulation and aberrant expression of E2F transcription factors has previously been reported in other cancers. This suggests the possibility that part of the dysregulation underlying MPNST is caused through changes in E2F transcriptional activity. Regulatory networks such as these can be useful because they can be applied to longitudinal datasets to follow disease progression.

CURRENT TREATMENT AND DISEASE MANAGEMENT

Despite significant effort, there is currently no effective treatment for NF1. The hallmark of the disease, dermal neurofibromas, although benign, can be painful, debilitating, disfiguring, and can grow large enough to encompass an entire body region. In addition to the high tumor burden for some patients, many cannot be surgically removed because of underlying nerve and vascular involvement. Furthermore, lesions frequently regrow after surgical resection.

It has been shown that dysfunction of several signal transduction elements, including the GTPase RAS, the kinase Src, the tumor suppressor p53 and mTOR-AKT are involved in neurofibroma development. Thus, reagents that block or reverse the abnormal activities related to these pathways may have therapeutic potential in NF1 treatment. As RAS pathway activation is an important feature involved in NF1 associated malignancies, several RAS pathway inhibitors are currently at different stages of clinical evaluation.

To date, there are a number of drugs and treatment strategies in various stages of clinical trials (<http://clinicaltrials.gov/ct2/results?term=NF1andpg=1>) (Table 1). A majority of the trials are aimed at patients with plexiform neurofibroma, including pirfenidone, PEG-interferon alpha-2b, imatinib, sirolimus, vinblastine with methotrexate, and photodynamic therapy [115]. Some of these trials are listed below.

Tipifarnib (IR115777): Tipifarnib is an inhibitor for farnesyl transferase, a key enzyme for RAS activation, which transfers a farnesyl group from farnesyl pyrophosphate (FPP) to the pre-Ras protein. A phase 1 clinical trial with tipifarnib including NF1 patients with plexiform

neurofibromas has been conducted by Widemann et al. [116]. A subsequent phase 2 study has been completed, but the results have not yet been published.

Table 1. Summary of clinical trials for NF1 and related conditions registered in US National Institutes of Health (<http://clinicaltrials.gov/ct2/results?term=NF1andpg=1>) and in the recent report from Children's Tumor Foundation Annual Meeting [87]

Condition	NF1	Plexiform neurofibroma	Low-grade glioma associated with NF1	Unresectable meningioma, intracranial hemangioblastoma	MPNST
Agent	Lovastatin	Imatinib Methylete	Tarceva and Rapamycin	Sunitinib Malate	Filgrastim, doxorubicin hydrochloride, etoposide, ifosfamide, conventional surgery, and radiation therapy
	Skincerity plus sirolimus/ rapamycin	Sorafenib	RAD001 (Everolimus)	Bevacizumab	Imatinib Mesylate
	Methyphenidate	Nilotinib	Carboplatin, Vincristine		Bevacizumab and Everolimus
	Pirfenidone	R115777 (Tipifarnib)	Sorafenib		
	Local injection of ranibizumab	Sirolimus	Lenalidomide		
	Imiquimod 5% Cream	Pirfenidone			
	Cediranib maleate	PEG-interferonalpha-2b			
	Erbium-YAG laser vaporization	Cediranib maleate			
	Sorafenib	Peg-Interferon alpha-2b, Celecoxib, Temozolomide, Vincristine			
	Vitamin D3	Mexotrexat, Vincristine			
		LS 11 (Talaporfin sodium)			

Sorafenib: Plexiform neurofibroma may be susceptible to the inhibition of Ras/Raf/mitogen-activated protein kinase (MAPK) pathway.

Thus, sorafenib, an oral receptor tyrosine kinase inhibitor has been evaluated in children with inoperable plexiform neurofibromas and MPNST [115], with publication of the final study results pending.

Sirolimus (Rapamycin): Sirolimus is an antifungal antibiotic. It inhibits cell cycle progression through the inactivation of mTOR kinase and has anti-proliferative effects [117]. Sirolimus is being evaluated for the treatment of plexiform neurofibromas in a phase 2 study.

Cediranib: Cediranib is a small molecule inhibitor of vascular endothelial growth factor receptor (VEGFR) and is being investigated in patients with NF1 and plexiform neurofibromas.

Pirfenidone: Pirfenidone, an inhibitor for TGF synthesis has been tested in adult patients with NF1 in an open label trial published by Babovic-Vuksanovic *et al.* [118]. Four out of 17 patients had a decrease in tumor volume by 15% or more [118].

Imatinib: Imatinib targets c-kit and PDGFR, and has been reported to be effective in tumor reduction in a subset of patients with plexiform neurofibroma. Symptom improvement was also reported in some patients [87].

HMG-CoA reductase inhibitor: Besides treating NF1 associated tumors, approximately 30% to 65% of children with NF1 suffer from a broad range of both nonverbal and verbal learning disabilities [119]. A Phase I trial of Lovastatin for children with NF1 and cognitive deficits was completed and found improvement of specific cognitive areas, such as memory, recall, and recognition [115].

CONCLUSION

Neurofibromatosis type 1 is a complex disorder, with widely variable clinical features. Anticipatory guidance, surveillance, symptom management and surgical and/or medical treatment of NF1-associated tumors are the main therapeutic approaches in the clinic. Current evidences suggest the haploinsufficiency of neurofibromin is the underlying key pathogenic factor for NF1 and its associated dysplasia and neoplasm. Understand how normal molecular network affected by dysfunctional neurofibromin through comprehensive systems biology studies may inform new therapeutic approaches in the future.

REFERENCES

- [1] Von Recklinghausen, F., *Über die Multiplen Fibrome der Haut und Ihre Beziehung zu Multiplen Neuromen. [About the Multiple Fibromas of the Skin and Their Relationship to Multiple Neuromas]*. 1882, Berlin: August Hirschwald.
- [2] Samuelsson, B. and R. Axelsson, Neurofibromatosis. A clinical and genetic study of 96 cases in Gothenburg, Sweden. *Acta Derm. Venereol. Suppl.* (Stockh.), 1981. 95: p. 67-71.
- [3] Huson, S. M., et al., A genetic study of von Recklinghausen neurofibromatosis in south east Wales. I. Prevalence, fitness, mutation rate, and effect of parental transmission on severity. *J. Med. Genet.*, 1989. 26(11): p. 704-11.
- [4] Fuller, L. C., B. Cox and R. J. Gardner, Prevalence of von Recklinghausen neurofibromatosis in Dunedin, New Zealand. *Neurofibromatosis*, 1989. 2(5-6): p. 278-83.
- [5] Clementi, M., et al., Neurofibromatosis-1: a maximum likelihood estimation of mutation rate. *Hum. Genet.*, 1990. 84(2): p. 116-8.

- [6] Huson, S. M. and R. Hughes, *The neurofibromatoses: a pathogenetic and clinical overview*. 1994, London: Chapman and Hall.
- [7] Riccardi, V., *Neurofibromatosis: phenotype, natural history, and pathogenesis*. 2nd ed. 1992, Baltimore: Johns Hopkins University Press.
- [8] Upadhyaya, M. and D. N. Cooper, *Neurofibromatosis type 1: from genotype to phenotype*. 1998, Guildford: Biddles.
- [9] Friedman, J. M., Epidemiology of neurofibromatosis type 1. *Am. J. Med. Genet.*, 1999. 89(1): p. 1-6.
- [10] National Institutes of Health Consensus Development Conference Statement on Acoustic Neuroma, December 11-13, 1991. The Consensus Development Panel. *Arch. Neurol.*, 1994. 51(2): p. 201-7.
- [11] DeBella, K., J. Szudek and J. M. Friedman, Use of the national institutes of health criteria for diagnosis of neurofibromatosis 1 in children. *Pediatrics*, 2000. 105(3 Pt 1): p. 608-14.
- [12] Friedman, J. M. and P. H. Birch, Type 1 neurofibromatosis: a descriptive analysis of the disorder in 1,728 patients. *Am. J. Med. Genet.*, 1997. 70(2): p. 138-43.
- [13] Niimura, M., Neurofibromatosis in Japan, In: *Tuberous sclerosis and neurofibromatosis: epidemiology, pathophysiology, biology, and management.*, Y. Ishibashi and Y. Hori, Editors. 1990, Elsevier: Amsterdam. p. p 23-31.
- [14] Obringer, A. C., A. T. Meadows and E. H. Zackai, The diagnosis of neurofibromatosis-1 in the child under the age of 6 years. *Am. J. Dis. Child*, 1989. 143(6): p. 717-9.
- [15] Wolkenstein, P., et al., Usefulness of screening investigations in neurofibromatosis type 1. A study of 152 patients. *Arch. Dermatol.*, 1996. 132(11): p. 1333-6.
- [16] Goldberg, Y., et al., Neurofibromatosis type 1--an update and review for the primary pediatrician. *Clin. Pediatr. (Phila.)*, 1996. 35(11): p. 545-61.
- [17] Korf, B. R., Diagnostic outcome in children with multiple cafe au lait spots. *Pediatrics*, 1992. 90(6): p. 924-7.
- [18] Riccardi, V. M., Von Recklinghausen neurofibromatosis. *N Engl. J. Med.*, 1981. 305(27): p. 1617-27.
- [19] Tonsgard, J. H., Clinical manifestations and management of neurofibromatosis type 1. *Semin. Pediatr. Neurol.*, 2006. 13(1): p. 2-7.
- [20] Shah, K. N., The diagnostic and clinical significance of cafe-au-lait macules. *Pediatr. Clin. North Am.*, 2010. 57(5): p. 1131-53.
- [21] Okazaki, M., et al., Correlation between age and the secretions of melanocyte-stimulating cytokines in cultured keratinocytes and fibroblasts. *Br. J. Dermatol.*, 2005. 153 Suppl. 2: p. 23-9.
- [22] Wehrle-Haller, B., The role of Kit-ligand in melanocyte development and epidermal homeostasis. *Pigment. Cell. Res.*, 2003. 16(3): p. 287-96.
- [23] Wehrle-Haller, B., M. Meller and J. A. Weston, Analysis of melanocyte precursors in Nf1 mutants reveals that MGF/KIT signaling promotes directed cell migration independent of its function in cell survival. *Dev. Biol.*, 2001. 232(2): p. 471-83.
- [24] Korf, B. R., Plexiform neurofibromas. *Am. J. Med. Genet.*, 1999. 89(1): p. 31-7.
- [25] Woodruff, J. M., Pathology of tumors of the peripheral nerve sheath in type 1 neurofibromatosis. *Am. J. Med. Genet.*, 1999. 89(1): p. 23-30.
- [26] Overdiek, A., et al., Diagnostic delay of NF1 in hemifacial hypertrophy due to plexiform neurofibromas. *Brain Dev.*, 2006. 28(5): p. 275-80.

- [27] Stucky, C. C., et al., Malignant peripheral nerve sheath tumors (MPNST): the Mayo Clinic experience. *Ann. Surg. Oncol.* 19(3): p. 878-85.
- [28] Crawford, A. H. and E. K. Schorry, Neurofibromatosis in children: the role of the orthopaedist. *J. Am. Acad. Orthop. Surg.*, 1999. 7(4): p. 217-30.
- [29] Riccardi, V. M., J. E. Womack and T. Jacks, Neurofibromatosis and related tumors. Natural occurrence and animal models. *Am. J. Pathol.*, 1994. 145(5): p. 994-1000.
- [30] Lee, M. J. and D. A. Stephenson, Recent developments in neurofibromatosis type 1. *Curr. Opin. Neurol.*, 2007. 20(2): p. 135-41.
- [31] Gilbert, A. and R. Brockman, Congenital pseudarthrosis of the tibia. Long-term followup of 29 cases treated by microvascular bone transfer. *Clin. Orthop. Relat. Res.*, 1995(314): p. 37-44.
- [32] Kuorilehto, T., et al., Decreased bone mineral density and content in neurofibromatosis type 1: lowest local values are located in the load-carrying parts of the body. *Osteoporos Int*, 2005. 16(8): p. 928-36.
- [33] Lee, M. J., et al., Identification of forty-five novel and twenty-three known NF1 mutations in Chinese patients with neurofibromatosis type 1. *Hum. Mutat.*, 2006. 27(8): p. 832.
- [34] Wang, W., et al., Mice lacking Nf1 in osteochondroprogenitor cells display skeletal dysplasia similar to patients with neurofibromatosis type I. *Hum. Mol. Genet.*, 2011. 20(20): p. 3910-24.
- [35] Norton, K. K., J. Xu and D. H. Gutmann, Expression of the neurofibromatosis I gene product, neurofibromin, in blood vessel endothelial cells and smooth muscle. *Neurobiol. Dis.*, 1995. 2(1): p. 13-21.
- [36] Tang, S. C., et al., Novel mutation of neurofibromatosis type 1 in a patient with cerebral vasculopathy and fatal ischemic stroke. *J. Neurol. Sci.*, 2006. 243(1-2): p. 53-5.
- [37] Gao, J., A. Fisher and J. Chung, Color duplex ultrasonography in detecting renal artery abnormalities in a patient with neurofibromatosis 1: a case report. *Clin. Imaging*, 2006. 30(2): p. 140-2.
- [38] Belzeaux, R. and C. Lancon, [Neurofibromatosis type 1: psychiatric disorders and quality of life impairment]. *Presse Med.*, 2006. 35(2 Pt 2): p. 277-80.
- [39] Page, P. Z., et al., Impact of neurofibromatosis 1 on Quality of Life: a cross-sectional study of 176 American cases. *Am. J. Med. Genet. A*, 2006. 140(18): p. 1893-8.
- [40] Dang, J. D. and P. R. Cohen, Segmental neurofibromatosis and malignancy. *Skinmed*, 2010. 8(3): p. 156-9.
- [41] Gripp, K. W., et al., High incidence of progressive postnatal cerebellar enlargement in Costello syndrome: brain overgrowth associated with HRAS mutations as the likely cause of structural brain and spinal cord abnormalities. *Am. J. Med. Genet. A*, 2010. 152A(5): p. 1161-8.
- [42] Dincer, A., U. Yener and M. M. Ozek, Hydrocephalus in patients with neurofibromatosis type 1: MR imaging findings and the outcome of endoscopic third ventriculostomy. *AJNR Am. J. Neuroradiol.*, 2011. 32(4): p. 643-6.
- [43] Bajenaru, M. L., et al., Neurofibromatosis 1 (NF1) heterozygosity results in a cell-autonomous growth advantage for astrocytes. *Glia*, 2001. 33(4): p. 314-23.
- [44] Ostertun, B., et al., Dysembryoplastic neuroepithelial tumors: MR and CT evaluation. *AJNR Am. J. Neuroradiol.*, 1996. 17(3): p. 419-30.

- [45] Prayson, R. A., Composite ganglioglioma and dysembryoplastic neuroepithelial tumor. *Arch. Pathol. Lab. Med.*, 1999. 123(3): p. 247-50.
- [46] Munchhof, A. M., et al., Neurofibroma-associated growth factors activate a distinct signaling network to alter the function of neurofibromin-deficient endothelial cells. *Hum. Mol. Genet.*, 2006. 15(11): p. 1858-69.
- [47] Teixeira, F., et al., Vascular changes in cutaneous neurofibromas. *Neurofibromatosis*, 1988. 1(1): p. 5-16.
- [48] Wu, M., M. R. Wallace and D. Muir, Nf1 haploinsufficiency augments angiogenesis. *Oncogene*, 2006. 25(16): p. 2297-303.
- [49] Riccardi, V., The potential role of trauma and mast cells in the pathogenesis of neurofibromas., In *Tuberous sclerosis and neurofibromatosis: epidemiology, pathophysiology, biology and management.*, Ishibashi, Y. and H. Y., Editors. 1990, Elsevier: Amsterdam. p. pp. 167-190.
- [50] Harkin, J. C. and R. J. Reed, Tumors of the peripheral nervous system., In: *Atlas of tumor pathology*. 1969, Armed Forces Institute of Pathology: Washington, D.C. p. PP. 67-106.
- [51] Jaakkola, S., et al., Type 1 neurofibromatosis: selective expression of extracellular matrix genes by Schwann cells, perineurial cells, and fibroblasts in mixed cultures. *J. Clin. Invest.*, 1989. 84(1): p. 253-61.
- [52] Staser, K., F. C. Yang and D. W. Clapp, Mast cells and the neurofibroma microenvironment. *Blood*, 2010. 116(2): p. 157-64.
- [53] Yang, F. C., et al., Nf1-dependent tumors require a microenvironment containing Nf1+/- and c-kit-dependent bone marrow. *Cell*, 2008. 135(3): p. 437-48.
- [54] Demestre, M., et al., Imatinib mesylate (Glivec) inhibits Schwann cell viability and reduces the size of human plexiform neurofibroma in a xenograft model. *J. Neurooncol.*, 2010. 98(1): p. 11-9.
- [55] Yang, F. C., et al., Neurofibromin-deficient Schwann cells secrete a potent migratory stimulus for Nf1+/- mast cells. *J. Clin. Invest.*, 2003. 112(12): p. 1851-61.
- [56] Viskochil, D., et al., Deletions and a translocation interrupt a cloned gene at the neurofibromatosis type 1 locus. *Cell*, 1990. 62(1): p. 187-92.
- [57] Wallace, M. R., et al., Type 1 neurofibromatosis gene: identification of a large transcript disrupted in three NF1 patients. *Science*, 1990. 249(4965): p. 181-6.
- [58] Wimmer, K., et al., Spectrum of single- and multiexon NF1 copy number changes in a cohort of 1,100 unselected NF1 patients. *Genes Chromosomes Cancer*, 2006. 45(3): p. 265-76.
- [59] Upadhyaya, M., et al., Characterisation of germline mutations in the neurofibromatosis type 1 (NF1) gene. *J. Med. Genet.*, 1995. 32(9): p. 706-10.
- [60] Toliat, M. R., et al., Analysis of the NF1 gene by temperature gradient gel electrophoresis reveals a high incidence of mutations in exon 4b. *Electrophoresis*, 2000. 21(3): p. 541-4.
- [61] Heim, R. A., et al., Distribution of 13 truncating mutations in the neurofibromatosis 1 gene. *Hum. Mol. Genet.*, 1995. 4(6): p. 975-81.
- [62] Messiaen, L. M., et al., Exhaustive mutation analysis of the NF1 gene allows identification of 95% of mutations and reveals a high frequency of unusual splicing defects. *Hum. Mutat.*, 2000. 15(6): p. 541-55.
- [63] Drmanac, R., et al., Sequencing by hybridization: towards an automated sequencing of one million M13 clones arrayed on membranes. *Electrophoresis*, 1992. 13(8): p. 566-73.

- [64] Schirinzi, A., et al., Combinatorial sequencing-by-hybridization: analysis of the NF1 gene. *Genet. Test.*, 2006. 10(1): p. 8-17.
- [65] Han, S. S., D. N. Cooper and M. N. Upadhyaya, Evaluation of denaturing high performance liquid chromatography (DHPLC) for the mutational analysis of the neurofibromatosis type 1 (NF1) gene. *Hum. Genet.*, 2001. 109(5): p. 487-97.
- [66] Hodges, E., et al., Genome-wide in situ exon capture for selective resequencing. *Nat. Genet.*, 2007. 39(12): p. 1522-7.
- [67] Kluwe, L., R. Friedrich and V. F. Mautner, Loss of NF1 allele in Schwann cells but not in fibroblasts derived from an NF1-associated neurofibroma. *Genes Chromosomes Cancer*, 1999. 24(3): p. 283-5.
- [68] Rutkowski, J. L., et al., Genetic and cellular defects contributing to benign tumor formation in neurofibromatosis type 1. *Hum. Mol. Genet.*, 2000. 9(7): p. 1059-66.
- [69] Serra, E., et al., Schwann cells harbor the somatic NF1 mutation in neurofibromas: evidence of two different Schwann cell subpopulations. *Hum. Mol. Genet.*, 2000. 9(20): p. 3055-64.
- [70] Knudson, A. G., Jr., Mutation and cancer: statistical study of retinoblastoma. *Proc. Natl. Acad. Sci. US*, 1971. 68(4): p. 820-3.
- [71] Legius, E., et al., Somatic deletion of the neurofibromatosis type 1 gene in a neurofibrosarcoma supports a tumour suppressor gene hypothesis. *Nat. Genet.*, 1993. 3(2): p. 122-6.
- [72] Shannon, K. M., et al., Loss of the normal NF1 allele from the bone marrow of children with type 1 neurofibromatosis and malignant myeloid disorders. *N Engl. J. Med.*, 1994. 330(9): p. 597-601.
- [73] Maertens, O., et al., Comprehensive NF1 screening on cultured Schwann cells from neurofibromas. *Hum. Mutat.*, 2006. 27(10): p. 1030-40.
- [74] De Raedt, T., et al., Somatic loss of wild type NF1 allele in neurofibromas: Comparison of NF1 microdeletion and non-microdeletion patients. *Genes Chromosomes Cancer*, 2006. 45(10): p. 893-904.
- [75] Vortmeyer, A. O., et al., Somatic point mutation of the wild-type allele detected in tumors of patients with VHL germline deletion. *Oncogene*, 2002. 21(8): p. 1167-70.
- [76] Whitcup, S. M., et al., Use of microdissection and molecular genetics in the pathologic diagnosis of retinoblastoma. *Retina*, 1999. 19(4): p. 318-24.
- [77] Xing, E. P., et al., Loss of heterozygosity of the Rb gene correlates with pRb protein expression and associates with p53 alteration in human esophageal cancer. *Clin. Cancer Res.*, 1999. 5(5): p. 1231-40.
- [78] Kobayashi, C., et al., Aberrant expression of CHFR in malignant peripheral nerve sheath tumors. *Mod. Pathol.*, 2006. 19(4): p. 524-32.
- [79] Muram-Zborovski, T. M., et al., SPRED 1 mutations in a neurofibromatosis clinic. *J. Child Neurol.*, 2010. 25(10): p. 1203-9.
- [80] Menko, F. H., et al., A homozygous MSH6 mutation in a child with cafe-au-lait spots, oligodendroglioma and rectal cancer. *Fam. Cancer*, 2004. 3(2): p. 123-7.
- [81] Ostergaard, J. R., L. Sunde and H. Okkels, Neurofibromatosis von Recklinghausen type I phenotype and early onset of cancers in siblings compound heterozygous for mutations in MSH6. *Am. J. Med. Genet. A*, 2005. 139A(2): p. 96-105; discussion 96.

- [82] Easton, D. F., et al., An analysis of variation in expression of neurofibromatosis (NF) type 1 (NF1): evidence for modifying genes. *Am. J. Hum. Genet.*, 1993. 53(2): p. 305-13.
- [83] Sabbagh, A., et al., Unravelling the genetic basis of variable clinical expression in neurofibromatosis 1. *Hum. Mol. Genet.*, 2009. 18(15): p. 2768-78.
- [84] Szudek, J., H. Joe and J. M. Friedman, Analysis of intrafamilial phenotypic variation in neurofibromatosis 1 (NF1). *Genet. Epidemiol.*, 2002. 23(2): p. 150-64.
- [85] Kayes, L. M., et al., Deletions spanning the neurofibromatosis 1 gene: identification and phenotype of five patients. *Am. J. Hum. Genet.*, 1994. 54(3): p. 424-36.
- [86] Upadhyaya, M., et al., An absence of cutaneous neurofibromas associated with a 3-bp inframe deletion in exon 17 of the NF1 gene (c.2970-2972 delAAT): evidence of a clinically significant NF1 genotype-phenotype correlation. *Am. J. Hum. Genet.*, 2007. 80 (1): p. 140-51.
- [87] Kalamarides, M., et al., Neurofibromatosis 2011: a report of the Children's Tumor Foundation annual meeting. *Acta Neuropathol.* 123(3): p. 369-80.
- [88] Nishi, T., et al., Differential expression of two types of the neurofibromatosis type 1 (NF1) gene transcripts related to neuronal differentiation. *Oncogene*, 1991. 6(9): p. 1555-9.
- [89] Suzuki, Y., et al., Brain tumors predominantly express the neurofibromatosis type 1 gene transcripts containing the 63 base insert in the region coding for GTPase activating protein-related domain. *Biochem. Biophys. Res. Commun.*, 1991. 181(3): p. 955-61.
- [90] Malhotra, R. and N. Ratner, Localization of neurofibromin to keratinocytes and melanocytes in developing rat and human skin. *J. Invest. Dermatol.*, 1994. 102(5): p. 812-8.
- [91] Xu, G. F., et al., The neurofibromatosis type 1 gene encodes a protein related to GAP. *Cell*, 1990. 62(3): p. 599-608.
- [92] Basu, T. N., et al., Aberrant regulation of ras proteins in malignant tumour cells from type 1 neurofibromatosis patients. *Nature*, 1992. 356(6371): p. 713-5.
- [93] Sherman, L. S., et al., Single cell Ras-GTP analysis reveals altered Ras activity in a subpopulation of neurofibroma Schwann cells but not fibroblasts. *J. Biol. Chem.*, 2000. 275(39): p. 30740-5.
- [94] Cichowski, K., et al., Dynamic regulation of the Ras pathway via proteolysis of the NF1 tumor suppressor. *Genes Dev.*, 2003. 17(4): p. 449-54.
- [95] McGillicuddy, L. T., et al., Proteasomal and genetic inactivation of the NF1 tumor suppressor in gliomagenesis. *Cancer Cell*, 2009. 16(1): p. 44-54.
- [96] DeClue, J. E., et al., Abnormal regulation of mammalian p21ras contributes to malignant tumor growth in von Recklinghausen (type 1) neurofibromatosis. *Cell*, 1992. 69(2): p. 265-73.
- [97] Yang, F. C., et al., Nf1^{+/-} mast cells induce neurofibroma like phenotypes through secreted TGF-beta signaling. *Hum. Mol. Genet.*, 2006. 15(16): p. 2421-37.
- [98] Guo, H. F., et al., Requirement of Drosophila NF1 for activation of adenylyl cyclase by PACAP38-like neuropeptides. *Science*, 1997. 276(5313): p. 795-8.
- [99] The, I., et al., Rescue of a Drosophila NF1 mutant phenotype by protein kinase A. *Science*, 1997. 276(5313): p. 791-4.
- [100] Tong, J., et al., Neurofibromin regulates G protein-stimulated adenylyl cyclase activity. *Nat. Neurosci.*, 2002. 5(2): p. 95-6.

- [101] Dasgupta, B., L. L. Dugan and D. H. Gutmann, The neurofibromatosis 1 gene product neurofibromin regulates pituitary adenylate cyclase-activating polypeptide-mediated signaling in astrocytes. *J. Neurosci.*, 2003. 23(26): p. 8949-54.
- [102] Harashima, T., et al., The kelch proteins Gpb1 and Gpb2 inhibit Ras activity via association with the yeast RasGAP neurofibromin homologs Ira1 and Ira2. *Mol. Cell*, 2006. 22(6): p. 819-30.
- [103] Dasgupta, B., et al., Proteomic analysis reveals hyperactivation of the mammalian target of rapamycin pathway in neurofibromatosis 1-associated human and mouse brain tumors. *Cancer Res.*, 2005. 65(7): p. 2755-60.
- [104] Johannessen, C. M., et al., The NF1 tumor suppressor critically regulates TSC2 and mTOR. *Proc. Natl. Acad. Sci. US*, 2005. 102(24): p. 8573-8.
- [105] Garami, A., et al., Insulin activation of Rheb, a mediator of mTOR/S6K/4E-BP signaling, is inhibited by TSC1 and 2. *Mol. Cell*, 2003. 11(6): p. 1457-66.
- [106] Tee, A. R., et al., Tuberous sclerosis complex gene products, Tuberin and Hamartin, control mTOR signaling by acting as a GTPase-activating protein complex toward Rheb. *Curr. Biol.*, 2003. 13(15): p. 1259-68.
- [107] Johansson, G., et al., Effective in vivo targeting of the mammalian target of rapamycin pathway in malignant peripheral nerve sheath tumors. *Mol. Cancer Ther.*, 2008. 7(5): p. 1237-45.
- [108] Cho, J. H., et al., Systems biology of interstitial lung diseases: integration of mRNA and microRNA expression changes. *BMC Med. Genomics*. 4: p. 8.
- [109] Hwang, D., et al., A systems approach to prion disease. *Mol. Syst. Biol.*, 2009. 5: p. 252.
- [110] Moussay, E., et al., MicroRNA as biomarkers and regulators in B-cell chronic lymphocytic leukemia. *Proc. Natl. Acad. Sci. US*. 108(16): p. 6573-8.
- [111] Ezzie, M. E., et al., Gene expression networks in COPD: microRNA and mRNA regulation. *Thorax*. 67(2): p. 122-31.
- [112] Subramanian, S., et al., Genome-wide transcriptome analyses reveal p53 inactivation mediated loss of miR-34a expression in malignant peripheral nerve sheath tumours. *The Journal of pathology*, 2010. 220(1): p. 58-70.
- [113] Chang, T. C., et al., Transactivation of miR-34a by p53 broadly influences gene expression and promotes apoptosis. *Molecular cell*, 2007. 26(5): p. 745-52.
- [114] Lee, M. J., et al., The systems biology of neurofibromatosis type 1--critical roles for microRNA. *Experimental neurology*, 2012. 235(2): p. 464-8.
- [115] Huson, S. M., et al., Back to the future: Proceedings from the 2010 NF Conference. *Am. J. Med. Genet. A*, 2010.
- [116] Widemann, B. C., et al., Phase I trial and pharmacokinetic study of the farnesyltransferase inhibitor tipifarnib in children with refractory solid tumors or neurofibromatosis type I and plexiform neurofibromas. *J. Clin. Oncol.*, 2006. 24(3): p. 507-16.
- [117] Van Rossum, H. H., et al., Everolimus and sirolimus antagonize tacrolimus based calcineurin inhibition via competition for FK-binding protein 12. *Biochem. Pharmacol.*, 2009. 77(7): p. 1206-12.
- [118] Babovic-Vuksanovic, D., et al., Phase II trial of pirfenidone in adults with neurofibromatosis type 1. *Neurology*, 2006. 67(10): p. 1860-2.
- [119] Rosser, T. L. and R. J. Packer, Neurocognitive dysfunction in children with neurofibromatosis type 1. *Curr. Neurol. Neurosci. Rep.*, 2003. 3(2): p. 129-36.

Chapter 107

IMAGING OF PLEXIFORM NEUROFIBROMAS IN NEUROFIBROMATOSIS TYPE 1

Eva Dombi¹, Nicholas Patronas² and Brigitte Widemann¹

¹National Cancer Institute, Pediatric Oncology Branch,
National Institutes of Health, Bethesda, MD, US

²Department of Radiology, Center for Cancer Research,
National Institutes of Health, Bethesda, MD, US

ABSTRACT

This chapter describes the radiographic appearance and distribution of plexiform neurofibromas (PNs) in NF1. Although PNs are histologically benign tumors, they are the source of significant morbidity including disfigurement, pain, functional impairment and they can undergo malignant transformation. Identifying targeted agents that slow the growth or decrease the size of PNs is an area of intense research. Volumetric MRI analysis of PNs is helpful in detecting treatment response and progression in these large and complex shaped tumors with greater sensitivity compared to 1-dimensional or 2-dimensional measurements.

INTRODUCTION

Plexiform neurofibromas are benign nerve sheath tumors affecting multiple fascicles and branches of peripheral nerves. They can develop on any segment of the nerve from the nerve root to the nerve endings. The prevalence of PNs is estimated around 25% in NF1 patients [1-3]. There is no sex predilection. Deep internal PNs are common and may not be clinically apparent [4]. In one study whole body MRI of 65 consecutive pediatric NF1 patients identified 37 patients with PN (57%), only 16 of those patients (25%) had symptomatic tumors [5]. Most PNs present at a young age, may be congenital, and have a tendency to increase in size, especially in early childhood [6]. Significant PN growth in adults is unusual. Overall tumor burden does not always correlate with clinical symptoms, but rapidly growing lesions are more

likely to become symptomatic. Complete surgical removal is rarely achievable and regrowth after surgery is common. Multiple medical treatment options are currently tested in clinical trials. The goal of this chapter is to illustrate the heterogeneity and complexity of these tumors.

IMAGING APPEARANCE OF PLEXIFORM NEUROFIBROMAS

On CT images, PNs present as low attenuation masses and appear hypodense relative to muscle. On post contrast CT scans, PNs enhance only faintly. MRI is the preferred imaging modality for the evaluation of PN burden. On the T1-weighted images (Figure 1A), large PNs demonstrate homogeneous low signal intensity, but smaller tumors can be isointense to muscles. On T2-weighted images PNs are hyperintense with signal intensity similar to or slightly higher than fatty tissue [7, 8].

The pulse sequence known as short TI inversion recovery (STIR) accentuates the long T2 value of the tumors while it provides fat suppression of the surrounding tissues. This technique allows clear separation of PNs from the adjacent tissues and therefore it is uniquely suitable for automatic tissue segmentation and volumetric analysis of these tumors. On the post contrast scans, peripheral PNs show a variable degree and often heterogeneous enhancement which depends on size and vascularity (Figure 1C). The extent of nerve overgrowth varies among patients. Mild segmental nerve enlargement can appear as string of beads alongside the normal distribution of the peripheral nerve (Figure 2A). Around larger solitary nodes the normal rim of fat layer is preserved (split fat sign) and the entering and exiting nerve can be seen centrally to the node. Massive overgrowth of a large segment of the nerve imparts a rope-like appearance (Figure 2B).

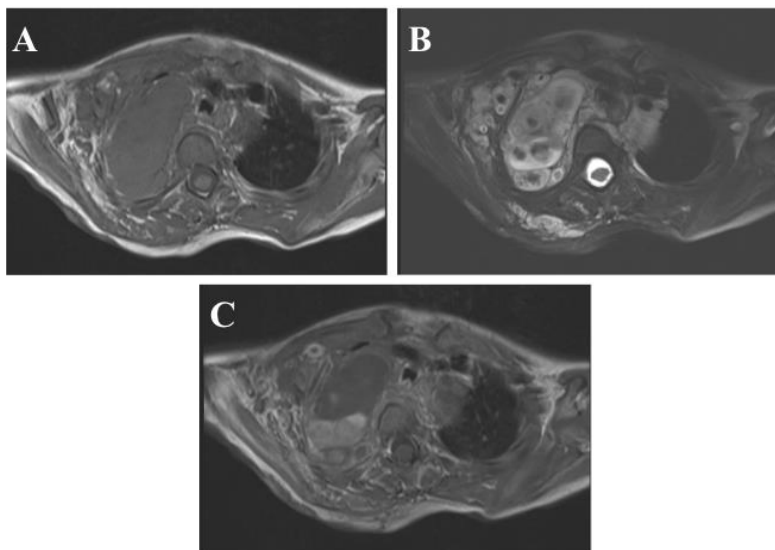


Figure 1. MRI characteristics of PN. Axial T1-weighted image of the chest (A) shows a homogeneous low signal intensity mass. On T2-weighted fat suppressed images (B) the mass appears multinodular with peripheral hyperintensity and central hypointensity of the nodes (target sign). On T1 post-contrast images (C) gadolinium enhancement is variable.

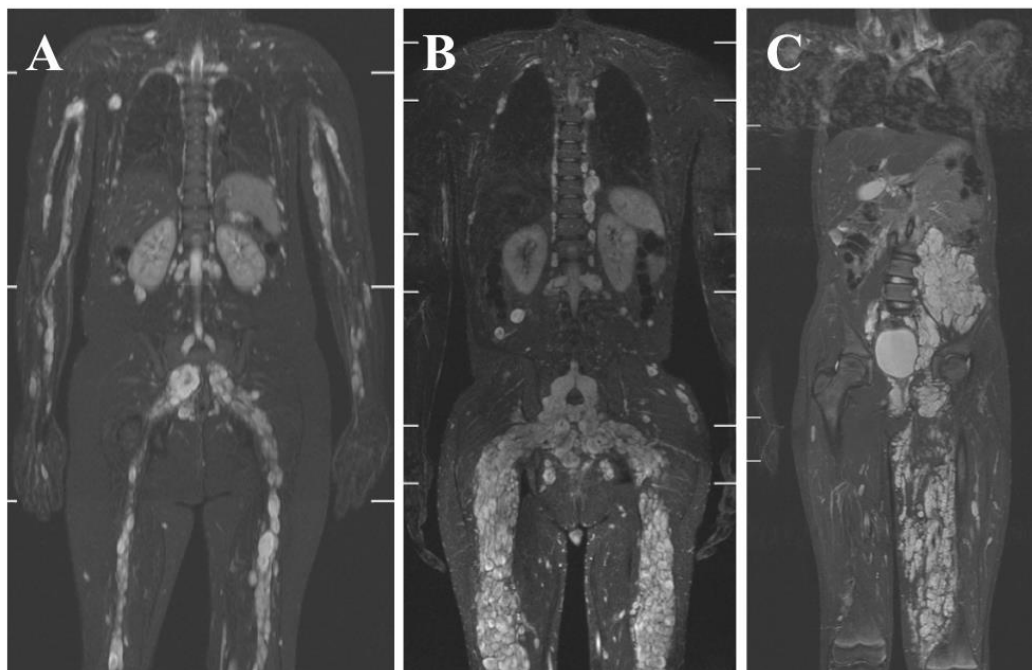


Figure 2. PN distribution. Coronal whole body STIR MR images demonstrate high PN burden in three patients. Case 1: (A) 14 year-old female diagnosed with NF1 at the age of 2 years when she was noted to have a neurofibroma on her scalp. Widespread nodular enlargement of peripheral nerves can be seen on the MRI. Her total PN volume measured by volumetric MRI is 2268 ml, accounting for 5.8% of her body weight. Case 2: (B) 34 year-old male diagnosed with NF1 when he suffered a sports injury as a high school athlete and found to have massive PN burden along his sciatic nerves and lumbosacral spine. He developed cervical cord compression at age 28 that required extensive decompression surgery. He has developed progressive weakness, requires wheelchair assistance, and suffers from chronic neuropathic pain. His whole body tumor burden is 6931 ml, 8.3% of his body weight. Case 3: (C) 13 year-old male with large paraspinal PN that extends to of the left thigh and diffusely infiltrates the thigh muscles. Note the leg length discrepancy and scoliosis of the lumbosacral spine. His whole body tumor burden is 3304 ml, 11.2% of his body weight.

Multiple tumors originating from adjacent nerve branches can form large conglomerate multinodular masses (Figure 2C). The center of each individual nodule contains densely packed collagen fibers and has lower signal intensity on T2-weighted and STIR images while the periphery of the tumor contains less organized myxoid intracellular matrix and appears hyperintense. The combined effect of the central hypointensity with the peripheral hyperintense region has been described as the target sign, or central dot sign and is a unique feature of PNs (Figure 1B). The target sign can also be appreciated on CT and ultrasound [9]. Most PNs display pronounced vascularity.

Plexiform Neurofibroma Types

On large nerves, the epineurium is preserved and fully encapsulates the tumor (displacing tumor type) (Figure 3A). Neurofibromas arising from smaller nerves are locally invasive and extend into the surrounding tissues (infiltrative tumor type) (Figure 3B-E).

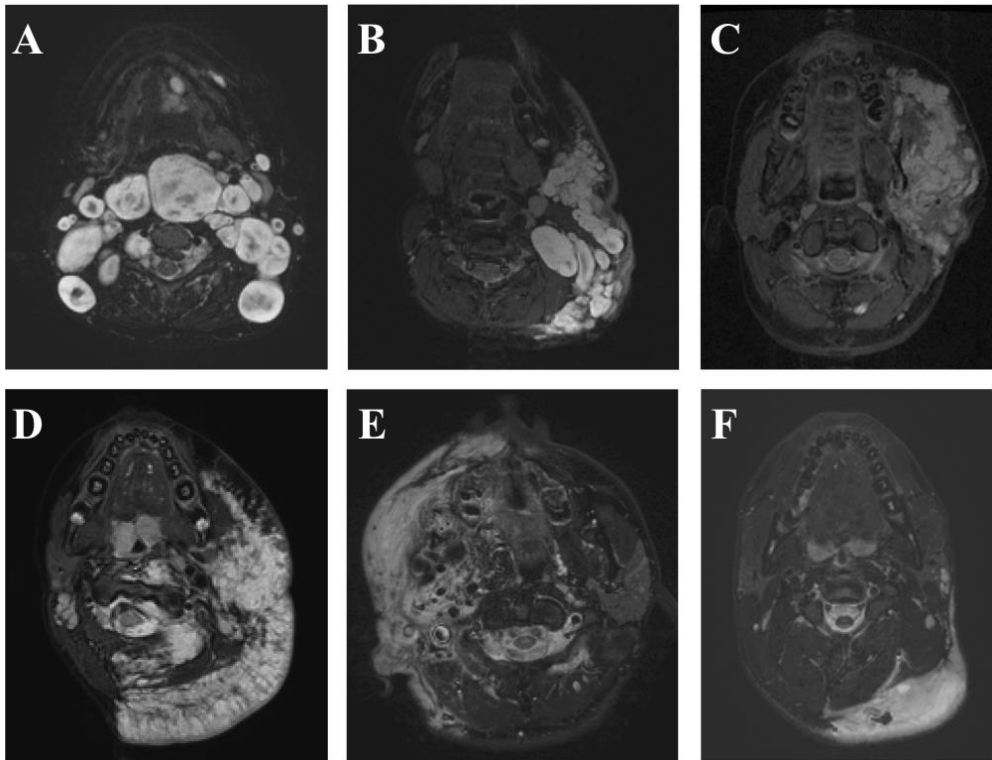


Figure 3. PN types. Imaging characteristics and variety of PN types on axial STIR MR images of the upper neck in 6 different patients. Panel A shows a fully encapsulated displacing multi-lobular tumor. The larger nodes appear heterogeneously hyperintense while the smaller nodes display the characteristic target sign. On panel B the nerve involvement is more peripheral, the nodes are smaller and interlock with the surrounding muscles. The bright signal in the overlaying skin indicates diffuse infiltration of the skin. The PN shown on panel C appears coarsely granular and infiltrative, involving mostly the muscle layer. On panel D a similar tumor is seen invading the skin, subcutaneous fat as well as facial muscles. Panel E shows a PN with a superficial and a highly vascular deep component. Panel F shows a diffuse superficial PN.

Superficial PNs are diffuse, infiltrative tumors involving the upper layers of skin or subcutaneous fat layer (Figure 3F) [7]. Some patients have a predilection for displacing or infiltrative tumor types, while others exhibit mixed features.

Volumetric Analysis

The size of PNs can range from small to very large, span a few centimeters or the full height of the patient. Most PNs have irregular, complex shape and assessing change over time with line measurements used in the response evaluation of solid malignant tumors is challenging [10-12]. Volumetric MRI analysis allows to sensitively and reproducibly determine small changes in PN size that would be impossible to quantify with conventional response criteria. There is a concerted effort (REiNS, Response Evaluation in Neurofibromatosis and Schwannomatosis) to standardize response criteria on clinical trials for NF1 and other related disorders.

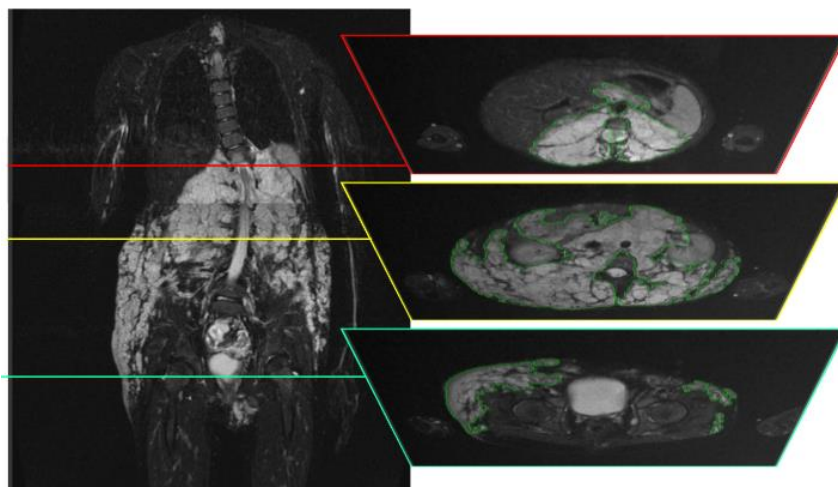


Figure 4. Volumetric MRI analysis of PNs. An outline is generated for every MRI slice that shows PN. The area within the border multiplied by the slice thickness gives the PN volume.

One of the recommendations from the REiNS group is the implementation of volumetric MRI analysis to measure PN size. Different tumor segmentation programs are now available [13-15]. Response evaluation on most clinical trials for PN to date is based on a semi-automatic lesion detection method developed on the MEDx platform [15]. On STIR MR images the bright tumor is separated from the surrounding normal tissue based on signal intensity histogram analysis. The resulting outline is projected over the original image for review (Figure 4). This method has been validated and used successfully in prior and ongoing clinical trials for PNs [16, 17] and also to study the growth behavior and natural history of these tumors [6]. Even with sensitive measurement methods size change is rarely detected within a few months and imaging intervals on clinical trials can be as long as 3-6 months, particularly when the goal of the trial is to increase time to disease progression.

Cranial Nerves

Among the cranial nerves PNs most often affect the trigeminal, glossopharyngeal and vagal nerves. Tumors of the trigeminal nerve can originate from the main trunk in the prepontine cistern or from one of its three branches within the cavernous sinus often causing expansion of this sinus (Figure 5A-B). Unlike other space occupying processes of the cavernous sinus, the slow growing PNs rarely result in overt symptoms. Tumors of the V1 branch can enter the orbit through the superior orbital fissure, leading to the remodeling of the bony structure. Intraorbital PNs of the V1 branch can compress the optic nerve resulting in visual loss. They also commonly produce proptosis of the globe (Figure 5C-D). Sphenoid wing dysplasia can be present with or without PN. Neurofibromas originating from the V2 or V3 branch of the trigeminal nerve may extend into the pterygopalatine fossa or into the soft tissues of the nasopharynx.

The V2 and V3 peripheral branches give rise to tumors in the maxillary or mandibular region. They often involve the tongue, gums and salivary glands. Facial PNs originating from the trigeminal nerve are almost always diffuse, highly invasive and prominently vascular.

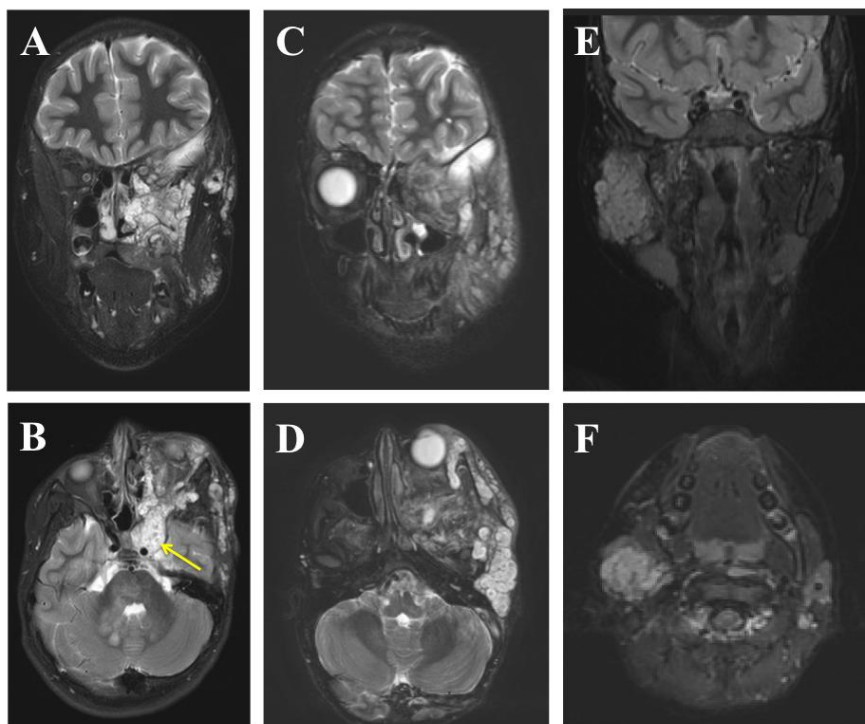


Figure 5. Cranial nerve PNs. Case 4: (A, B) 7 year-old female. At birth protuberance of the left eye was noted leading to the diagnosis of NF1 and orbital PN. Her left face became increasingly disfigured. She underwent a partial resection of the tumor at age 5. She has no vision in her left eye and complains of frequent headaches. Coronal (A) and axial (B) STIR MR images through the orbit show a nodular, invasive PN in the inferior aspect of the orbit and facial muscles. Note the marked expansion of the cavernous sinus (arrow). Total volume of the mass is 162 ml. Case 5: (C, D) 13 year-old male with family history of NF1. At birth only a slight fullness around the left eye was noted, but soon the size of his orbital PN increased and he underwent multiple debulking surgeries by age 3. In addition to the orbital lesion he has asymptomatic deep abdominal and paraspinal PNs. His total tumor burden at age 3 was 269 ml that increased to 800 ml currently, about 29% increase in volume per year. MRI shows a diffuse infiltrative PN of the orbit and deep structures of the face with nodular components in the temporal area. The associated sphenoid wing dysplasia is better appreciated on CT (not shown). Case 6: (E, F) 8 year-old female, diagnosed with NF1 at 15 months of age based on skin findings. Right facial swelling was noted at 18 months. Her PN is confined within the parotid gland. The tumor is associated with mandibular remodeling, temporo-mandibular joint asymmetry, malposition of the lower molars, an open bite and asymmetric smile. She has mild conductive low-frequency hearing loss in the right ear due to narrowing of the external auditory canal. Although the tumor is relatively small, 42 ml, it profoundly impacts her quality of life.

They are often associated with varying degrees of cosmetic deformity and functional impairment. Neurofibromas of the facial nerve are rare; some arise within the petrous bone or the facial canal. PNs of the intracanalicular segment can compress the eighth nerve and clinically present with hearing loss or dizziness. When the origin is at the geniculate ganglion PNs cause bone destruction and project into the middle cranial fossa as an extraaxial mass. Neurofibromas of cranial nerve VII have also been described within the parotid gland (Figure 5E-F) arising from the parotid plexus [18]. Facial nerve injury and paralysis most often develops secondary to surgical procedures. The 9th and 10th nerve can give rise to periauricular, retropharyngeal, occipital and neck masses.

Spinal Nerves

Neurofibromas of the spinal canal can be intradural, extradural or both. The intradural tumors are most often found in the cervical and in the lumbar canal. These tumors are best shown on post contrast T1-weighted scans where they show intense enhancement. The intradural tumors, spinal cord and cerebrospinal fluid are also well visualized using Balanced Fast Field Echo (BFFE) technique (Figure 8). They are rounded or fusiform in configuration and, depending on their size, and deform or compress the spinal cord (Figure 6A–C).

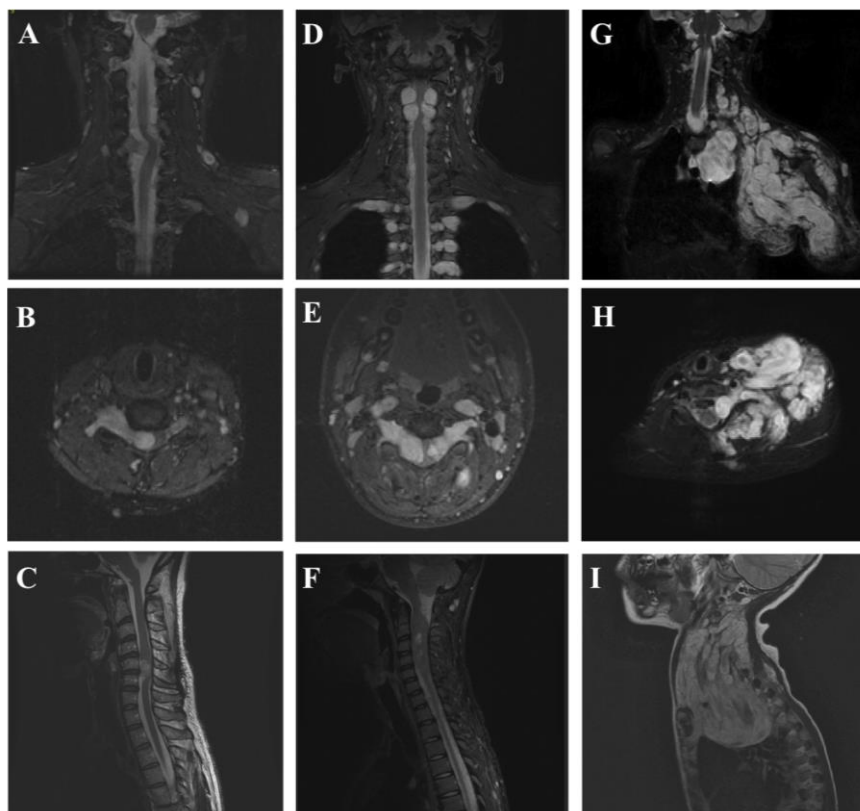


Figure 6. Spinal neurofibromas. Case 7: (A, B, C) 40 year-old female with long-standing history of NF1 who presented with right-sided shoulder pain. Coronal (A), axial STIR (B) and sagittal T2-weighted images (C) of the cervical spine show an intradural neurofibroma on the right at C5/6 level that occupies more than 50% of the canal and displaces the cord to the left. Additional smaller intraforaminal neurofibromas and scattered subcutaneous neurofibromas are also present. Case 8: (D, E, F) 15 year-old male diagnosed with NF1 at a young age. His spinal neurofibromas first manifested at the age of 13 years. The MRI shows large bilateral dumbbell lesions at virtually every nerve root with severe cord compression at C2-5 level resulting in pancake like flattening of the spinal cord. After repeated decompression surgeries he developed swan-neck deformity. Now at age 22 he has limited mobility of the neck, but otherwise good performance status. Case 9: (G, H, I) 5 year-old male diagnosed with NF1 at 6 weeks of age based on multiple cafe-au-lait macules. Increase in the circumference of the left upper arm was first noted at age 1. An MRI obtained at that time showed a large left brachial plexus PN with extension into the left hemithorax. Between age 3 and 5 his PN volume increased more than 3-fold from 498 ml to 1513 ml. The MRI shows a large nodular, fascicular PN extending to the upper arm.

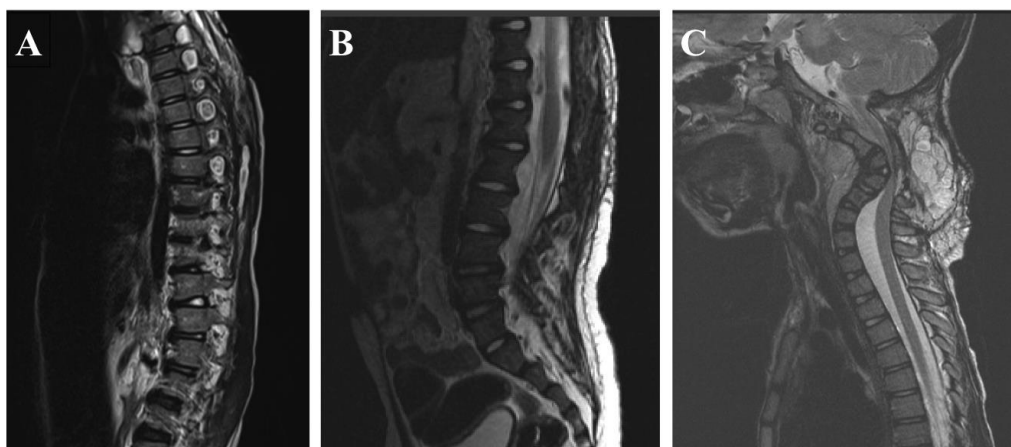


Figure 7. Spine deformities associated with PNs. Case 10: (A) 9 year-old male with multiple paraspinal neurofibromas. Enlargement of the neural foramina is seen on the sagittal T2-weighted MR image of the spine at T2-T6 level. Case 11: (B) 4 year-old male with dural ectasia and scalloping of the thoracic and lumbar vertebrae. Case 12: (C) 5 year-old male with large neck PN and history of cord compression. He developed swan neck deformity after multiple laminectomies.

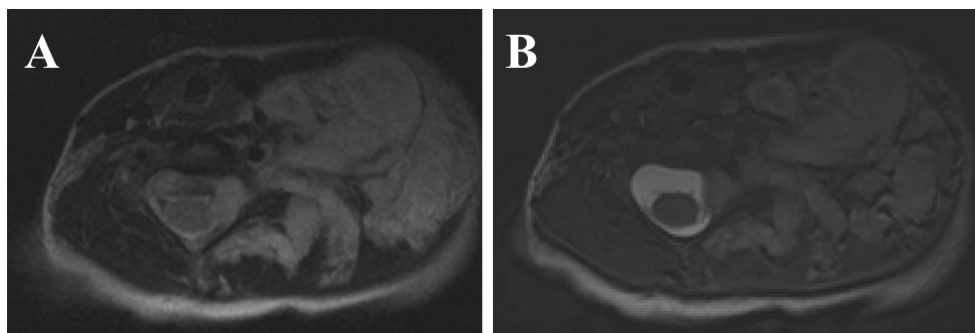


Figure 8. Evaluation of spinal neurofibromas. On axial T2-weighted MR image of the cervical spine (A) the relationship between the spinal cord and C5/6 nerve root is obscured by cerebrospinal fluid (CSF) flow-related signal loss. Balanced Fast Field Echo (BFFE) sequence (B) shows clear definition of the CSF space, spinal cord and extradural neurofibroma on the left side.

Mild spinal cord compression by small tumors can be asymptomatic, while large tumors often produce myelopathy requiring surgical intervention. Extradural tumors are usually located within the neural foramina and often project into the spinal canal causing variable degree of spinal canal stenosis. They are best evaluated by post contrast T1-weighted technique with or without fat suppression. Since they are slowly growing, these tumors tend to gradually erode the bone and produce enlargement of the neural foramina (Figure 7A) and scalloping of the adjacent vertebral body (Figure 7B). Because of the tight confines of the neural foramina, neurofibromas in this location typically assume a dumbbell shape. Spinal cord compression can also occur by extradural tumors, particularly when they happen to be in two neural foramina at the same level and have acquired significant size to project within the spinal canal on both sides (Figure 6D-F). Compression of the spinal cord, either by intradural or extradural tumors, is often accompanied by abnormal signal changes within the cord parenchyma as a result of edema. Cord signal changes that are secondary to edema are reversible after surgical resection

of the tumor and decompression of the cord. Chronic edema however usually leads to alteration of myelin, resulting in myelomalacia that is not reversible and clinically presents with chronic myelopathy. Careful serial neurological evaluation is particularly important in patients with spinal cord compression since development of neurologic deficits can occur with minimal or absent visible changes on imaging studies and may require prompt surgical intervention. Tumor regrowth after surgery occurs frequently and may require multiple revisions. Extensive laminectomies carry the risk of spinal instability that is compounded by NF1 related primary bony dysplasia and manifests as progressive scoliosis, kyphoscoliosis or in most severe form swan-neck deformity of the cervical spine (Figure 7C). Many patients require spine stabilization with metal implants. With current imaging technology, metal artifacts from those implants make radiological evaluation of the spine challenging. The use of titanium reduces this artifact. Enlargement of spinal nerve roots beyond the neural foramina can present as round solitary, fusiform or multilobular masses in the paraspinal soft tissues. In some cases, tumors of the brachial and lumbosacral plexus form giant masses that extend to the extremity and may cause overgrowth of the affected limb (Figure 6G–I). The neurofibromas in the paraspinal regions are best evaluated by STIR MRI technique.

Neck, Mediastinum, Chest

Paraspinal neurofibromas in the thoracic area tend to be less bulky than cervical and lumbosacral tumors, they are most often displacing although invasion into the paraspinal muscles can occur (unpublished observation).

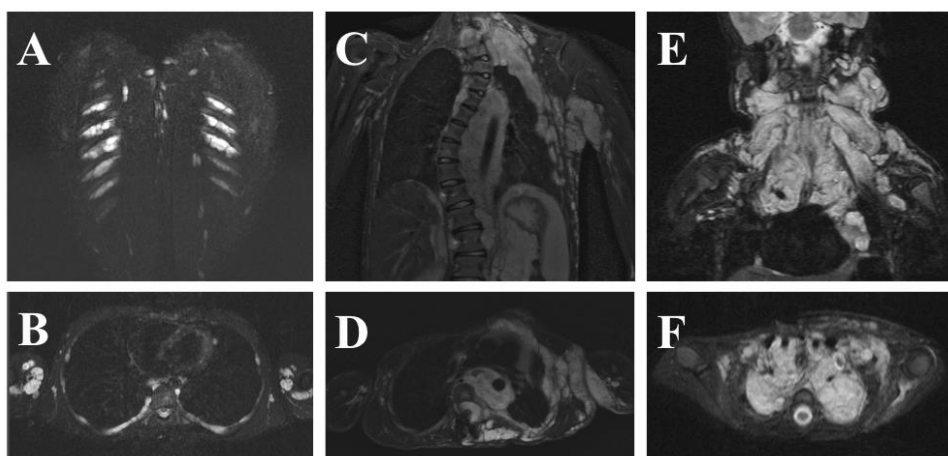


Figure 9. Neck and chest PNs. Case 13: (A, B) 15 year-old male, diagnosed with NF1 at age 10, based on skin findings including subcutaneous neurofibromas. The peripheral nerves throughout his body are enlarged; his whole body tumor burden is 4630 ml, 7.9% of his body weight. The coronal and axial MR images of the chest highlight the thickening of intercostal nerves. Case 14: (C, D) 10 year-old female, diagnosed with NF1 at 18 months of age. A large PN is seen along the descending aorta in the mediastinum and abdomen. She has severe scoliosis and chest deformity resulting in restrictive lung disease. Her brachial plexus PN causes pain in the left axilla and arm. The volume of her tumor is 1175 ml, increased from 263 ml at age 4, at a rate of 75% per year. Case 15: (E, F) 3 year-old male with neck and chest PN and severe airway compression requiring tracheostomy.

Tumors growing along the intercostal nerves (Figure 9A,B) can erode the lower edge of the ribs (ribbon ribs). Mediastinal tumors originate from the vagal nerve or the sympathetic chain (Figure 9C,D). They can be extreme in size, are mostly diffuse, and may cause airway compromise (Figure 9E,F, 10A,B) or swallowing difficulties.

Abdomen, Pelvis

PNs of the autonomic plexus are mostly retroperitoneal [19], some may extend into the mesentery alongside the mesenteric vessels (Figure 11A,B) or more rarely into the liver through the portal system (Figure 11C,D). Bowel obstruction is a rare, but serious complication. Diffuse pelvic tumors can invade the muscle layer of the bladder and cause urinary obstruction (Figure 11E,F) in particular at the vesicoureteric junction. The dense neural network of the uterine wall or the soft tissues of perineum can also give rise to PNs in that area. Tumors of the lumbosacral plexus can be unilateral (Figure 12A,B) or bilateral (Figure 12C,D), and may cause erosive changes to the spine. The concurrent scoliosis can be severe. PN expansion into the paraspinal, iliopsoas, abdominal wall and gluteal musculature is common (Figure 12E,F). Involvement of the sciatic nerve is more frequent than the femoral nerve.

Malignant Peripheral Nerve Sheath Tumors (MPNSTs)

Malignant peripheral nerve sheath tumors are aggressive soft tissue sarcomas with poor prognosis. Rarely seen in the general population, about half the cases are diagnosed in patients with NF1. The lifetime risk of MPNST in NF1 is 8-13% (20, 21). That risk is further elevated if the patient received radiation therapy for optic glioma or other malignancy. The 5-year survival with combination therapy is 21-52% and appears to be worse for individuals with NF1 associated versus sporadic MPNSTs [20, 22-24].

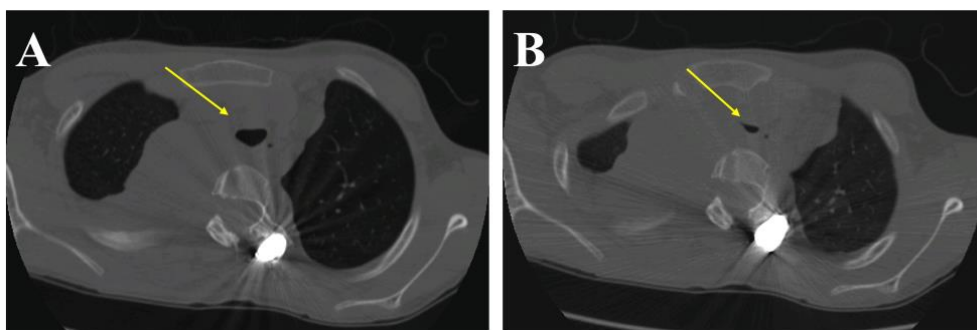


Figure 10. Tracheomalacia. Case 16: 16 year-old male with severe tracheomalacia as a result of a large neck and chest PN. CT of the chest during inspiration (A) shows mild narrowing of the airway (cross-sectional area 1.2 cm²). In expiration phase (B) the narrowing becomes severe (0.2 cm²). He underwent a Y stent placement in the distal trachea that temporarily relieved the symptoms of respiratory distress. He died of probable cardiac arrest a year later.

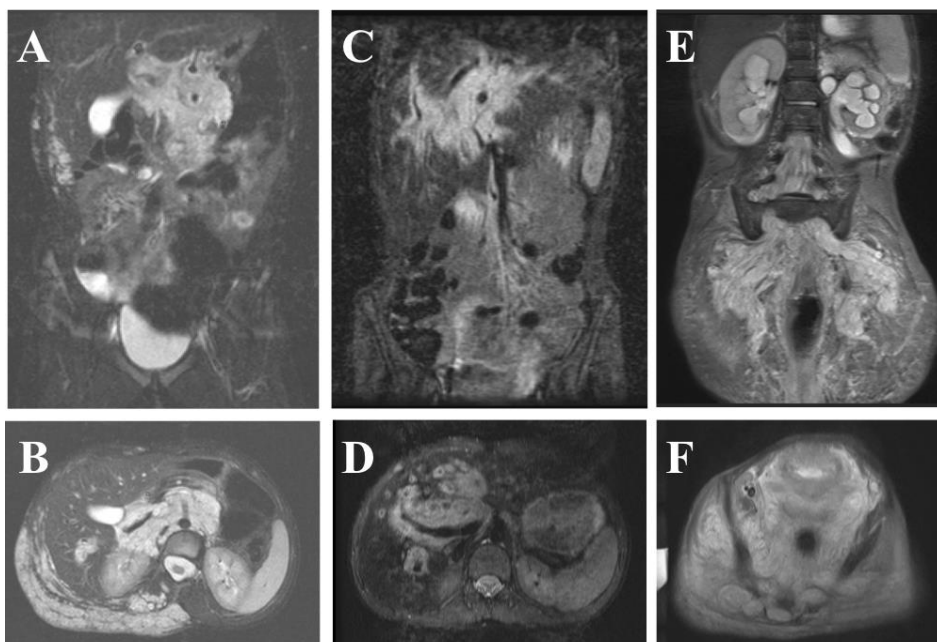


Figure 11. Abdomino-pelvic PNs of the autonomous nervous system. Case 17: (A, B) 8 year-old female. Multiple areas of skin hyperpigmentation were noted at birth; on the right chest wall one large patch of the hyperpigmentation was associated with roughening of the skin indicating a superficial PN. In addition to the PN in the paraspinal muscles and subcutaneous tissues of the back she also has a large retroperitoneal lesion originating from the sympathetic chain. The PN encases the descending aorta, celiac artery, superior mesenteric artery and renal arteries. Case 18: (C, D) 10 year-old female with family history of NF1. She complains of intermittent abdominal pain since age five; her height and weight are around the 10th percentile. The MRI reveals a PN around the porta hepatis and expanding along the biliary system. The size of the PN has not changed significantly within the past 5 years and she maintains normal liver function. Case 19: (E, F) 3 year-old female, multiple café-au-lait macules were noted at birth raising the suspicion of NF1. The large pelvic PN was discovered at age 1 confirming the NF1 diagnosis. The PN surrounds the bowels and invades the bladder wall causing urinary obstruction. She subsequently underwent ureterostomy and the hydronephrosis resolved. Additional neurofibromas fill the lumbosacral thecal sac, expand the spinal canal, the sacral neuroforamina and infiltrate the gluteal muscles. Between age 3 and 8 the volume of her PNs increased from 808 ml to 3685 ml, more than 90% volume change per year.

In the setting of NF1, most MPNSTs (65-88%) arise in preexisting PNs [20, 25-27] and can be large, invasive or metastatic at diagnosis [22]. Patients with high PN burden may be at higher risk for MPNST [28, 29]. The goal of surveillance is early detection with the hope of curative surgery, which requires complete resection of the MPNST. The most predictable clinical sign of malignant degeneration is new or increasing pain.

The pain can present locally at the site of disease, radiate along the affected nerve or project distally from the lesion. The imaging field should be selected to include the entire nerve from the site of pain to the matching nerve root. Numbness, paresthesia, weakness, other functional deficits, change in consistency on palpation or swelling are also common. Change in size, especially if it is focal and out of proportion to the rest of the PN is concerning. Volumetric MRI analysis allows to accurately establish the tumor growth rate and detect accelerating growth sooner compared to line measurements.

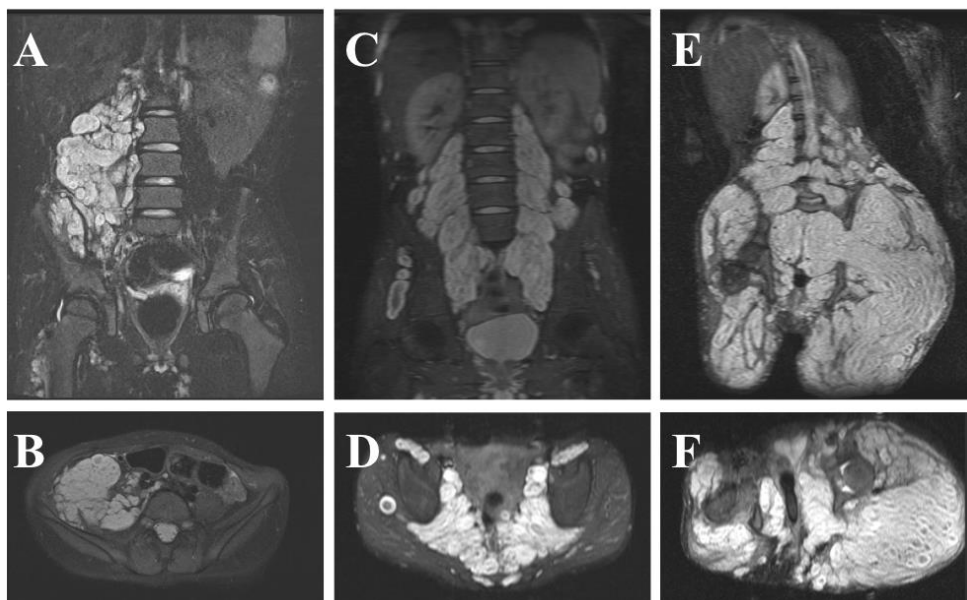


Figure 12. Lumbosacral PNs. Case 20: (A, B) 6 year-old male with family history of NF1. The coronal (A) and axial (B) STIR MR images demonstrate a right-sided lumbosacral and pelvic PN. His tumor first became symptomatic at age 3. He started to turn the right leg outward and his ankle collapsed. Currently he walks with a lower leg brace and requires regular pain medication. Case 21: (C, D) 11 year-old male diagnosed with NF1 in early infancy. By age 5 he was complaining of pain in his legs when walking. His current symptoms include lower extremity weakness that prevents him from bending the knees and severe radiating pain in the back of the legs that he rates 8 out of 10. The MRI shows bilateral paraspinal and deep pelvic PNs. Case 22: (E, F) 7 year-old male with extreme tumor burden; the 5629 ml PN accounts for more than 25% of his body weight. The muscles are atrophied on both lower extremities and he is unable to walk. The bulky gluteal mass makes sitting in the wheelchair uncomfortable and the family is considering amputation.

Other imaging signs that indicate malignant transformation (Figure 13) include changes in the characteristics in some part of the PN, such as disappearance of the target sign, inhomogeneity, patchy contrast enhancement, intra-tumor hemorrhage, necrosis or edema.

None of these signs are specific enough to reliably separate MPNST from a PN. The usefulness of ^{18}F -fluorodeoxyglucose positron emission tomography (FDG-PET) in the detection of MPNSTs has been extensively studied [30-32]. Maximum FDG uptake in benign neurofibromas is generally below 2.0 g/mL, while high grade MPNSTs demonstrate FDG uptake consistently above 3.5 g/mL. However above-background focal uptake can be present in histologically confirmed benign neurofibromas. Therefore, selecting a single diagnostic cut-off that is unequivocal for MPNST is not possible [33]. Diagnostic specificity may be increased with delayed imaging, performed 4 hours after FDG injection. Further increase in uptake on delayed imaging is suggestive of a malignant lesion [32]. FDG-PET is a helpful tool to identify concerning areas in a patient with high PN burden and can be used as a guide to targeted biopsy (Figure 14). Benign lesions with increased FDG uptake, especially those that show cellular atypia, should be closely monitored. There is evidence that atypical neurofibromas harbor chromosomal aberrations reminiscent of MPNST and therefore can be considered premalignant tumors [34].

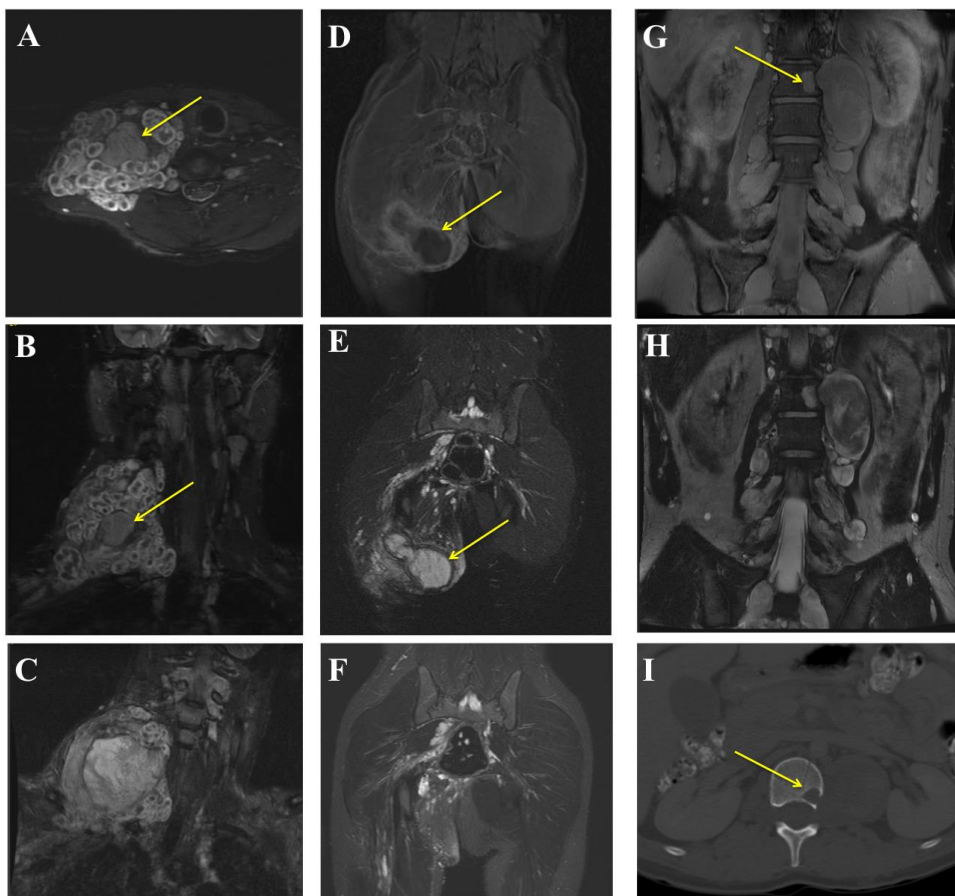


Figure 13. MPNST. Case 23: (A, B, C) 15 year-old male, diagnosed with NF1 and a neck PN at age 7. Axial (A) and coronal (B) STIR MR images show a right brachial plexus tumor, with nodules displaying the characteristic target sign at the periphery and a distinct homogeneous area in the center. The area where the target sign is lost (arrow) corresponds to high grade MPNST within the PN. The tumor was unresectable at diagnosis. Despite intensive chemotherapy the patient had rapid progression (C) and died of the disease 14 months after diagnosis. Case 24: (D, E, F) 15 year-old male, with long standing history of NF1 and large PN burden. He complained of a fast growing painful node in his right buttock. The fat suppressed T1 post contrast MR image (D) shows avid enhancement at the periphery of the nodule (arrow). On STIR sequence (E) the lesion shows inhomogeneous bright signal and there is edema in the surrounding tissues. Biopsy confirmed a low grade MPNST; the lesion was resected with negative margins. One year after surgery there was no sign of recurrence (F), and four years after diagnosis the patient is alive with no evidence of malignancy. Case 25: (G, H, I) 20 year-old male, diagnosed with NF1 in infancy. Persistent lower back pain prompted the MRI that revealed multiple dumbbell lesions and an 8 cm left paraspinal mass eroding the L1 vertebral body (arrow). On T1 post contrast images (G) his tumors show avid enhancement with some inhomogeneity within the larger node. On T2-weighted images (H) the large node displays variable high signal intensity. CT of the spine demonstrates bony erosion without sclerotic rim (I) suggestive of a fast growing tumor. Histology of the completely resected lesion was consistent with low grade MPNST. Four years after surgery the patient is alive with no evidence of malignancy.

Surgical removal of these lesions should be considered, if feasible without untoward morbidity. FDG-PET may also have a role in detecting treatment response and as a prognostic tool for long-term survival in NF1 patients with MPNST [35]. It remains to be seen whether

alternative PET tracers, for example ^{18}F -fluoro-L-thymidine (FLT) that localizes to sites of high cell proliferation, better differentiate benign from malignant nerve sheath tumors.

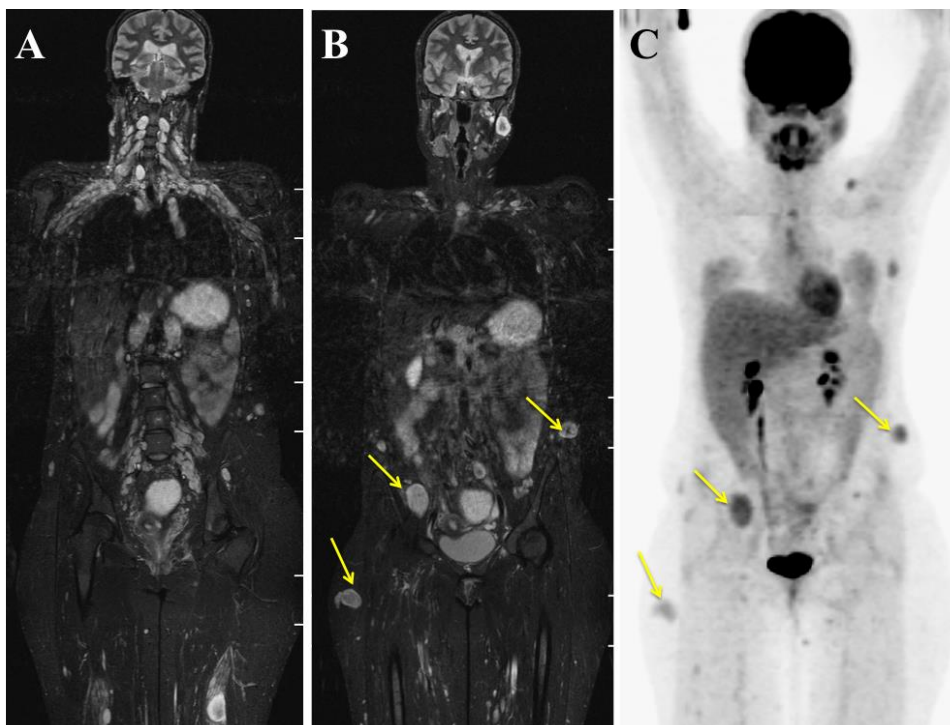


Figure 14. FDG-PET imaging. Case 26: 24 year old female with family history of NF1 and large PN burden involving all the major nerves (A). At age 14 she developed a low grade MPNST in the right calf, underwent surgical resection of the tumor followed by radiotherapy. 4 years later she was found to have a high grade MPNST in the right sacral/ischial area; after complete resection she received adjuvant chemotherapy and radiation. She is at high risk of recurrence and is closely followed by MRI and *yearly* PET scans. The FDG uptake in the majority of her benign tumors is similar to background, however several areas of increased uptake are noted (C), SUV 1.9 g/mL in right thigh, 4.0 g/mL in right iliac fossa, 3.6 g/mL in left flank, corresponding to circumscribed nodules within the PNs (B). The size, imaging appearance and PET avidity of these lesions has been stable in the past 6 years.

CONCLUSION

PNs are the second most frequently diagnosed NF1-associated tumors after dermal neurofibromas. PNs involve multiple nerve branches, tend to be large, and often have a complex shape. PNs are best imaged by MRI. On STIR sequences the clear separation of high signal intensity tumor from the surrounding low signal intensity normal tissue allows for automatic tissue segmentation and volume measurement. Volumetric analysis helps to assess the overall tumor burden that can range from a few milliliters to several liters. In one of the above examples PNs account for 25% of the patient's body weight. Over time volumetric analysis can detect small incremental changes in PN size. Fast growing PNs can as much as double in volume within a year; typically PNs grow more slowly or remain stable for prolonged time periods. Other than young age we do not know what influences PN growth rates. Growth behavior may

depend on tumor type, location, size, vascularity, genetic background, hormone levels or environmental exposures. We hope to get better insight about these factors through our ongoing longitudinal natural history study. The extent of PN involvement often can't be predicted from clinical exam. Screening for asymptomatic internal PNs is not recommended at this time due to the lack of effective treatment options. Recently published clinical trials show promise of identifying active medical treatments for PNs. Some of these therapies appear to be more effective in slowing the tumor growth rather than shrinking the tumors and therefore could be tested as preventive interventions before symptoms appear. The recommendation for screening may be reconsidered if prevention of PN growth proves to be feasible and safe. PNs have a high risk for malignant transformation and there is a need to develop better diagnostic tools for the early detection of MPNSTs.

REFERENCES

- [1] Huson SM, Harper PS, Compston DA. Von Recklinghausen neurofibromatosis. A clinical and population study in south-east Wales. *Brain: a journal of neurology*. 1988;111 (Pt 6):1355-81. Epub 1988/12/01.
- [2] Friedman JM, Birch PH. Type 1 neurofibromatosis: a descriptive analysis of the disorder in 1,728 patients. *American journal of medical genetics*. 1997;70(2):138-43. Epub 1997/05/16.
- [3] Waggoner DJ, Towbin J, Gottesman G, Gutmann DH. Clinic-based study of plexiform neurofibromas in neurofibromatosis 1. *American journal of medical genetics*. 2000;92(2):132-5. Epub 2000/05/08.
- [4] Tonsgard JH, Kwak SM, Short MP, Dachman AH. CT imaging in adults with neurofibromatosis-1: frequent asymptomatic plexiform lesions. *Neurology*. 1998;50(6):1755-60. Epub 1998/06/20.
- [5] Nguyen R, Kluwe L, Fuensterer C, Kentsch M, Friedrich RE, Mautner VF. Plexiform neurofibromas in children with neurofibromatosis type 1: frequency and associated clinical deficits. *The Journal of pediatrics*. 2011;159(4):652-5 e2. Epub 2011/05/31.
- [6] Dombi E, Solomon J, Gillespie AJ, Fox E, Balis FM, Patronas N, et al. NF1 plexiform neurofibroma growth rate by volumetric MRI: relationship to age and body weight. *Neurology*. 2007;68(9):643-7.
- [7] Friedrich RE, Korf B, Funsterer C, Mautner VF. Growth type of plexiform neurofibromas in NF1 determined on magnetic resonance images. *Anticancer research*. 2003;23(2A):949-52.
- [8] Rodriguez D, Poussaint TY. Neuroimaging findings in neurofibromatosis type 1 and 2. *Neuroimaging clinics of North America*. 2004;14(2):149-+.
- [9] Lin J, Jacobson JA, Hayes CW. Sonographic target sign in neurofibromas. *Journal of ultrasound in medicine: official journal of the American Institute of Ultrasound in Medicine*. 1999;18(7):513-7. Epub 1999/07/10.
- [10] Miller AB, Hoogstraten B, Staquet M, Winkler A. Reporting results of cancer treatment. *Cancer*. 1981;47(1):207-14. Epub 1981/01/01.

- [11] Eisenhauer EA, Therasse P, Bogaerts J, Schwartz LH, Sargent D, Ford R, et al. New response evaluation criteria in solid tumours: revised RECIST guideline (version 1.1). *Eur. J. Cancer*. 2009;45(2):228-47. Epub 2008/12/23.
- [12] Sasaki T. [New guidelines to evaluate the response to treatment "RECIST"]. *Gan to kagaku ryoho Cancer & chemotherapy*. 2000;27(14):2179-84. Epub 2001/01/06.
- [13] Cai W, Kassarijian A, Bredella MA, Harris GJ, Yoshida H, Mautner VF, et al. Tumor burden in patients with neurofibromatosis types 1 and 2 and schwannomatosis: determination on whole-body MR images. *Radiology*. 2009;250(3):665-73.
- [14] Poussaint TY, Jaramillo D, Chang Y, Korf B. Interobserver reproducibility of volumetric MR imaging measurements of plexiform neurofibromas. *AJR Am. J. Roentgenol*. 2003;180(2):419-23. Epub 2003/01/24.
- [15] Solomon J, Warren K, Dombi E, Patronas N, Widemann B. Automated detection and volume measurement of plexiform neurofibromas in neurofibromatosis 1 using magnetic resonance imaging. *Comput Med. Imaging Graph*. 2004;28(5):257-65.
- [16] Widemann B, Babovic-Vuksanovic D, Dombi E, Goldman S, Goodspeed W, Goodwin A, et al., editors. Phase II trial of pirfenidone in children and young adults with neurofibromatosis type 1 (NF1) and progressive plexiform neurofibromas (PN). *Children's Tumor Foundation*; 2009; Portland, OR.
- [17] Widemann B, Dombi E, Gillespie A, Belasco J, Goldman S, Korf B, et al., editors. Phase II randomized, flexible cross-over, double-blinded, placebo-controlled trial of the farnesyltransferase inhibitor (FTI) tipifarnib (R115777) in pediatric patients (pts) with neurofibromatosis type 1 (NF1) and progressive plexiform neurofibromas (PN). *Children's Tumor Foundation*; 2009; Portland, OR.
- [18] McGuirt WF, Sr., Johnson PE, McGuirt WT. Intraparotid facial nerve neurofibromas. *The Laryngoscope*. 2003;113(1):82-4. Epub 2003/01/07.
- [19] Zacharia TT, Jaramillo D, Poussaint TY, Korf B. MR imaging of abdominopelvic involvement in neurofibromatosis type 1: a review of 43 patients. *Pediatr Radiol*. 2005;35(3):317-22. Epub 2004/11/02.
- [20] Ducatman BS, Scheithauer BW, Piepgras DG, Reiman HM, Ilstrup DM. Malignant peripheral nerve sheath tumors. A clinicopathologic study of 120 cases. *Cancer*. 1986;57(10):2006-21.
- [21] Evans DG, Baser ME, McGaughran J, Sharif S, Howard E, Moran A. Malignant peripheral nerve sheath tumours in neurofibromatosis 1. *J. Med. Genet*. 2002;39(5):311-4.
- [22] Carli M, Ferrari A, Mattke A, Zanetti I, Casanova M, Bisogno G, et al. Pediatric malignant peripheral nerve sheath tumor: the Italian and German soft tissue sarcoma cooperative group. *J. Clin. Oncol*. 2005;23(33):8422-30.
- [23] deCou JM, Rao BN, Parham DM, Lobe TE, Bowman L, Pappo AS, et al. Malignant peripheral nerve sheath tumors: the St. Jude Children's Research Hospital experience. *Ann. Surg. Oncol*. 1995;2(6):524-9.
- [24] Wong WW, Hirose T, Scheithauer BW, Schild SE, Gunderson LL. Malignant peripheral nerve sheath tumor: analysis of treatment outcome. *Int. J. Radiat. Oncol. Biol. Phys*. 1998;42(2):351-60.
- [25] Hruban RH, Shiu MH, Senie RT, Woodruff JM. Malignant peripheral nerve sheath tumors of the buttock and lower extremity. A study of 43 cases. *Cancer*. 1990;66(6):1253-65.

- [26] Meis JM, Enzinger FM, Martz KL, Neal JA. Malignant peripheral nerve sheath tumors (malignant schwannomas) in children. *Am. J. Surg. Pathol.* 1992;16(7):694-707.
- [27] King A, Listerick R, Charrow J, Piersall L, Gutmann DH. Optic pathway gliomas in neurofibromatosis type 1: the effect of presenting symptoms on outcome. *Am. J. Med. Genet A.* 2003;122(2):95-9.
- [28] Ferner RE, Gutmann DH. International consensus statement on malignant peripheral nerve sheath tumors in neurofibromatosis. *Cancer Res.* 2002;62(5):1573-7.
- [29] Tucker T, Wolkenstein P, Revuz J, Zeller J, Friedman JM. Association between benign and malignant peripheral nerve sheath tumors in NF1. *Neurology.* 2005;65(2):205-11.
- [30] Benz MR, Czernin J, Dry SM, Tap WD, Allen-Auerbach MS, Elashoff D, et al. Quantitative F18-fluorodeoxyglucose positron emission tomography accurately characterizes peripheral nerve sheath tumors as malignant or benign. *Cancer.* 116(2):451-8.
- [31] Ferner RE, Lucas JD, O'Doherty MJ, Hughes RA, Smith MA, Cronin BF, et al. Evaluation of (18)fluorodeoxyglucose positron emission tomography ((18)FDG PET) in the detection of malignant peripheral nerve sheath tumours arising from within plexiform neurofibromas in neurofibromatosis 1. *J. Neurol. Neurosurg. Psychiatry.* 2000;68(3):353-7.
- [32] Warbey VS, Ferner RE, Dunn JT, Calonje E, O'Doherty MJ. [18F]FDG PET/CT in the diagnosis of malignant peripheral nerve sheath tumours in neurofibromatosis type-1. *Eur. J. Nucl. Med. Mol. Imaging.* 2009;36(5):751-7.
- [33] Meany H, Dombi E, Reynolds J, Whatley M, Kurwa A, Tsokos M, et al. 18-fluorodeoxyglucose-positron emission tomography (FDG-PET) evaluation of nodular lesions in patients with neurofibromatosis type 1 and plexiform neurofibromas (PN) or malignant peripheral nerve sheath tumors (MPNST). *Pediatric blood & cancer.* 2012. Epub 2012/05/31.
- [34] Beert E, Brems H, Daniels B, De Wever I, Van Calenbergh F, Schoenaers J, et al. Atypical neurofibromas in neurofibromatosis type 1 are premalignant tumors. *Genes, chromosomes & cancer.* 2011;50(12):1021-32. Epub 2011/10/12.
- [35] Brenner W, Friedrich RE, Gawad KA, Hagel C, von Deimling A, de Wit M, et al. Prognostic relevance of FDG PET in patients with neurofibromatosis type-1 and malignant peripheral nerve sheath tumours. *Eur. J. Nucl. Med. Mol. Imaging.* 2006;33(4):428-32. Epub 2006/01/13.

Chapter 108

SURGICAL TREATMENT OF GIANT NEUROFIBROMAS

***Stamatis Sapountzis¹, Ji Hoon Kim²,
Abid Rashid¹ and Hung-Chi Chen¹***

¹Department of Plastic and Reconstructive Surgery, China Medical University Hospital,
Taichung, Taiwan

²Department of Plastic and Reconstructive Surgery, Seoul National University Bundang
Hospital, South Korea

ABSTRACT

Neurofibromatosis Type 1 (NF1) is an autosomal dominant systemic disease. Up to fifty percent of patients with NF1 are reported to have concomitant vascular abnormalities, incidence which increases the mortality especially in young patients. Among the complications of giant neurofibromas, spontaneous rupture and massive hemorrhage has been reported, leading to limb loss and death as well. The resection of larger neurofibromas is a challenging procedure because the risk of uncontrolled hemorrhage is much higher. In this chapter, we present a review of the medical literature of the methods that have been used to control intraoperative bleeding in large neurofibromas. We also discuss a novel surgical technique, which includes the ligation of the base of the giant neurofibroma tissue using a continuous loop-shaped suture. This method makes the operation field relatively bloodless and facilitates identification and ligation of the intralesional vessels. In addition, this surgical technique is less complicated and easier to perform and compared to others.

INTRODUCTION

Neurofibromatosis type 1 (NF1) is an autosomal dominant disorder with an incidence of approximately one in 3,000 live births [1]. The gene locus of NF1 is localized to chromosome 17. Patients with this disorder develop Schwann cell tumors, called neurofibromas, and skin abnormalities. The most characteristic features are the “café-au-lait” pigmented skin spots, Lisch nodules (iris hamartomas), and multiple neurofibromas, including cutaneous and plexiform neurofibromas [2].

In NF1 patients, transformation to a malignant peripheral nerve sheath tumor (MPNST) should be suspected if there is progressive enlargement and pain related to a neurofibroma. If this occurs, surgical excision is absolutely indicated, and also yields the best cosmetic results. Even though hemorrhage following trauma is a rare complication of neurofibromatosis, vascular abnormalities have been found in almost half of neurofibromatosis patients [3]. Macroscopically, these can be vascular stenosis, aneurysms, and arteriovenous fistulae. Microscopically, the most vulnerable are the small and the medium-sized vessels, in which the intima becomes thicker and the media thinner and fibrotic. Abnormal vascular structures are also observed in the neurofibromatous tissue. There are thin-walled, ecstatic blood vessels lying in a loose neural stroma, which replace the normal adipose tissue. For this reason, the risk for severe bleeding during the surgical excision is high, especially in giant neurofibromas [4].

This chapter presents a review in medical literature about the reconstruction options, as well as novel methods to control intraoperative bleeding. Using these methods, the authors present their personal experience in reconstructive surgery of giant neurofibromas and their novel techniques to reduce intraoperative bleeding.

SURGICAL TECHNIQUES

From 2004 to 2012, 16 patients (10 men and 6 women) with giant neurofibromas were operated on in our department. The mean age of the patient was 34 year old. The location of the giant neurofibroma was in the upper limb in 6 patients, in the head and neck region in 5 patients, in the lower limb in 3 patients, and in the chest wall and back in 2 patients. In two patients, one with neurofibroma in the face and the other with it in the left upper limb, the neurofibroma was associated with AVM malformation.

In 12 patients, the neurofibroma was excised and primary closure of the surgical wound was achieved. In the remaining patients, grafting of the surgical skin defect was required. In one patient with neurofibroma in the upper limb, the ulnar nerve was affected and a segment of the nerve was sacrificed. In order to reconstruct the nerve defect, a sural nerve graft was obtained from the right leg and was used to bridge the defect between the ulnar nerve proximally to the common digital nerve of the 4th and 5th digits distally. Another patient with neurofibroma of the upper limb required functional reconstruction of the hand after surgical excision. In this case the extensor carpi radialis longus and radiobrachialis tendon were transferred from the ulnar site to the first metacarpal head, in order to restore the normal function of the hand. In patients with head and neck region involvement, all of the tumors were excised successfully and primary closure of the surgical defect could be obtained. Increased attention should be given to the branches of the facial nerve because of this particular region. In all cases the operations were started with the identification of the marginal mandibular branch of the facial nerve, which crosses the facial artery. Then, working from distal to proximal, the dissection was advanced to the main trunk of the facial nerve as well as to the rest of the branches. In all patients, a total parotidectomy was performed as the tumor extended into the parotid gland (Figures 1–10).



Figure 1. A 53 year-old patient with neurofibroma of the face — front view.



Figure 2. Same patient — lateral view.



Figure 3. A 14 year-old boy with giant neurofibroma of the left upper limb associated with AVM malformation.

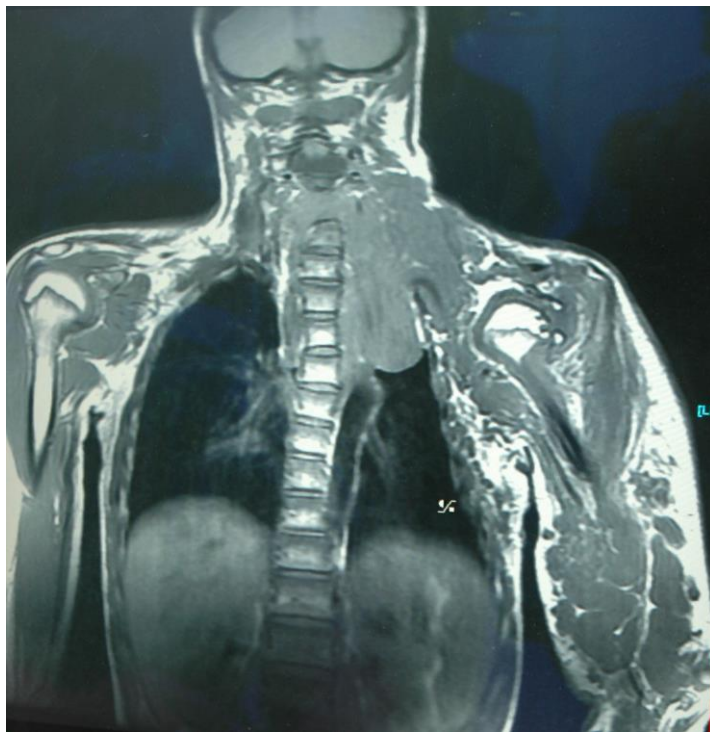


Figure 4. MRI revealed the intrathoracic extension of the neurofibroma.

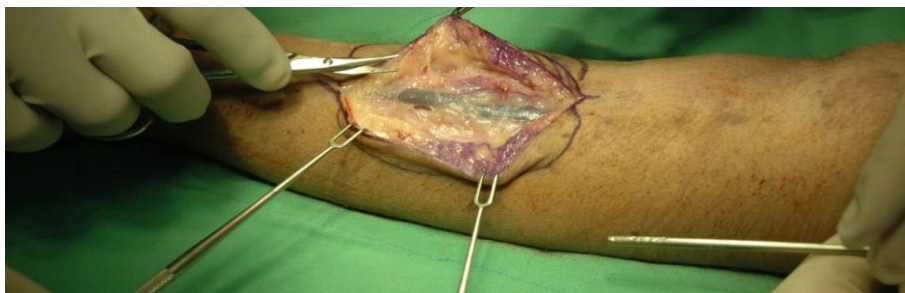


Figure 5. Intraoperative photo showing dilated abnormal superficial venous system of the upper limb.



Figure 6. Excision of the AVM of the upper limb and partial excision of the neurofibroma.



Figure 7. Postoperative photo after the first operation.



Figure 8. A 38 year-old woman with giant neurofibroma of the left upper limb.

A novel technique for prevention of the intraoperative bleeding was used by the authors [5] on a 46-year-old female patient with giant neurofibromas in the bilateral hip region. In the past, the patient had had three surgical excisions of subcutaneous nodules in different anatomical regions for cosmetic purposes, but never in the gluteal region because of the high tendency of intraoperative bleeding. The operation was performed under general anesthesia with the patient in prone position and the hips bent about 15 degrees. After performing the elliptical resection margin, the tumor base was tightened by use of a continuous loop shaped suture ligation. The thread was woven up and down in a loop shaped pattern, with a space of 2 cm between each loop, using a straight needle and prolene 2-0. After skin incision, the dissection was obliquely carried out toward the central and inferior sides of the mass using Metzenbaum scissors, while entering into large vascular sinuses was avoided. The tumor was resected in a wedge shape. The small vessels, which were visualized during the dissection, were

coagulated, and the small bleeding sites were also controlled with the electrocautery. After the skin closure, the loop-shaped ligation of the base was removed, and compressive dressing was performed with gauzes and elastic bandages. The size of the resected tumor was 26x31x21 cm and 24x30x19 cm on the left and right sides, respectively. Due to moderate blood loss, the patient was transfused with two units of packed red blood cells after the operation. The pressure dressing was maintained until the 7th post-operative day. The sciatic nerve was also evaluated during this period, with no evidence of nerve damage. The postoperative period was uneventful with no sign of infection, bleeding, or hematoma. The sutures were removed on the 13th postoperative day, and the patient was discharged the next day (Figure 11–14).



Figure 9. The excised tumor of the left upper limb.



Figure 10. Postoperative photo after total excision of the giant neurofibroma and coverage of the defect with split-thickness skin graft.

In medical literature there are few reports of successful surgical treatment of giant neurofibromas. The high incidence of postoperative complications, the necessity for preserving vital structures such as nerves in head and neck area, and at the same time the importance if the achievement good cosmetics result make the surgical treatment quite challenging. The surgical excision of giant neurofibromas and coverage of the defect using surrounding tissue remains the most popular technique. However, special considerations should be given to the control of intraoperative bleeding and the management of postoperative complications.

Ross [6] presented their experience with the management of a giant 49-kg neurofibroma of the lower extremity in a 37-year-old male with NF1 associated with right thigh pain, paresthesias, increasing edema, and accelerated growth of the mass. In this case the tumor was excised using electrocautery, and hemostasis was maintained with bipolar cautery, ligature and clips. Intraoperative EMG stimulation was used to avoid further nerve injury. An 18-cm rectangular fasciocutaneous flap along the medial proximal thigh was created to cover the defect. The estimated blood loss was 800ml, and the patient received two units of packed-red blood cells. During the following months, the patient was admitted to the hospital three times due to fever, signs of infection and poor wound healing. However, after the appropriate management no further problems occurred, and at the one year follow-up the wound had successfully healed with some residual lymphedema.

Liu et al. [7] reported a 13-year old girl with a large neurofibroma on the back extending into the midaxillary line and measuring 42x48x5cm³. The MRI revealed a diffuse neurofibroma with associated dysplastic blood vessels exhibiting irregular areas of tunica media and sinusoidal-like vascular channels (pseudohemangioma). An angiography demonstrated that the blood vessels of the tumor originated from the intercostal arteries and the transverse cervical artery. Interventional radiologists were unable to embolize such complex vasculature. The neurofibroma tissue was dissected and elevated above the deep muscle fascia, and a large split-thickness skin graft was harvested from the resected tumor tissue in order to cover the entire back.

Less frequently, large neurofibroma can be developed in genitalia. Sadeghi et al. [8] reported a case of giant neurofibroma of the labia major, associated with pain and bleeding from an ulcer on its lateral surface. The size of the tumor was 16 cm in length and 2 to 3 cm in thickness.

Special consideration should be given to head and neck giant neurofibromas. Cheng et al [9] reported a 45-year-old man with a giant tumor in the head, face, and neck. Facial asymmetry was found at birth and became more and more serious with aging. Subsequently, the tumor gradually extended to the head, face and neck and resulted in obvious deformity. This giant tumor was 50 cm in length and 44 cm in perimeter and led to a dislocated and deformed earlap and a longer external auditory channel. The tumor was resected successfully and eyelid anaplasty was performed.

However, apart from the surgical excision of the giant neurofibroma, a novel approach has been presented by Hamdoon et al. [10] using photodynamic therapy in a patient with giant solitary neurofibroma. Their report describes a 70-year old female with painful neck mass associated with slight shortness of breath and dysphagia. Examination revealed a large mass in the neck with no vascular compromise. A core biopsy was performed and histopathological examination revealed a disorganized array of peripheral nerve fascicles. The patient elected to receive photodynamic therapy as the primary intervention. The photosensitizing agent was mTHPC (0,15mg.KG), which was undertaken under general anaesthesia. Post-PDP follow-up

showed that the patient's pain, dysphagia and shortness of breath issues had improved. The disfigurement of the neck caused by the mass was no longer a problem. Three months post-PDT, MRI revealed a significant reduction in the neurofibroma size.

DISCUSSION

Neurofibromatosis type I (NF-1), or von Recklinghausen disease, is an autosomal dominant disorder affecting approximately 1 in 3000 individuals [4]. Typically the disease is diagnosed clinically during childhood, as the first sign in 80% of NF1 individuals is usually café-au-lait macules by 1 year of age. NF1 is a progressive and unpredictable disease that is associated with a variety of clinical outcomes and complications; prognosis is dependent on age, severity of the disease, and organ involvement by the growth of neurofibromas. The life expectancy of individuals with neurofibromatosis may be reduced by 10 to 15 years, most often as a result of malignancies such as peripheral nerve sheath tumors or soft-tissue sarcoma [11]. Furthermore, vascular disease appears to contribute to the excess mortality of young patients [12].

In addition, about the half of the patients also have vascular lesions, however the pathogenesis of these lesions are not well defined. The vascular lesions have been described in the entire arterial tree, but involvement of the renal arteries is most common. NF1 vasculopathy of the cerebrum, endocrine system, gastrointestinal tract, and heart have also been reported. Frequently multiple vessels are involved [13]. In 1944 Reubi [14] classified the vascular histology into intimal, aneurismal and fusocellular forms. A common finding between the types is spindle cell proliferation. Other theories support that the vascular lesions are due to nerve proliferation within the vessel wall or from compression or invasion by neural tumors. More often, the histologic feature is fibromuscular dysplasia with a predominance of intimal thickening.

Salyer and Salyer [15] suggested that intimal thickening in NF1 vasculopathy is the result of proliferation of Schwann cells within the arteries. This implies a pathogenic relationship between these lesions and the neurofibromas that characterize this disease. Riccardi has suggested that NF1 vasculopathy results from a dysplastic process, in which abnormal function of neurofibromin alters vascular histogenesis.

Macroscopically, these lesions can take a variety of forms including vascular stenosis, aneurysms, and arteriovenous fistulae. In histological examination the small and medium sized vessels are presented with thickening of the intima and a thin but fibrotic media. The vascular structure in neurofibromatous tissue also has histological alterations with thin walled ecstatic blood vessels lying in a loose neural stroma, which replace the normal adipose tissue [2].

Coexisting coagulopathies may increase the complication rates. These may be inherited or secondary to vascular anomalies causing platelet trapping and consumption of clotting factors. Hemostasis can be difficult to obtain, following injury to these lesions. Diathermy is of limited use as the tissue is very friable [16-18].

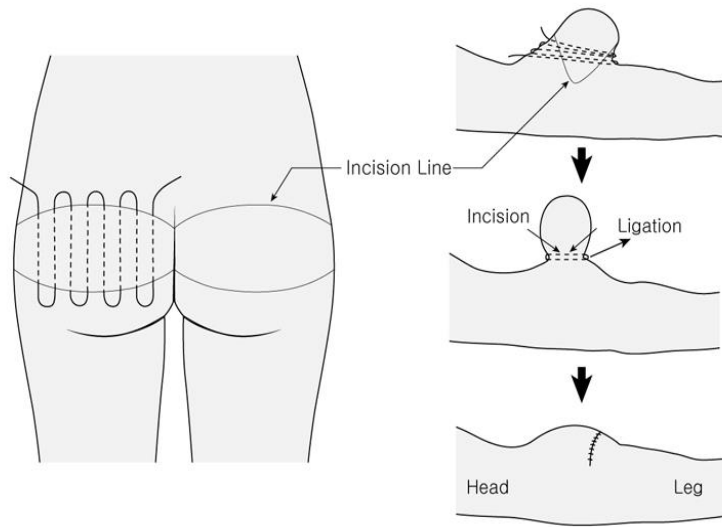


Figure 12. The schematic view of a continuous loop-shaped suture ligation, from ref. [5].



Figure 13. Postoperative view 2 days after surgery, from ref. [5].



Figure 14. Postoperative result 12 months after surgery, from ref. [5].

Preoperative angiography and superselective embolization have been used in elective excision of vascular neurofibromas. The results of such treatment without associated surgery have been disappointing, because the tumors tend to revascularize quickly. It has been suggested that embolization therapy by a skilled interventional radiologist should be considered prior to elective excision of neurofibromas to reduce intraoperative blood loss. The most effective way to control the bleeding seems to be the ligation of the neurofibromata's vascular pedicle under direct vision of the feeding vessels and the external compression [3].

CONCLUSION

This chapter presents the experiences of the authors in 16 cases of giant neurofibroma, as well as a literature review of the surgical methods that have been used in giant neurofibroma surgery. The preservation of significant structures such as the facial nerve and the control of the intraoperative bleeding in the cases that, associated with vascular malformations, are important steps when excision of these tumors is decided. In our cases, two patients with neurofibroma of the upper limb required further reconstruction after the tumor excision; one of them required nerve graft to bridge the nerve defect, and a second patient required tendon transfer in order for the normal function of the hand to be restored. In one case of giant neurofibroma of the buttock, a novel method was performed to obtain a bloodless field using a continuous loop-shaped suture in order to ligate the base of the tumor. In addition, we presented several other approaches, methods and tools facilitating successful tumor surgery.

REFERENCES

- [1] Choi KR, Yoon KM, Kim KS, Lee SY, Cho BH: Histopathological study of the vascular lesions in neurofibromatosis. *J Korean Soc Plast Reconstr Surg* 24: 45, 1997.
- [2] White N, Gwanmesia I, Akhtar N, Withey SJ: Severe haemorrhage in neurofibromatoma: a lesson. *Br J Plast Surg* 57: 456, 2004.
- [3] Tung TC, Chen YR, Chen KT, Chen CT, Bendor-Samuel R: Massive intratumor hemorrhage in facial plexiform neurofibroma. *Head Neck* 19: 158, 1997.
- [4] Choi HY, Lee HY, Kim YS, Kim SI, Kim JT: The treatment of neurofibromatosis involving trigeminal ganglion. *J Korean Soc Plast Reconstr Surg* 26: 263, 1999.
- [5] Stamatis Sapountzis, Ji Hoon Kim, Jin Sik Burm, Pham Minh, Chan Yeong Heo: A safe method for excision of a giant neurofibroma on both buttocks using a loop-shaped suture. *European Journal of Plastic Surgery*, 2012, Volume 35, Number 5, Pages 389-393.
- [6] Ross AL, Panthaki Z, Levi AD: Surgical management of a giant plexiform neurofibroma of the lower extremity. *World Neurosurg*. 2011 May-Jun;75(5-6):754-7.
- [7] Liu L, Fan J, Gan C, Tian J, Jiao H.: Total resection of giant plexiform neurofibroma of the entire back in one stage by autologous tumor skin graft. *J Plast Reconstr Aesthet Surg*. 2012 Feb;65(2):276-7. Epub 2011 Sep 8.
- [8] Sadeghi E, Johari HG, Seifi M, Rahmani M, Madani SH, Shemshadi M, Eskandari S.: Giant Neurofibroma of Labia Major. *J Low Genit Tract Dis*. 2012 Jan 5. [Epub ahead of print].

- [9] Cheng XB, Lei DL, Li YP, Tian L, Liu YP, Feng XH, Hu XG, Sun MY, Ma Q, Mao TQ, Liu BL, Zhao YM, Feng ZH, Xu LX, Zhang H, Zhang TC, Liu R, Shen L.: Surgical treatment of a giant neurofibroma. *J Craniofac Surg*. 2011 Nov;22(6):2244-6.
- [10] Hamdoon Z, Jerjes W, Al-Delayme R, Hopper C.: Solitary giant neurofibroma of the neck subjected to photodynamic therapy: case study. *Head Neck Oncol*. 2012 Jun 6;4(1):30. [Epub ahead of print].
- [11] Viskochil DH. Neurofibromatosis Type 1. *Management of Genetic Syndromes*. 2001; 229-251. New York: Wiley-Liss, Inc.
- [12] Rasmussen S, Yang QH, Friedman JM. Mortality associated with neurofibromatosis 1 in the United States from 1983 to 1995: an analysis using data from death certificates. *Am J Hum Genet* 1999; 65: A49.
- [13] Halpern M, Currarino G. Vascular lesions causing hypertension in neurofibromatosis. *N Engl J Med* 1965; 273 (5): 248–252.
- [14] Reubi F. *Les vaisseaux et les glandes endocrines dans la neurofibromatose: le syndrome sympathicotonique dans le maladie de Recklinghausen*. [Vessels and endocrine glands in neurofibromatosis: sympathicotonic syndrome in Recklinghausen's disease]. *Schweiz Z Pathol Bacteriol* 1944;7:168-236.
- [15] Salyer WR, Salyer DC. The vascular lesions of neurofibromatosis. *Angiology* 1974; 25: 510–519.
- [16] Kitao T, Miyabo S, Hittori K. Hemophillia associated with von Recklinghausens disease. *South Med J* 1976;69:16—39.
- [17] Farah GR, Awidi AS. Massive bleeding in neurofibromatosis with congenital hypofibrinogenaemia. *Eur J Surg Oncol* 1985; 11:57—60.
- [18] Francis DMA, Mackie W. Life threatening haemorrhage in patients with neurofibromatosis. *Aust NZ J Surg* 1987;57: 679—82.

Chapter 109

NEUROFIBROMATOSIS TYPE 1-ASSOCIATED NERVOUS SYSTEM TUMORS AND CURRENT TREATMENT STRATEGIES

Nilika Shah Singhal¹ and Sabine Mueller^{1,2}

¹Department of Neurology, Division of Child Neurology,
University of California San Francisco, San Francisco, CA, US

²Department of Pediatrics and Neurosurgery, University of California San Francisco,
San Francisco, CA, US

ABSTRACT

Patients with Neurofibromatosis 1 (NF 1) suffer from a myriad of medical problems including benign as well as malignant tumors. Malignancy is the most common cause of death in affected individuals, and reduces life expectancy by 10-15 years. Amongst other tumors, patients with NF1 carry an elevated risk of specific central nervous system (CNS) tumors. The most common NF1-associated CNS tumors are low-grade gliomas (LGG) such as optic pathway gliomas, hypothalamic gliomas, and other parenchymal gliomas located in the brainstem, cerebellar peduncles, globus pallidus, and midbrain. In addition to CNS tumors, patients with NF1 also develop peripheral nervous system (PNS) tumors such as neurofibromas and schwannomas. These tumors arise most commonly from major peripheral nerves such as the radial and ulnar nerve. These are benign tumors however can cause significant disfigurement, pain and depending on the location specific neurological complications. Malignant transformation of these tumors leads to malignant peripheral nerve sheath tumors (MPNST). These tumors occur in less than 5% of children with NF1, but are the leading cause of mortality in adult patients with NF1.

Significant advances in our understanding of LGG have been achieved over the last several years. Several genetic aberrations, which activate the MAPK pathway have been identified in the majority of LGGs, most notably the BRAF-KIA1549 fusion protein and the activating BRAFV600E mutation. Specific inhibitors of MEK1/2, the immediate downstream target of BRAF, are currently being tested in the clinic for children with LGGs. Treatment for symptomatic plexiform neurofibromas remains a clinical challenge for most pediatric neuro-oncologists. Recent studies have shown promise to use pegylated (PEG)-interferon-alpha-2b. Other treatment strategies also target the MAPK pathway for

these tumors. Treatment of MPNST remains under investigation and outcome remains poor. Most patients are currently treated with a combination of surgery, radiation and chemotherapy.

In this chapter, we will outline NF1 associated CNS and PNS tumors and discuss current treatment approaches.

INTRODUCTION

Neurofibromatosis 1 (NF1) is a genetic syndrome due to loss-of-function of the NF1 gene that encodes neurofibromin [1]. NF1 is a tumor suppressor gene that is a negative regulator of the RAS signaling cascade [2]. Loss of neurofibromin results in increased activity of ras, which leads to tumorigenesis [3]. The incidence of NF1 is estimated to be about 1 in 3000 [4]. Patients with NF1 suffer from a myriad of medical problems, including benign as well as malignant tumors [5].

Malignancy is the most common cause of death in affected individuals and reduces life expectancy by 10-15 years [6]. Amongst other tumors, patients with NF1 carry an elevated risk of specific central nervous system (CNS) tumors. The most common NF1-associated CNS tumors are low-grade gliomas (LGG), such as optic pathway gliomas, hypothalamic gliomas, and other parenchymal gliomas located in the brainstem, cerebellar peduncles, globus pallidus, and midbrain [7, 8, 9]. In addition to CNS tumors, patients with NF1 also develop peripheral nervous system (PNS) tumors, such as neurofibromas and schwannomas [6]. These tumors arise most commonly from major peripheral nerves such as the radial and ulnar nerve. While these tumors are typically low-grade, they can cause significant disfigurement, pain and other neurological complications. Further, malignant transformation of these tumors leads to malignant peripheral nerve sheath tumors (MPNST). MPNSTs occur in less than 5% of children with NF1, but are the leading cause of mortality in adult patients with NF1 [6].

Significant advances in our understanding of these tumors have been achieved in recent years. Several genetic aberrations, which upregulate the RAS pathway and activate VEGF, play a role in the tumorigenesis of these tumors. In addition, mutations that activate the MAPK pathway have been identified in pediatric LGGs, most notably the BRAF-KIAA1549 fusion protein and the activating BRAF^{V600E} mutation [10, 11, 12, 13, 14], however these might be less common in patients with NF1 [12, 15, 16, 17]. These identified mutations might play a role as biomarkers, and identification of these pathways has led to new therapeutic developments that are currently being tested in clinical trials. For instance, specific inhibitors of MEK1/2, the immediate downstream target of BRAF, are currently being tested in the clinic for children with LGGs via the Pediatric Brain Tumor Consortium (PBTC). Treatment for symptomatic plexiform neurofibromas remains a clinical challenge for pediatric neuro-oncologists [18], but recent studies have shown promise in the use of PEG-interferon-alpha-2b [19, 20]. To date we have not yet identified an effective treatment strategy for patients with MPNST, and patients are often treated on clinical trials with a combination of surgery, radiation and chemotherapy.

In this chapter, we outline NF1 associated CNS and PNS tumors, and discuss current treatment approaches.

CENTRAL NERVOUS SYSTEM TUMORS IN CHILDREN WITH NF1

The most common NF1-associated CNS tumors are optic pathway gliomas (OPG) and hypothalamic gliomas, as well as other parenchymal gliomas located in the brainstem, cerebellar peduncles, globus pallidus, and midbrain [7, 8, 9]. OPGs are the most common CNS tumors in patients with NF1, occurring in up to 15% of children with NF1 [21, 22, 23]. In patients with symptomatic OPGs, the risk of developing other glial tumors in the CNS is nine times more frequent than in patients without symptomatic OPGs [24].



Figure 1. Optic Glioma. Coronal T2-weighted magnetic resonance imaging (MRI) scan demonstrating right optic glioma. Note enlargement of the optic nerve.

Low Grade Glioma: Clinical Presentation

OPGs can occur anywhere along the optic tract. In most patients, these tumors are asymptomatic, but about one third of patients will develop symptoms, including vision loss or proptosis. Other clinical findings include abnormal color vision, optic atrophy and afferent pupillary defect. In addition, precocious puberty can be seen in OPGs involving the optic

chiasm and the hypothalamus [25, 26]. In general, children become symptomatic of these tumors within the first decade of life [27]. About 50% of children with NF1 and OPG will develop vision loss. Consequently, close clinical observation is typically done for initial management [23]. A recent retrospective, multi-center study evaluated visual outcomes in patients following chemotherapy for NF1-associated OPG. Mean age of diagnosis of OPG was 2.66 years, age of first chemotherapy dose median was 4.04 years with a median range of diagnosis to treatment being 133 days. Outcomes were assessed at the completion of therapy: about one third of patients had improved visual acuity and 40% had stable visual acuity. While most subjects experienced stabilization or improvement in vision after therapy, visual acuity worsened in 28% of subjects after treatment. The most consistent prognostic factor for poor visual outcome was tumor involvement of the optic tracts or radiations ($P = .02$ per subject) and optic pallor at the start of treatment ($P = .005$ per eye). Age was another prognostic factor for poor visual outcome, with subjects younger than 2 years or older than 5 years more likely to have a decline in visual acuity. Further, none of the children younger than 2 years of age demonstrated any improvement in visual acuity with treatment. Of note, there was a poor correlation between radiographic outcome and visual acuity outcome; as many as 25% of patients with complete response or partial response by MRI had worsening of visual acuity, and 29% with stable disease or minor response showed worsening of visual acuity, thus calling into question the utility of MRI as adequate measure of visual outcome [23].

Other LGGs that occur in patients with NF1, though less commonly than OPGs, include astrocytomas and ependymomas, and are located within the cerebellum, brainstem, deep grey nuclei, and spinal cord [28]. Like OPGs, these low-grade gliomas tend to have a better prognosis in NF-1 patients as compared to non-NF1 patients [29]. This is true regardless of location including brainstem tumors [30, 31]. NF1 patients with brainstem gliomas tend to have more indolent courses. A study of 17 patients with NF1 and brainstem tumors indicated that the medulla was most commonly involved structure, in 82% of these patients, in contrast to the pontine location as more commonly occurs in non-NF1 patients [30]. Median follow-up was 52 months; 15 of 17 patients remained alive, and 14 of those did not require adjuvant chemotherapy [30]. Another study described 23 NF1 patients with brainstem tumors with a median follow-up period of 67 months for 17 of 23 patients who were untreated and 102 months for the six of 23 patients who received therapy; only one previously untreated patient experienced radiographic and clinical progression, and all but one remain alive [32].

Low Grade Glioma: Imaging Characteristics

The characteristics of optic pathway glioma are often apparent on head CT scans, however; MRI is the preferred method and current clinical standard for diagnosis and follow up imaging. CT characteristics include enlargement of the optic nerve(s), chiasm or tract. Further, there can be changes to the bony structures of the optic canal and sphenoid wing [33, 34]. MRI T2-weighted imaging more clearly delineates anatomic changes including enlargement of the optic nerve(s) (Figure 1), chiasm (Figure 2) or tract. In addition, both homogenous and heterogenous contrast enhancement can be seen [33, 35, 36]. Other characteristics include a cystic component to the tumor, as well as mass effect on adjacent structures [37]. Other parenchymal lesions, such as gliomas involving the thalami and basal ganglia, appear as hyperintense T2 lesions that may or may not be enhancing [35, 36].

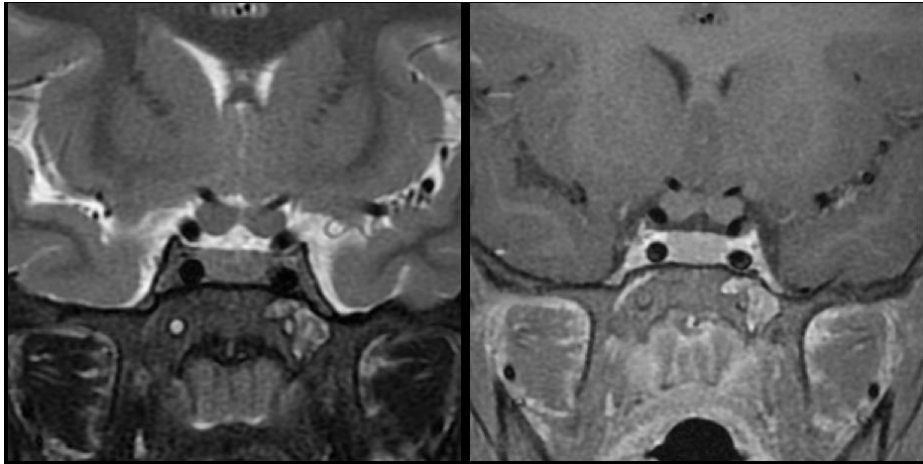


Figure 2. Low-grade Glioma. Coronal T2-weighted MRI (left) and T1-weighted MRI (right) demonstrating low-grade glioma of optic chiasm. Note enlargement of the optic chiasm.

Low Grade Glioma: Prognosis

OPG tumors are common and the natural history demonstrates often follows a benign course and some tumors may even regress after initial diagnosis [38]. The overall 5-year survival rate is 90% at 5 years [37]. However, these tumors may lead to symptoms, typically visual loss and hypothalamic dysfunction such as precocious puberty. The highest-risk period for the development of symptomatic OPGs in NF1 patients is during the first 6 years of life [39], and further, ages <2 years or >5 years are poor prognostic factors [23]. Screening MRIs of asymptomatic children remains controversial since it has not been proven effective in reducing complications related to OPGs [39]. Further, a recent study demonstrated a poor correlation between radiographic outcome and visual acuity outcome [23]. Thus, functional outcomes are of highest importance, and yearly ophthalmological evaluation is indicated in younger children, and older children should undergo sporadic ophthalmological evaluations [39]. OPGs that present in older patients are more likely to progress compared to younger patients under age 10 years [39]. Other features associated with shorter survival aside from diagnosis in adulthood include extra-optic location and symptomatic tumors [37]. Of note, symptomatic and/or progressive tumors are often found to be of more aggressive histologic subtype [37]. Brainstem glioma and optic pathway glioma tend to be less aggressive in NF1 patients than in non-NF1 patients [30, 40]; in a study of 23 patients with OPG, only 1 patient with NF1 and OPG died due to disease progression, whereas 7 of 39 non-NF1 patients died ($p=.045$) [40].

Low Grade Glioma: Treatment Modalities

1. Role of Surgery

For lesions that are easily resectable, such as LGGs located within the posterior fossa, surgery is standard of care and often curative. Conservative treatment such as serial

neuroimaging and serial ophthalmological evaluations, however, is often preferred for most asymptomatic LGGs for which complete surgical resection cannot be achieved without significant morbidity such as OPG or midline LGG. In these cases, surgery is reserved for those with radiographic progression or associated severe symptoms [7], e.g., significant or complete visual loss, severe proptosis or strabismus, or associated hydrocephalus [7]. Cerebrospinal fluid shunts are used when patients are symptomatic from hydrocephalus [37].

2. Radiation Therapy

Most centers avoid radiation therapy as a first-line treatment in young children with NF1, and only patients who have failed initial chemotherapy options are considered for radiation therapy. Various doses of radiation therapy have been used in the treatment of OPGs, and while tumor control is attained in up to 80-90% of patients with NF1 [41, 42], this type of treatment is associated with significant delayed morbidity, including neuro-cognitive decline, radiation-induced vasculopathies and endocrinopathies, such as deficiency in growth hormone [43]. Currently the rate of malignant transformation in pediatric LGG is not well characterized since biopsies at recurrence are rarely performed for these patients. While the risk of spontaneous transformation of LGG might be rare in pediatric patients [44], it has been noted in patients treated with radiation therapy [41, 45]. A study examining 82 tumors in 18 NF1 patients at a median age of 25 years (range, 4.3-64 years) showed that the most common reason to initiate radiation therapy was radiographic evidence of growth. 67% of patients received radiation therapy. Of all NF1 patients, the overall survival at 5 years was 94%. Five patients with NF1 were treated for LGG, which accounted for 11% of the tumors. Two of these patients developed malignant transformation of tumors, and both of these patients subsequently died [41]. Given these significant morbidities, radiation therapy is often avoided in children with NF1, unless they fail other treatment modalities.

3. Chemotherapy

Conventional chemotherapy is used effectively, particularly in young patients under age 5 years old in whom radiation therapy poses significant morbidity [27, 46]. Carboplatin and vincristine are the two agents most commonly used together, resulting in stabilization or reduction of tumor, with a beneficial 3-year progression-free survival (PFS) [19]. A recent study involving children with progressive or residual LGGs were randomly assigned to receive either a combination of carboplatin and vincristine, or a combination of thioguanine, procarbazine, lomustine, and vincristine (TPCV). Among all children, the 5-year event-free survival and overall survival were both higher for those receiving TPCV. This study also included a nonrandomized arm for patients with NF1, the results of which are to be reported separately [47]. Given the loss of a known tumor suppressor gene as occurs in NF1, there is concern that NF1 patients are at higher risk for transformation of residual or progressive tumor to a more aggressive phenotype after alkylating chemotherapy, as this has been reported in two patients occurring in the absence of radiation therapy and within 4 months of receiving chemotherapy [48]. Other agents that are currently being used for patients with LGG include vinblastine, the combination of bevacizumab and irinotecan, as well as temozolomide [49, 50, 51]. Weekly treatment with vinblastine for one year resulted in a 5-year PFS in a phase II trial of $42.3\% \pm 7.2\%$ and was in general well-tolerated, thus seems to be a good alternative to radiation therapy in pediatric patients with LGG who have failed other chemotherapy [50]. In children with multiple recurrent LGG treated with the combination of bevacizumab and

irinotecan every 2 weeks, 7 of 10 had an objective neuroradio-graphic response after two cycles of treatment, and 7 of 10 demonstrated clinical improvements, the majority with a lasting response, and this combination is reasonably well-tolerated. Of these patients, 6 of 10 remained on treatment for up to 22 months after treatment initiation [51].

4. Targeted Therapy

PI3K/mTOR Pathway: TOR is an evolutionarily conserved serine/threonine protein kinase; it plays an important role in cell growth and proliferation across species [52]. Mammalian TOR (mTOR) controls a variety of cellular processes involving regulation of growth factors and energy [52]. mTOR signals downstream targets within the PI3K/AKT pathway. This pathway is known to be dysregulated in a wide spectrum of human cancers through loss/mutation of PTEN as a negative regulator, PI3K mutation/amplification, AKT/PKB overexpression, and modulation of TSC1/TSC2 tumor suppressors. In addition, activation of the PI3K/AKT/mTOR pathway is frequently a characteristic of worsening prognosis through increased aggressiveness, resistance to treatment and progression. As such, mTOR inhibition might be an effective treatment strategy for LGG [53]. NF1-related low-grade astrocytomas exhibit increased levels of mTOR pathway activation, demonstrated by western blot analysis assessing phospho-S6, a marker reflecting mTOR activation [54]. A current trial is investigating the potential benefit of mTOR inhibitor everolimus in children with chemotherapy-refractory progressive or recurrent LGGs, including children with NF1. Results of this trial are pending. (www.poeticphase1.org).

RAS-MAPK Pathway

Neurofibromin, encoded by the NF1 gene, is a GTPase-activating protein (GAP). It acts as a tumor suppressor by negatively regulating RAS pathway output by the conversion of active RAS-GTP to inactive RAS-GPD [55]. Thus lack of NF1 results in increased RAS activity leading to dysregulated cell growth. There are three important isoforms of RAS: K-RAS, N-RAS, and H-RAS. It has been demonstrated that activation of the K-RAS isoform leads to a proliferative advantage seen in astrocytes [56]. Further, K-RAS has been shown to modulate the RAF/mitogen-activated protein kinase (MAPK)/extracellular signal-regulated kinase (ERK) pathway; typically, increased RAS activity results in increased signal transduction via this pathway. One avenue for possible therapeutics for LGGs involves targeting the MAPK pathway. Antibodies that block the MAPK-ERK pathway in the Nf1:p53 mouse tumor model cell lines *in vivo* have been demonstrated to block EGFR-stimulated tumor cell line growth [57]. Recent advances investigating the underlying oncogenic events in pediatric LGG have led to the discovery of activating mutations in BRAF, a member of the RAF family of serine/threonine protein kinases. The RAS-MAPK signaling pathway utilizes a series of kinases to relay signals from the cell membrane to the nucleus, playing important roles in cell proliferation, differentiation, cell-cycle arrest, and apoptosis [58]. In pediatric LGG, particularly in pilocytic astrocytomas, a copy number gain of 2 Mb at the chromosomal region 7q34 has been identified. This copy number results in a fusion of the genes KIAA1549 and BRAF, and leads to constitutive activation of BRAF due to loss of the Ras-binding domain of BRAF [13, 16, 59]. BRAF activation leads to phosphorylation of MEK1/2, which in turn phosphorylates ERK. The various KIAA1549-BRAF fusion genes are thought to be functionally similar, all possessing the BRAF kinase domain, while the NH2 terminus is

replaced by KIAA1549, implying that the fusion protein is constitutively activated, leading to high levels of phosphorylated MEK and ERK [60].

In a study of 106 LGG (grade I-II) from primarily pediatric patients, including pilocytic astrocytomas, low-grade glioneuronal/neuroepithelial tumors, pleomorphic xanthoastrocytomas, diffuse astrocytomas, pilomyxoid astrocytoma, and unclassifiable low-grade gliomas, KIAA1549-BRAF fusions were identified by sequencing in nearly 50% of tumors overall and were found in the majority of pilocytic astrocytomas [61]. Importantly, no KIAA1549: BRAF fusions were seen as higher-grade tumors (five high-grade astrocytomas, four medulloblastomas, two ependymomas, and one dysembryoplastic neuroectodermal tumor were tested) [61]. Of note, in this cohort of patients, 5% had a clinical diagnosis of neurofibromatosis, and all of these lacked BRAF alterations [61]. Identification of KIAA1549 - BRAF fusion may be an important molecular marker to help in distinguishing pilocytic astrocytomas and other tumors such as low-grade diffuse astrocytomas and mixed oligoastrocytomas grade II, as these fusion products are detected in 80% of pilocytic astrocytomas but only infrequently in these other tumors [60]. However, classifying pediatric LGG based on specific molecular aberrations is still subject of debate and additional research is required.

BRAF rearrangements have been shown to be an independent favorable prognostic factor in pediatric LGG [62]. In a study of 70 pediatric low-grade astrocytomas compared with 76 control tumors, five-year PFS was 61% for BRAF-KIAA1549 fusion-positive compared with 18% for fusion-negative patients ($P=0.0004$) [62]. Further, it has been demonstrated that pediatric gliomas demonstrate activating point mutations of BRAF, the BRAF^{V600E} mutation [12]. The BRAF^{V600E} mutation is found in approximately 60% of pleomorphic xanthoastrocytomas and 10–15% of pediatric astrocytomas, but are also seen in pediatric gangliogliomas [12, 15, 16, 17]. In addition to BRAF mutations, other targets along the MAPK signaling pathway include K-RAS, PTPN11, MEK1/2, and H-RAS. In some pilocytic astrocytomas, the SRGAP3-RAF1 gene fusion, as well as activating mutations in KRAS, have been identified [13, 59]. Additional studies are needed to inquire how common these mutations are in NF1-associated tumors.

MEK inhibitors are currently introduced in clinical trials for other types of tumors, such as melanoma and lung cancer. Given the upregulation of the MAPK pathway in pediatric LGG, an ongoing study from the Pediatric Brain Tumor consortium is currently testing a MEK inhibitor for the treatment of pediatric progressive LGG including children with NF1 (www.pbtc.org).

PERIPHERAL NERVOUS SYSTEM TUMORS IN CHILDREN WITH NF1

In addition to CNS tumors, patients with NF1 also develop peripheral nervous system (PNS) tumors such as neurofibromas and schwannomas [6]. These tumors arise most commonly from major peripheral nerves such as the radial and ulnar nerve. While benign, these tumors can cause significant disfiguration, pain and other neurological complications. Malignant transformation to malignant peripheral nerve sheath tumors (MPNST) can occur. These occur in less than 5% of children with NF1, but are the leading cause of mortality in adult patients with NF1 [6].

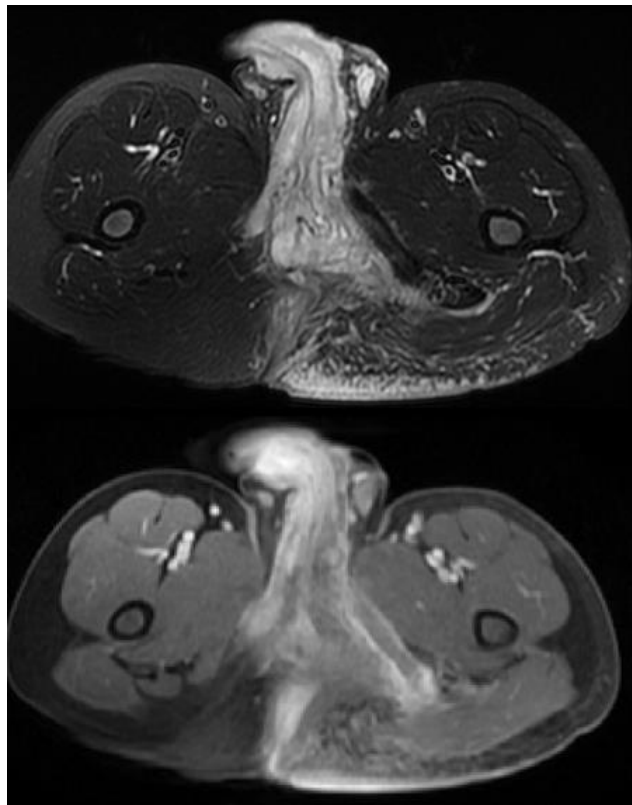


Figure 3. Plexiform Neurofibroma. Axial T2-weighted MRI (top) and T1-weighted post-contrast MRI (bottom) demonstrating large, diffuse, infiltrating plexiform neurofibroma involving the left sacral plexus extending into the left sacral neural foramina.

Plexiform Neurofibromas: Clinical Presentation

Neurofibromas are comprised of a mixture of fibroblasts, pericytes and proliferated Schwann cells [63]. Schwann cells are thought to be the primary pathogenic cell in neurofibromas [64]. These tumors originate from within the nerve, expanding that portion of nerve, and are well-circumscribed. Neurofibromas in NF1 patients can be classified as cutaneous neurofibromas, subcutaneous neurofibromas and plexiform neurofibromas (PNs).

Cutaneous and subcutaneous neurofibromas arise at the nerve terminus or just beneath the skin. The size ranges from only millimeters to large enough to be disfiguring. Localized cutaneous neurofibromas are the most common clinicopathologic subtype. These can be colored and are typically soft. The subcutaneous neurofibromas are sometimes firm, tender, palpable nodules under the skin. Both can vary in number from just a few to innumerable in any given patient and typically appear in late childhood or early adolescence [65]. PNs occur in over 50% of people with NF1 [66]. These can be nodular or diffuse. Nodular PNs arise from nerves or organs beneath the skin, commonly clustering around proximal nerve roots with the potential to extend through the neural foramina and compress the spinal cord. Diffuse PNs involve multiple nerves, often expanding into surrounding tissues and organs. The overlying

skin is often thickened and hyperpigmented. Diffuse PNs tend to enlarge with age and can become disfiguring. As such, they may limit range of motion or function of involved limbs or organs [27].

Plexiform Neurofibromas: Imaging Characteristics

Neuroimaging can determine the precise nerves involved or the extent of involvement of peripheral nerve sheath tumors (Figure 3). Certain subsets of these tumors have a typical appearance on imaging. For instance, nodular PN take on a classic “dumbbell” appearance on neuroimaging [27]. Imaging can sometimes be helpful for determining the transformation from benign to malignant peripheral nerve sheath tumors, as characteristics such as rapid growth compared to prior radiographic appearance, avid enhancement and invasion of surrounding structures correlate with malignancy. However, these qualitative measures may be difficult to assess at times, thus necessitating biopsy [7, 67].

Plexiform Neurofibromas: Prognosis

Treatment of plexiform neurofibromas remains challenging. The natural history of these tumors is not well understood. One study showed that PN can remain stable in older individuals, but have a higher tendency for growth in children < 21 years of age. These tumors tend to grow and enlarge over time, with no proven medical therapies to reduce or prevent their growth [27]. A recent study followed 34 NF1 patients with a total of 44 tumors with longitudinal MRI scans over 16 years. Most of the PN in this study were small (<10 cm²), but notably the largest tumors occurred in children younger than 13 years of age. Further, neurofibromas in younger patients grew faster than neurofibromas from older patients; in children younger than 13 years of age, growth of 2 cm²/ year or more in 30% of tumors was observed. This study followed 18 patients with NF1 younger than 10 years of age; only 33% of the 23 tumors were symptomatic. These findings may suggest that MRI screening for PN in younger NF1 patients may be valuable. Also in this study, superficial plexiform neurofibromas were observed to grow more rapidly than deep PNs ($p=.034$), implying that tumor location may affect growth. In patients with multiple PNs, the growth of each tumor often differed; this underscores the importance of following each tumor within an individual separately, as rate of growth of one tumor cannot be used to represent the rate of growth of all tumors within any given patient [66]. Of all of the patients followed, one adult man with a superficial and invasive PN developed pain and was found to have substantial tumor growth; this tumor was biopsied and found to be an MPNST [66].

Plexiform Neurofibromas: Treatment Modalities

1. Role of Surgery

Surgical excision has been the mainstay of treatment of PNS tumors such as neurofibromas, particularly for those tumors causing severe pain or weakness, thus indicating compression of

nerves or the spinal cord itself, or involvement of surrounding tissues. However, the associated morbidities of nerve injury, hemorrhage and tumor recurrence are significant, and thereby limit surgical management only to those patients with intolerable symptoms [27]. As they infiltrate surrounding tissues, complete surgical resection is often not possible [27] and tumor regrowth is common. In a retrospective review of 121 NF1 patients with 168 tumors over 20 years at a pediatric referral center who underwent a total of 302 surgical procedures, 74 of the 168 tumors (44%) all progressed after the first procedure [68]. A subgroup analysis revealed that 60.2% of children age 10 years or less had tumor progression after the first procedure, compared with 31.2% of children older than 10 years of age ($p=.0004$). Of the 25 cases of complete tumor resection, only 5 (20%) progressed, compared with 39.5% progression among those tumors with a near-total resection. By contrast, 74 tumors had a subtotal resection, defined as 50-90% resection, with 33 (44.6%) of these tumors progressing; 21 of 31 (67.7%) tumors with a <50% resection progressed after the first procedure ($p<.0001$ log rank). Of note, for those tumors that did progress, the median time to progression was increased the more extensive the resection (<2 years time to progression for those with <50% resection compared with >10 year time to progression for those tumors with near-total resection) [68].

2. Radiation Therapy

Certain patients are not ideal candidates for surgery, particularly those with advanced age, tumor location precluding surgical resection or other comorbidities. Stereotactic radiosurgery is a potential treatment option for benign lesions, though less data exist in support of this as compared with radiotherapy for malignant tumors. Further, secondary neoplasia and radiation-induced changes, such as myelopathy, remain a concern [69]. Despite these potential morbidities, image-guided radiosurgery is being used to treat benign tumors, including neurofibromas, particularly in those patients for whom surgery is not an option. A study reported greater than 95% long-term control as long as 48 months, as demonstrated by imaging, of 46 benign spinal cord tumors including neurofibromas, schwannomas and meningiomas treated with radiosurgery [69]. In this study, nearly 90% of the neurofibromas occurred in confirmed NF1 patients. Of all of the tumors in this study, the neurofibromas showed the poorest clinical response, with only 18% showing reduction in tumor volume [69]. This may indicate that neurofibromas respond more poorly to radiation therapy than do other tumor histologies.

3. Chemotherapy

Chemotherapy is generally not effective, and various clinical trials have not led to a standard medical treatment [18]. Early studies were performed using the antihistamine agent ketotifen fumarate, but variability in entry criteria as well as subjective endpoints make interpretation of efficacy of this medication not possible [18]. A prospective trial for treatment was a randomized phase II trial for PN, which examined cis-retinoic acid or interferon. This trial enrolled both children and adults, though the mean age of participants was 11 years. Participants were enrolled if they had progressive PN at time of entry, and tumors were monitored with serial imaging. While 96% of patients treated with interferon were stable at 18 months, no patient had a significant radiographic response; 14% of patients described symptomatic improvement, and 8% demonstrated minor tumor shrinkage on imaging [18]. A phase I trial using thalidomide, an anti-angiogenic agent, in patients with PN was recently completed [18, 70].

4. Targeted Therapy

a) mTOR Pathway

Blocking mTOR may slow or stop tumor growth in patients with NF1. An ongoing trial is examining the role of sirolimus, an mTOR inhibitor, to treat PN in patients with NF1 (study number NCT00652990).

Interferon Treatment

Interferons have been shown to have antitumor effects *in vitro* by both antiproliferative and antiangiogenic properties. These cytokines are secreted proteins that defend the host against infectious agents, such as viruses and parasites, but also affect immune functioning, cell proliferation, and cell differentiation. Treatment with interferons have been proven effective *in vivo* in a variety of human diseases, including hairy cell leukemia, Kaposi's sarcoma, non-Hodgkin's lymphoma, and chronic myelogenous leukemia [19]. A recent study looked at the role of interferons in neurofibromas. Thirty patients were enrolled, with a median age of 9.3 years, with progressive, symptomatic, unresectable, or life-threatening PN. Twelve of these patients received their recommended phase II dose of pegylated interferon- α -2b (PI) of 1 μ g/kg/wk. The PI was administered weekly and continued for up to 2 years. This study demonstrated that children with PN showed disease stabilization; further, 29% of patients who underwent volumetric analysis had a 15-22% decrease in volume with interferon [20].

b) Anti-angiogenic Treatment

Angiogenesis plays a role in tumor formation in NF1. RAS mutations upregulate vascular endothelial growth factor (VEGF) expression [71]. In addition, fibroblast growth factor (FGF) is highly expressed in neurofibromas [72]. Thus, anti-angiogenic treatment is another potential treatment intervention. A current clinical trial is examining local injection of ranibizumab, a VEGF inhibitor, into neurofibromas compared with control (saline) on tumor volume (study number NCT00657202). Inhibition of survival pathways, such as blocking the epidermal growth factor receptor (EGFR) signaling cascade, has been shown to be effective in *in vitro* models of neurofibromas. Growth of neurofibroma cells *in vitro* in the Nf1:p53 mouse tumor model could be blocked by an EGFR antagonist AG1478, which is a cell-permeable member of the tyrphostin family that binds tightly to the catalytic domain of the EGFR, inhibiting its tyrosine-specific protein kinase activity [56]. Further, NF1+/- p53 +/- mice that develop sarcomas have shown improved survival when EGFR expression was reduced genetically [73].

c) Tyrosine Kinase Inhibitor Treatment

Imatinib mesylate is a receptor tyrosine kinase inhibitor that targets platelet-derived growth factor receptor (PDGFR) expressed on Schwann cells and blood vessels. Imatinib significantly reduced Schwann cell viability in those cells derived from human PN *in vitro* after treatment for 28 days [74]. A pilot study is currently ongoing to examine the efficacy of imatinib mesylate in NF1 patients with plexiform neurofibromas (study number NCT01140360). A phase I trial of the raf kinase and receptor tyrosine kinase inhibitor sorafenib in children and young adults with NF1 and inoperable neurofibroma is completed with results pending (study number NCT00727233). Further, a phase II trial of pirfenidone, an anti-inflammatory and anti-fibrotic agent, in children is completed with results pending (study number NCT00076102).

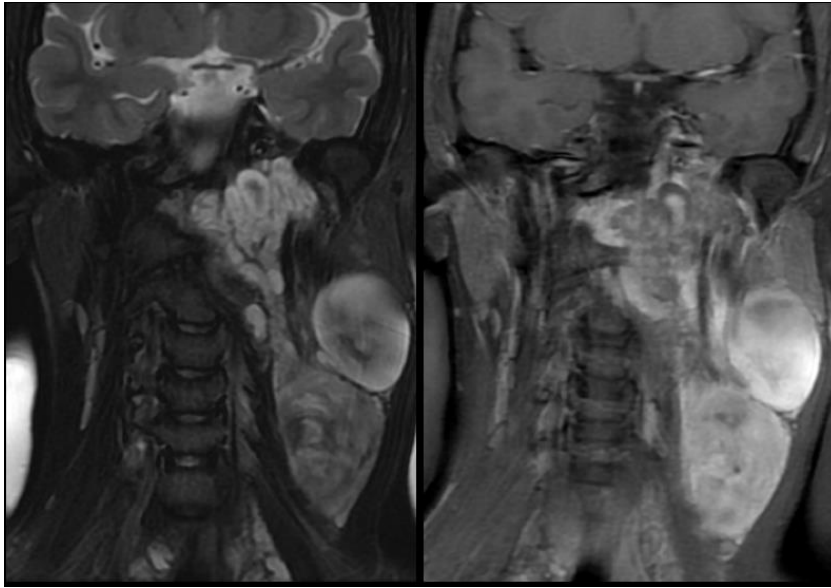


Figure 4. Malignant Peripheral Nerve Sheath Tumor. Coronal T2-weighted MRI (left) and T1-weighted post-contrast MRI (right) demonstrating large, extensive, multilobar conglomerate mass of the left neck involving the left retropharyngeal, carotid, and suboccipital spaces. Note the heterogeneous enhancement of the mass seen on the T1-weighted post-contrast scan.

Malignant Peripheral Nerve Sheath Tumor (MPNST): Clinical Presentation

PN can undergo malignant transformation to MPSNTs. These are aggressive tumors that are often fatal despite intensive therapy. These occur in less than 5% of patients with NF1 [6, 67] at a mean age of 28.7 years [75]. Rapid or asymmetric growth, often accompanied by pain, can be signs of malignant transformation. Despite imaging or tissue sampling, diagnosis can be difficult, as biopsy may miss a malignant component of a large tumor [7]. Further, the ability to diagnose MPNST has suffered from lack of specific morphological, immunohistochemical and molecular criteria and tests [76]. Neurofibromas can occur anywhere in the body in patients with NF1, and at least 40% of affected adults have internal neurofibromas, which may not be apparent on physical examination, thus the use of whole-body MRI to characterize tumor burden may be helpful. In a study of NF1 patients with MPNSTs and age- and sex-matched NF1 patients without MPNSTs, internal neurofibromas were detected with whole-body MRI in 100% of the NF1 patients with MPNST under the age of 30, thus advocating the use of whole-body MRI in young patients with NF1 to detect internal tumors which have malignant potential. In addition, there is an association between the median number of subcutaneous neurofibromas in NF1 patients and the occurrence of MPNSTs overall [77].

A grading system separates low-grade from high-grade MPNST. Roughly 85% of these tumors fall into the high-grade category, characterized by cytologic atypia, high numbers of mitoses, and hypercellularity with or without necrosis [76].

MPNST: Imaging Characteristics

Malignant transformation of PNs to MPNSTs remains a challenging problem to diagnose and manage. PNs are carefully followed by MRI or CT, and patients are monitored for features of malignancy including unremitting pain or a rapid increase in tumor size [78]. MRI is used for tumor surveillance but often does not reliably differentiate malignant from non-malignant tissue [7]. Imaging characteristics of malignant transformation include rapid growth of tumors and ill-defined margins, as well as invasion into surrounding structures (Figure 4) [7, 79].

A recent study examined the predictive value of [18 F]-fluorodeoxyglucose (FDG) positron emission tomography (PET) to identify malignant transformation of PN in children with NF1 and found that with a pre-defined standard uptake value, sensitivity and specificity for determining benign from malignant tumors were 1.0 and 0.94, respectively [80]. This indicates that FDG-PET may be a useful imaging modality to follow PNs at risk for malignant transformation.

MPNST: Prognosis

These tumors are the leading cause of mortality in adult patients with NF1, as the lifetime risk increases to up to 13% [6, 67]. Median survival is about 30 months among children 1-17 years of age with MPNST [81], and five-year survival is only 16% [75]. Of note, NF1 patients tend to have a worse outcome with statistically significant lower disease-free survival than non-NF1 patients, with more recurrences and ultimately deaths [82]. Numerous studies cite disease-specific survival for NF1 patients with MPNST 16-38%, compared with 42-57% in non-NF1 patients [83].

MPNST: Treatment Modalities

1. Role of Surgery

Complete surgical resection of MPNST is the mainstay of treatment and is a strong predictor of long-term survival. Achieving tumor-free histological margins is the goal in every NF1 patient. However, MPNSTs are very aggressive tumors with common local recurrences even in spite of tumor-free margins. In addition, the location of these tumors sometimes can preclude complete resection; MPNSTs affecting the head and neck are rarely amenable to gross total resection without resulting in significant associated disfigurement and deformation of surrounding structures [84].

2. Radiation Therapy

Radiotherapy is generally recommended for all MPNST [82]. This may be an important adjunct to surgical therapy to prevent local recurrence. A recent study examining 175 patients with MPNST over 25 years revealed higher disease-specific survival (DSS) of 60% at 5 years and 45% 10 years, compared with DSS reported in the literature of 34-52% and 22-34% respectively, likely related to better local control attributable to aggressive surgical resection followed by radiation therapy [83].

3. Chemotherapy

The role of chemotherapy in treatment of MPNST remains controversial, given the lack of sufficient data. There have been no controlled studies evaluating the role of chemotherapy alone in patients with MPNST [82]. Despite this, chemotherapy is generally used in combination with surgery and radiation to treat MPNST. Several agents have been reported including gemcitabine, docetaxel, caboplatin, etoposide, dactinomycin, cisplatin, vincristine, cyclophosphamide, imidazole carboxamide, doxorubicin, and ifosfamide, but overall, the clinical benefits of each agent have been variable and thus inconclusive [82]. As these tumors have such a poor prognosis, most centers recommend early, aggressive, multi-modal therapy, including chemotherapy. A study evaluating the roles of the chemotherapeutic agents doxorubicin and ifosfamide, along with surgery and radiation therapy, demonstrated a two-year survival of 80% in patients with MPNST, which is higher than the reported rates of 54-73% from other studies of non-chemotherapeutic treatments [82].

4. Targeted Therapy

a) Tyrosine Kinase Inhibitor Treatment

A recent trial using imatinib mesylate to treat patients with MPNST was terminated early due to slow recruitment and lack of response (study number NCT00427583).

b) Angiogenic Pathways

VEGF expression is also increased in MPNSTs [85]. It was shown *in vitro* that tumor angiogenesis is related to the malignant potential of MPNSTs, which correlates to increased VEGF expression [85]. *In vivo* treatment of human NF-1 malignant neurogenic sarcoma xenografts with VEGF receptor inhibitor (SU5416) demonstrated reduced angiogenesis and reduced tumor volume, with reduction in tumor cell proliferation and increase in apoptosis [85].

c) Cytoskeletal Remodeling

A recent study examined single-nucleotide polymorphism (SNP) genotyping and copy number alteration (CNA), loss-of-heterozygosity (LOH), and copy number neutral-LOH (CNN-LOH) analyses of DNA from 15 MPNSTs, five PNs, and patient-matched lymphocyte DNA. MPNSTs, but not PNs, were found to have high-level CNN-LOH. The genes involved in the MPNST CNAs included the ITGB8, PDGFA, Ras-related C3 botulinum toxin substrate 1 (RAC1) (7p21-p22), PDGFRL (8p22-p21.3), and matrix metalloproteinase 12 (MMP12) (11q22.3) genes. The pathways involved in alteration of these genes included amplification of the Ras-homologous (Rho-) GTPase signaling pathway. Rho-GTPases are regulators of actin organization, cell motility, cell-cell adhesion, cell-extracellular matrix adhesion, cell cycle progression, and apoptosis, which are key processes involved in malignant transformation and metastasis. Thus, these copy number gains may play a role in malignant transformation of benign tumors such as PNs. Further, specific inhibitors of parts of the Rho-GTPase pathway may reduce the malignant potential of these cells [86]. These are potential new treatment strategies for treating MPNST.

CONCLUSION

In recent years, a number of novel molecular targets have been identified in tumors associated with NF1. Developing targeted therapies that exploit these biological vulnerabilities holds the promise of not only prolonging the life of patients affected with NF1, but also reducing the significant side effects associated with current therapies such as traditional chemotherapy and radiation therapy.

REFERENCES

- [1] Jett K., Friedman J. M. Clinical and genetic aspects of neurofibromatosis 1. *Genet Med* 2010;12: 1–11.
- [2] Denayer E., de Ravel T., Legius E. Clinical and molecular aspects of RAS related disorders. *J Med Genet* 2008;45:695–703.
- [3] Basu T. N., Gutmann D. H., Fletcher J. A. et al. Aberrant regulation of ras proteins in malignant tumor cells from type 1 neurofibromatosis patients. *Nature*. 1992;356:713-715.
- [4] Szudek J. P., Birch, et al. Associations of clinical features in neurofibromatosis 1 (NF1). *Genet epidemiol* 2000(19);4:429-439.
- [5] Walker L., Thompson D., Easton D. et al. A prospective study of neurofibromatosis type 1 cancer incidence in the UK. *Br J Cancer* 2006;95:233–238.
- [6] Rasmussen S. A., Yang Q., Friedman J. M. Mortality in neurofibromatosis 1: an analysis using U.S. death certificates. *Am J Hum Genet* 2001;68:1110–1118.
- [7] Korf B. R. Malignancy in neurofibromatosis type 1. *Oncologist* 2000;5(6):477-485.
- [8] Turgut M., Ozcan O. E. et al. Central neurofibromatosis. Case report and review of the literature. *Eur Neurol* 1991;31(4):188-192.
- [9] Listernick R. J., Charrow et al. Intracranial gliomas in neurofibromatosis type 1. *Am J Med Genet* 1999;89(1):38-44.
- [10] Dougherty M. J., Santi M., Brose M. S. et al. Activating mutation in BRAF characterizes a spectrum of pediatric low-grade gliomas. *J Neuro oncol.* 2010;12:621-630.
- [11] Pfister S., Janzarik W. G., Remke M. et al. BRAF gene duplication constitutes a mechanism of MAPK pathway activation in low-grade astrocytomas. *J Clin Invest* 2008; 118: 1739-1749.
- [12] Schiffman J. D. et al. Oncogenic BRAF mutation with CDKN2A inactivation is characteristic of a subset of pediatric malignant astrocytomas. *Cancer Res* 2010;70: 512–519.
- [13] Jones D. T. W., Kocialkowski S., Liu L., Pearson D. M., Ichimura K., Collins V. P. Oncogenic RAF1 rearrangement and a novel BRAF mutation as alternatives to KIAA1529:BRAF fusion in activating the MAPK pathway in pilocytic astrocytoma. *Oncogene* 2009;28(20): 2119-2123.
- [14] Jones D. T. W., Kocialkowski S., Liu L. et al. Tandem duplication producing a novel oncogenic BRAF fusion gene defines the majority of pilocytic astrocytomas. *Cancer Res.* 2008;68:8673-8677.

- [15] Hawkins C., Walker E., Mohamed N. et al. BRAF-KIAA1549 fusion predicts better clinical outcome in pediatric low-grade astrocytoma. *Clin Cancer Res* 2011; 17(14): 4790-4798.
- [16] Sievert A. J., Jackson E. M., Gai X. et al. Duplication of 7q34 in pediatric low-grade astrocytomas detected by high-grade single nucleotide polymorphism-based genotype arrays results in a novel BRAF fusion gene. *Brain Pathol.* 2009;19(3):449-458.
- [17] Schindler G., Capper D., Meyer J. et al. Analysis of BRAF V600E mutation in 1,320 nervous system tumors reveals high mutation frequencies in pleomorphic xanthroastrocytoma, ganglioglioma and extracerebellar pilocytic astrocytoma. *Acta Neuropathol* 2011;121(3):397-405.
- [18] Packer R. J., Ater J., Allen J. et al. Carboplatin and vincristine chemotherapy for children with newly diagnosed progressive low-grade gliomas. *J Neurosurg.* 1997;5: 747-754.
- [19] Sen G. C., Lengyel P. The interferon system: a bird's eye view of it's biochemistry. *J Biol Chem* 1992;267:5017-5020.
- [20] Jakacki R. I., Dombi E., Potter D. M. et al. Phase I trial of pegylated interferon-alpha-2b in young patients with plexiform neurofibromas. *Neurology* 2011;76:265.
- [21] Listernick R., Charrow J., Greenwald M. et al. Natural history of optic pathway tumors in children with neurofibromatosis type 1: a longitudinal study. *J Pediatr.* 1994;125:63-66.
- [22] Listernick R., Charrow J., Greenwald M. J. et al. Optic gliomas in children with neurofibromatosis type 1. *J Pediatr.* 1989;114:788-792.
- [23] Fisher M. J., Loguidice M., Gutmann D. H. et al. Visual outcomes in children with neurofibromatosis type 1-associated optic pathway glioma following chemotherapy: a multicenter retrospective analysis. *Neuro Oncol* 2012;14(6):790-797.
- [24] Friedman, J. M. Neurofibromatosis 1: Clinical Manifestations and Diagnostic Criteria. *J Child Neurol* 2002(17):548-554.
- [25] Cnossen M. H., Stam E. N., Cooman L. C. et al. Endocrinologic disorders and optic pathway gliomas in children with neurofibromatosis type 1. *Pediatrics* 1997;100:667-670.
- [26] Habiby R., Silverman B., Listernick R. et al. Precocious puberty in children with neurofibromatosis type 1. *J Pediatr* 1995;126:364-367.
- [27] Yohay K. Neurofibromatosis types 1 and 2. *Neurologist* 2006;12(2):86-93.
- [28] Ilgren E. B., Kinnier-Wilson L. M., Stiller C. A. Gliomas in neurofibromatosis: a series of 89 cases with evidence for enhanced malignancy in associated cerebellar astrocytomas. *Pathol Annu* 1985;20(part 10):331-358.
- [29] Rosser T., Packer R. J. Intracranial neoplasms in children with neurofibromatosis 1. *J Child Neurol* 2002;17(8):630-637.
- [30] Molloy P. T., Bilaniuk L. T., Vaughn S. N. et al. Brainstem tumors in patients with neurofibromatosis type 1: a distinct clinical entity. *Neurology* 1995;45:1897-1902.
- [31] Milstein J. M., Geyer J. R., Berger M. S. et al. Favorable prognosis for brain stem gliomas in neurofibromatosis. *J Neurooncol* 1989;7:367-371.
- [32] Ullrich N. J., Raja A. I., Irons M. B., Kieran M. W. Goumnerova L. Brainstem lesions in neurofibromatosis type 1. *Neurosurgery* 2007;61(4):766-767.
- [33] Balestri P. L., Calistri, et al. Central nervous system imaging in reevaluation of patients with neurofibromatosis type 1. *Childs Nerv Syst* 1993;9(8):448-451.
- [34] Mukonoweshuro W., Griffiths P. D. et al. Neurofibromatosis type 1: the role of neuroradiology. *Neuropediatrics* 1999;30(3):111-119.

- [35] Aoki S., Barkovich A. J. et al. Neurofibromatosis types 1 and 2: cranial MR findings. *Radiology* 1989;172(2):527-534.
- [36] Fischbein N. J., Dillon W. P. et al. *Teaching atlas of brain imaging*. 2000, New York, Thieme.
- [37] Guillamo J. S., Creange A., Kalifa C. et al. Prognostic factors of CNS tumours in Neurofibromatosis 1 (NF1): A retrospective study of 104 patients. *Brain* 2003;126:152-160.
- [38] Parsa C. F., Hoyt C. S., Lesser R. L. et al. Spontaneous regression of optic gliomas: thirteen cases documented by serial neuroimaging. *Arch Ophthalmol*. 2001;119:516-529.
- [39] Listernick R., Ferner R. E., Piersall L. et al. Late-onset optic pathway tumors in children with neurofibromatosis 1. *Neurology* 2004;63(10):1944-1946.
- [40] Pollack I. F., Mulvihill J. J. Special issues in the management of gliomas in children with neurofibromatosis 1. *J Neurooncol* 1996;28(2-3):257-268.
- [41] Wentworth S., Pinn M., Bourland J. D. et al. Clinical experience with radiation therapy in the management of neurofibromatosis-associated central nervous system tumors. *Int J Radiat Oncol Biol Phys* 2009;73:208-213.
- [42] Jenkin D., Angyalfi S., Becker L. et al. Optic glioma in children: surveillance, resection, or irradiation? *Int J Radiat Oncol Biol Phys* 1993;25:215-225.
- [43] Grill J., Couanet D., Cappelli C. et al. Radiation-induced cerebral vasculopathy in children with neurofibromatosis and optic pathway glioma. *Ann Neurol* 1999;45:393-396.
- [44] Broniscer A., Baker S. J., West A. N. et al. Clinical and molecular characteristics of malignant transformation of low-grade gliomas in children. *J Clin Oncol* 2007;25:682-689.
- [45] Rodriguez F. J., Scheithauer B. W., Burger P. C. et al. Anaplasia in pilocytic astrocytoma predicts aggressive behavior. *Am J Surg Pathol*. 2010;34:147-160.
- [46] Kortmann R. D., Timmermann B., Taylor R. E. et al. Current and future strategies in radiotherapy of childhood low-grade glioma of the brain. Part II: Treatment-related late toxicity. *Strahlenther Onkol* 2003;179:585-597.
- [47] Ater J. L., Zhou T., Holmes E. et al. Treatment of Low-Grade Glioma in Young Children: A Report From the Children's Oncology Group. *J Clin Oncol* 2012;30:2641-2647.
- [48] Peters K. B., Cummings T. J., Gururangan S. Transformation of Juvenile Pilocytic Astrocytoma to Anaplastic Pilocytic Astrocytoma in Patients With Neurofibromatosis Type 1. *J Pediatr Hematol Oncol* 2011;33(5):198-201.
- [49] Nicholson H. S., Kretschmar C. S., Krailo M. et al. Phase 2 Study of Temozolomide in Children and Adolescents With Recurrent Central Nervous System Tumors: a report from the Children's Oncology Group. *Cancer* 2007;110(7):1542-1550.
- [50] Bouffet E., Jakacki R., Goldman S. et al. Phase II Study of Weekly Vinblastine in Recurrent or Refractory Pediatric Low-Grade Glioma. *J Clin Oncol* 2012(30):1358-1363.
- [51] Packer R. J., Jakacki R., Horn M. et al. Objective Response of Multiply Recurrent Low-Grade Gliomas to Bevacizumab and Irinotecan. *Pediatr Blood Cancer* 2009;52:791-795.
- [52] Johannessen C. M., Reczek E. E., James M. F., Brems H., Legius E. and Cichowski K. The NF1 tumor suppressor critically regulates TSC2 and mTOR. *Proc Natl Acad Sci USA* 2005;102:8573-8578.
- [53] Dasgupta B., Yi Y., Chen D. Y., Weber J. D. and Gutmann D. H. Proteomic analysis reveals hyperactivation of the mammalian target of rapamycin pathway in

- neurofibromatosis 1-associated human and mouse brain tumors. *Cancer Res* 2005b;65:2755-2760.
- [54] Jentoft M. C., Giannini C., Cen L. et al. Phenotypic variations in NF1-associated low grade astrocytomas: possible role for increased mTOR activation in a subset. *Int J Clin Exp Pathol* 2011;4(1):43-57.
- [55] de Clue J. E., Papageorge A. G., Fletcher J. A., Diehl S. R., Ratner N., Vass W. C. and Lowy D. R. Abnormal regulation of mammalian p21ras contributes to malignant tumor growth in von Recklinghausen (type 1) neurofibromatosis. *Cell* 1992;69:265-273.
- [56] Dasgupta B., Li W., Perry A. and Gutmann D. H. Glioma formation in neurofibromatosis 1 reflects preferential activation of K-RAS in astrocytes. *Cancer Res* 2005a; 65: 236-245.
- [57] Li H., Velasco-Miguel S., Vass W. C., Parada L. F. and de Clue J. E. Epidermal growth factor receptor signaling pathways are associated with tumorigenesis in the Nf1:p53 mouse tumor model. *Cancer Res* 2002;62:4507-4513.
- [58] Peyssonnaud C., Eychene A. The Raf/MEK/ERK pathway: new concepts of activation. *Biol Cell*. 2001;93:53-62.
- [59] Forsheew T., Tatvossian R. G., Lawson A. R. J. et al. Activation of the ERK/MAPK pathway: a signature defect in posterior fossa pilocytic astrocytomas. *J Pathol*. 2009; doi: 10.1002/path.2558.
- [60] Jeuken J., Wesseling P. MAPK pathway activation through BRAF gene fusion in pilocytic astrocytomas; a novel oncogenic fusion gene with diagnostic, prognostic, and therapeutic potential. *J Pathol* 2010; 222:324-328.
- [61] Lin A., Rodriguez F. J., Karajannis M. A. et al. BRAF Alterations in Primary Glial and Glioneural Neoplasms of the Central Nervous System With Identification of 2 Novel KIAA1549:BRAF Fusion Variants. *J Neuropathol Exp Neurol* 2012;71(1):66-72.
- [62] Horbinski C., Hamilton R. L., Nikiforov Y., Pollack I. F. Association of molecular alterations, including BRAF, with biology and outcome in pilocytic astrocytomas. *Acta Neuropathol* 2010;119(5):641-649.
- [63] Lott I. T., Richardson E. P. Jr. Neuropathological findings and the biology of neurofibromatosis. *Adv Neurol*. 1981;29:23-32.
- [64] Cichowski K., Jacks T. NF1 tumor suppressor gene function: narrowing the GAP. *Cell* 2001;104:593-604.
- [65] de Bella K., Szudek J., Friedman J. M. Use of the national institutes of health criteria for diagnosis of neurofibromatosis 1 in children. *Pediatrics*. 2000;105(3 pt 1):608-614.
- [66] Tucker T., Friedman J. M., Friedrich R. E., Wenzel R., Fünsterer C., Mautner V.-F. Longitudinal study of neurofibromatosis 1 associated plexiform neurofibromas. *J Med Genet* 2009;46:81-85.
- [67] Evans D. G., Baser M. E., McGaughan J. et al. Malignant peripheral nerve sheath tumours in neurofibromatosis 1. *J Med Genet*. 2002;39:311-314.
- [68] Needle M. N., Cnaan A., Dattilo J. et al. Prognostic signs in the surgical management of plexiform neurofibroma: the Children's Hospital of Philadelphia experience, 1974–1994. *J Pediatr* 1997;131:678–682.
- [69] Sachdev S., Dodd R. L., Chang S. D. et al. Stereotactic Radiosurgery Yields Long-Term Control for Benign Intradural, Extramedullary Spinal Tumors. *Neurosurgery* 2011;69:533-539.

- [70] Gupta A., Cohen B. H., Ruggieri P., Packer R. J., Phillips P. C. Phase I study of thalidomide for the treatment of plexiform neurofibroma in neurofibromatosis 1. *Neurology* 2003;60:130–132.
- [71] Rak J., Mitsuhashi Y., Bayko L., Filmus J., Shirasawa S., Sasazuki T. and Kerbel R. S. Mutant ras oncogenes upregulate VEGF/VPF expression: implications for induction and inhibition of tumor angiogenesis. *Cancer Res* 1995;55:4575–4580.
- [72] Kawachi Y., Xu X., Ichikawa E., Imakado S. and Otsuka F. Expression of angiogenic factors in neurofibromas. *Exp Dermatol* 2003;12:412–417.
- [73] Ling B. C., Wu J., Miller S. J., Monk K. R., Shamekh R., Rizvi T. A. Decourten-Myers G., Vogel K. S., de Clue J. E. and Ratner N. Role for the epidermal growth factor receptor in neurofibromatosis-related peripheral nerve tumorigenesis. *Cancer Cell* 2005;7:65–75.
- [74] Demestre M., Herzberg J., Holtkamp N. et al. Imatinib mesylate (Glivec) inhibits Schwann cell viability and reduces the size of human plexiform neurofibroma in a xenograft model. *J Neurooncol* 2010;98:11–19.
- [75] Ducatman B., Scheithauer B. W., Piepgras D. G. et al. Malignant peripheral nerve sheath tumors. A clinicopathological study of 120 cases. *Cancer* 1986;57:2006–2021.
- [76] Rodriguez F. J., Folpe A. L., Giannini C., Perry A. Pathology of peripheral nerve sheath tumors: diagnostic overview and update on selected diagnostic problems. *Acta Neuropathol* 2012;123:295–319.
- [77] Mautner V. F., Asuagbor F. A., Dombi E. et al. Assessment of benign tumor burden by whole-body MRI in patients with neurofibromatosis 1. *Neuro Oncol* 2008;10(4):593–598.
- [78] Ferner R. E., Gutmann D. H. International consensus statement on malignant peripheral nerve sheath tumors in neurofibromatosis. *Cancer Res* 2002;62:1573–1577.
- [79] Chhabra A., Soldatos T., Durand D. J., Carrino J. A., McCarthy E. F., Belzberg A. J. The role of magnetic resonance imaging in the diagnostic evaluation of malignant peripheral nerve sheath tumors. *Indian J Cancer* 2011;48(3):328–334.
- [80] Tsai L. L., Drubach L., Fahey F., Irons M., Voss S., Ullrich N. J. [18F]-Fluorodeoxyglucose positron emission tomography in children with neurofibromatosis type 1 and plexiform neurofibromas: correlation with malignant transformation. *J Neurooncol* 2012;108:469–475.
- [81] Frederich R. E., Hartmann M., Mautner V. G. Malignant peripheral nerve sheath tumors (MPNST) in NF1-affected children. *Anticancer Res* 2007;27(4A):1957–1960.
- [82] Moretti V. M., Crawford E. A., Staddon A. P., Lackman R. D., Ogilvie C. M. Early outcomes for malignant peripheral nerve sheath tumor treated with chemotherapy. *Am J Clin Oncol* 2011;34(4):417–421.
- [83] Stucky C. H., Johnson K. N., Gray R. J. et al. Malignant Peripheral Nerve Sheath Tumors (MPNST): The Mayo Clinic Experience. *Ann Surg Oncol* 2012;19:878–885.
- [84] Ferner R. E. and Gutmann D. H. International consensus statement on malignant peripheral nerve sheath tumors in neurofibromatosis. *Cancer Res* 62: 1573–1577, 2002.
- [85] Angelov L., Sahlia B., Roncari L., McMahon G. and Guha A. Inhibition of angiogenesis by blocking activation of the vascular endothelial growth factor receptor 2 leads to decreased growth of neurogenic sarcomas. *Cancer Res* 1999;59:5536–5541.
- [86] Upadhyaya M., Spurlock G., Thomas L. et al. Microarray-Based Copy Number Analysis of Neurofibromatosis Type-1 (NF1)-Associated Malignant Peripheral Nerve Sheath Tumors Reveals a Role for Rho-GTPase Pathway Genes in NF1 Tumorigenesis. *Hum Mutat.* 2012 Apr;33(4):763–76.

Chapter 110

OUTCOME MEASURES FOR OPTIC PATHWAY GLIOMAS

Peter M. K. de Blank¹, Grant T. Liu² and Michael J. Fisher³

¹Division of Pediatric Hematology-Oncology, Rainbow Babies & Children's Hospital, Cleveland, OH; Department of Pediatrics, Case Western Reserve University School of Medicine, Cleveland, OH, US

²Neuro-Ophthalmology Service, The Children's Hospital of Philadelphia; Departments of Neurology and Ophthalmology, The Perelman School of Medicine at the University of Pennsylvania, Philadelphia, PA, US

³Division of Oncology, The Children's Hospital of Philadelphia, PA; Department of Pediatrics, The Perelman School of Medicine at the University of Pennsylvania, Philadelphia, PA, US

ABSTRACT

Optic pathway gliomas (OPG) occur in 15% of children with Neurofibromatosis type 1 (NF1) and will lead to visual deficits in up to half. Because this tumor rarely threatens life but often leads to uncorrectable and permanent vision loss, preserving vision is the primary objective in management of OPGs. However, measuring visual function with ophthalmology exams can be challenging in young children with NF1-associated OPG. Recently, a variety of surrogate outcome measures have been investigated to determine if reliable quantifiable markers of vision can be found.

In this chapter, we review the presentation and management of OPGs in children with NF1. We discuss endpoints commonly used in clinical trials, including assessments of visual function and radiologic progression. Finally, we investigate and compare recent putative surrogates of vision in OPG, and we discuss their utility in identifying and following vision loss in children with NF1.

Currently, visual acuity is the most reliable, comparable and quantifiable outcome measure available in the assessment of patients with OPG. However, novel physiologic and radiologic markers of vision loss may offer an important adjunct to the assessment of the child with NF1-associated OPG in the future.

INTRODUCTION

The most common brain tumor in children with Neurofibromatosis type 1 (NF1) is the optic pathway glioma (OPG), [1] a low-grade neoplasm that can involve any part of the precortical visual pathway, including the optic nerves, chiasm, tracts and radiations. This tumor accounts for 2 to 5% of all brain tumors in childhood, [2] but it affects as many as 15–20% of children with NF1. [3-5] OPGs can cause uncorrectable vision loss, proptosis, and hypothalamic dysfunction. Some degree of vision loss occurs in up to 50% of children with this tumor. [6, 7] Identifying individuals that require treatment, and in some patients measuring the extent of visual deficits, are challenges in NF1-associated OPGs.

The development of clinically useful endpoints for OPGs is hampered by a lack of consensus in the management of these tumors. In a survey of 10 leading NF1 centers, Fisher et al. [8] demonstrated the variability in treatment indications for OPG. Although most OPG patients were treated for some degree of vision loss, treatment was often initiated for a variety of other indications, including tumor characteristics (growth, size, location, or enhancement on MRI), optic disc changes (pallor or swelling) and unreliability of the visual exam. There also appeared to be a lack of consensus among institutions regarding the relative importance of these variables. Thus, it has been difficult to develop meaningful and uniform outcome measures to help guide treatment decisions. Many clinical trials use radiologic progression as their sole endpoint — a less clinically relevant endpoint than visual acuity loss or other neurologic or endocrine dysfunction. Currently, an international collaborative effort is underway to better define clinically relevant endpoints in OPG with an emphasis on visual acuity (Response Evaluation in Neurofibromatosis and Schwannomatosis, REiNS, <http://www.reinscollaboration.org>). In addition, Avery et al. have recently outlined appropriate visual acuity testing methods across the age spectrum to further the development of OPG research and therapies. [9]

Designing clinically relevant outcome markers for OPGs is vital to therapeutic progress in this field. This chapter will briefly review the epidemiology, presentation and management of NF1-associated OPGs. It will also discuss the current outcome measures for OPGs and review potential OPG surrogate endpoints that may help determine or predict vision loss in children with these tumors.

EPIDEMIOLOGY OF OPGS

OPGs most commonly arise in young children, with a median age at diagnosis of 4.9 years. [3, 10] Although it is unusual for new symptoms to emerge after 6 years of age, [4] it is now clear that older children and adults may also develop symptomatic OPGs. [11] Some studies report an increased incidence in females; [3, 12] however, others show the ratio of females to males to be equal. [5, 13-15] Most tumors involve the anterior visual pathways, including optic nerves and chiasm, [3] but approximately 50% also involve the posterior visual pathways, including optic tracts and radiations. [7] Although estimates of the incidence of NF1 among patients with OPGs varies widely, [13, 16] one large, single-institution study suggests that up to 58% of OPGs are found in children with NF1. [17]

PRESENTATION OF OPGs

Clinical Symptoms

OPGs may or may not be symptomatic at diagnosis. Asymptomatic tumors may be identified during routine ophthalmologic screening (e.g., optic pallor), during an MRI evaluation of the brain for another reason, or at institutions where MRI's are performed routinely as part of an NF1 screening protocol. It is unclear what proportion of NF1 children diagnosed without visual deficits are actually treated pre-symptomatically and how this strategy affects visual outcomes.

Symptomatic OPGs often present with vision loss, although young children with symptomatic tumors rarely complain of decreased acuity. [6, 18] Other signs and symptoms can include headache, ocular misalignment (strabismus), nystagmus, and proptosis. Precocious puberty can occur when the tumor involves the hypothalamus. [5, 19]

Diagnostic Imaging

The diagnosis of OPG is made primarily from MR imaging of the brain and orbits, preferably with thin slices through the optic nerve and chiasm. Biopsy is not necessary for lesions with characteristic clinical and imaging features, particularly in patients with NF1, and is reserved for unusual presentations in which the diagnosis is in doubt. [20] MRI is also the standard modality for surveillance of known tumors.

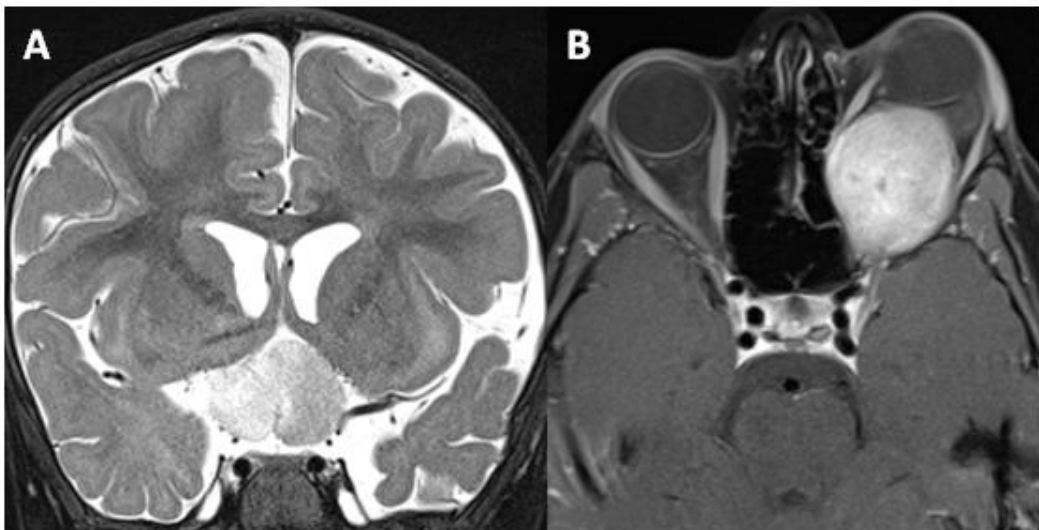


Figure 1. (A) Coronal T2-weighted MRI demonstrating an optic chiasm glioma. (B) Axial postgadolinium T1-weighted MRI of an optic nerve glioma.

OPGs generally appear isointense on T1-weighted sequences and iso- to hyperintense on T2-weighted sequences (Figure 1). Enhancement with gadolinium is common, but is more frequently seen in sporadic rather than NF1-associated OPGs. [21] Tumors of the optic nerve

may be fusiform, and kinking may cause a downward turn of the intraorbital segment of the optic nerve. [22] Although commonly seen in OPGs, nerve tortuosity alone is not sufficient for a diagnosis of an optic nerve glioma. [23] Dilatation of the subarachnoid sleeve surrounding the optic nerve can cause expansion of the dural sheath, although this "dural ectasis" may be more a developmental than a neoplastic manifestation in NF1 patients. Bilateral optic nerve tumors are almost exclusively seen in NF1-associated OPGs. Tumors of the posterior optic pathway may be diffuse, bilateral and without clear borders; they are most apparent on T2 sequences with variable degrees of enhancement. [21, 24]

PATHOLOGY

Optic pathway gliomas are histopathologically low-grade gliomas, typically pilocytic astrocytomas (WHO grade I), although other low-grade glioma variants are reported as well, [25-27] and NF1-associated and sporadic versions are histologically identical. [1] Microscopically, these tumors are composed of glial fibrillary acidic protein (GFAP) staining astrocytes, often with characteristic Rosenthal fibers. Tumors may lack these characteristic findings and be classified as fibrillary (WHO grade II) tumors; however, tumor grade has not been correlated with tumor progression or behavior. [28, 29]

MANAGEMENT OF OPGs

Clinical Course

Although OPGs are generally slow growing, [5, 15] individual tumors can show extreme variability in growth velocity. Some may have long periods of quiescence, followed by periods of rapid symptomatic growth. Occasional cases of spontaneous tumor regression have also been reported. [30-36] Unfortunately, improvement in MRI abnormalities does not always correlate with visual improvement in children. [30]

Multiple studies have investigated potential prognostic factors for NF1-associated OPG growth or vision loss. Anterior location of the OPG is associated with less aggressive behavior in some series, [7, 37] but not all. [5] Age (<2 years or >5years) has been associated with worse outcomes in some series. [8, 38, 39] Histology has not been shown to predict OPG behavior. [28, 29] In one of the largest studies, characteristics at presentation were investigated as prognostic factors for eventual treatment need in 90 children with NF1-associated OPG (51 symptomatic and 29 asymptomatic); [5] no factor (including tumor location, age at diagnosis, presenting symptoms, NF1 symptoms, or associated features) was found to be prognostic. Since there is currently no reliable indicator to predict future vision loss in OPGs, identifying visual deficits early may be the most important factor in preserving vision.

Surveillance

Child-reported vision loss is an unreliable marker of a symptomatic OPG, since young children rarely complain of vision loss. [6] Instead, children with NF1 should have annual comprehensive ophthalmologic evaluations by a pediatric ophthalmologist or neuro-ophthalmologist familiar with NF1. Although many guidelines recommend annual eye exams until 7 years of age, [6, 40] monitoring every other year should continue into adulthood given that some older children may remain at risk. [11] Comprehensive ophthalmology evaluations should include monocular visual acuity testing appropriate for the patient's age and cognitive abilities (e.g., Teller Acuity Cards, HOTV, Lea figures or Snellen, see Table 1), [9] an assessment of visual fields, color vision, and pupillary responses, and fundus examination to evaluate the optic discs.

Once symptoms or abnormal findings are identified on ophthalmologic evaluation, MRI of the brain and orbit should be used for diagnosis and localization. In general, MRI screening in asymptomatic individuals is not recommended because identification of asymptomatic OPGs does not affect visual or overall mortality outcomes. [4, 40] However, radiologic screening may be valuable if a patient is too young for a reliable ophthalmology examination. There is no clear consensus on imaging frequency once an OPG is identified, but expert reviews suggest MRI evaluations every three months for the first year after detection, with increasing intervals thereafter for stable lesions. [6]

Table 1. Age-based visual acuity testing and norms. Adapted from Listernick et al, *Ann Neurol* 2007;61:189-198

Age (years)	Recommended Testing Method	Normal Visual Acuity
0.5 – 2	Teller Acuity Test	age-based norms
3	Lea figure	20 / 40
4	HOTV matching	20 / 30
5	Snellen	20 / 25
≥6	Snellen	20 / 20

Treatment

Treatment is reserved for clinical and sometimes radiologic progression. Clinical progression consisting of a two-line decrease in visual acuity in one or both eyes on the Snellen, HOTV or Lea charts is suggested before considering treatment in an expert review,⁶ although a new visual field quadrant deficit may also prompt treatment. Caution should be applied, as determining a significant change in vision can be difficult since subject cooperation and age-defined normal acuities may vary between assessments. Many clinicians will also treat for progressive tumor growth on MRI, although radiographic growth has not necessarily correlated with progressive vision loss in multiple studies. [8, 41, 42]

Chemotherapy with carboplatin and vincristine is considered first-line treatment for NF1-associated OPGs. This regimen results in a 5-year progression free survival of 69% in children with NF1; [43] however, a high rate of allergic reactions to carboplatin led to discontinuation

of therapy in 19% of subjects. Unfortunately, for patients unable to receive carboplatin, there is no consensus of second-line treatment regimens in NF1-associated OPGs. The TPCV regimen (thioguanine, procarbazine, lomustine (CCNU) and vincristine) has equivalent efficacy to carboplatin/vincristine in subjects without NF1, [44] but this regimen is often avoided in patients with tumor predisposition syndromes such as NF1 due to the risk of second malignancy after exposure to the alkylators lomustine and procarbazine. [45] Cisplatin and etoposide have been successful in low grade gliomas, demonstrating a 3-year progression free survival of 73%, but etoposide has also been shown to increase the risk of secondary leukemia and cisplatin can cause hearing loss, which is of particular concern in patients with diminished vision. [46] Bevacizumab/irinotecan, [47] vinblastine, [48] and temozolomide [49] have also been used as second-line agents in recurrent and refractory OPGs; however temozolomide is also associated with an increased risk of developing secondary malignancy. Other agents currently under investigation in early clinical trials for NF1-associated OPGs include lenalidomide and everolimus. [50]

Despite their efficacy for tumor control, the morbidity associated with radiation therapy and surgical excision make them therapies of last resort for NF1-associated OPGs. Radiotherapy is associated with a 10-year progression free survival between 65 and 90%, [51-54] but is avoided due to the increased risk of second malignancy, [55] cerebrovascular disease, [56-58] endocrine complications [51, 52, 59] and neurocognitive deficits. [27, 54, 60] Although surgery is the mainstay of therapy for pediatric low-grade gliomas in the cerebellum and superficial cortex, resection of optic pathway gliomas can lead to further visual decline, as well as risks of endocrine and cerebrovascular complications. [45] Surgical resection may be indicated for unilateral optic nerve gliomas when vision has been lost and extensive proptosis results in corneal exposure. [6] Tumor debulking of hypothalamic tumors may also decompress hydrocephalus by removing tumor obstructing flow at the foramina of Monro. [6]

OUTCOME MEASURES

The past two decades have led to remarkable advances in our understanding of the natural history, biology and treatment of NF1-associated OPGs. Clinical trials have identified chemotherapy regimens that are safe and efficacious in preventing death or radiologic progression. Recently, however, the focus of clinical research in OPGs has shifted from preventing radiologic progression to avoiding progressive vision loss, since this is the ultimate goal of treatment. [9] Therefore, determining reliable and clinically meaningful endpoints for visual outcome is a vital part of future clinical trial design in OPGs. At this time, visual acuity is the best outcome measure for OPGs. However, surrogate endpoints and biomarkers for visual function in children with OPGs are nevertheless desirable for two major reasons: 1) some children may not cooperate sufficiently for reliable visual assessments, and 2) a test which might indicate impending vision loss before it occurs could guide treatment decisions and might preserve normal vision. Ideally, such novel surrogate markers would be generally available, minimally invasive, inexpensive, rapid and reliable. Most importantly, they need to correlate closely with visual function. In many cases, the development of potential surrogate markers is already underway. However, more study will be necessary to validate these methods prospectively before they may be useful for the clinical management of OPGs.

Anatomical Outcomes (MRI)

Because an OPG is rarely fatal, [43] most clinical trials evaluate radiologic progression free survival or tumor response to determine treatment efficacy. [61-63] Modern MRI techniques with thin slices, high resolution and gadolinium contrast agents have been a boon to the diagnosis, localization and assessment of OPGs. MRI can measure tumor volume and detect tumor growth reliably, reproducibly, and independent of patient factors such as cooperation or attention. However, traditional measures of imaging response (complete response, partial response, stable disease and progressive disease) may be difficult to measure accurately in children with NF1 who often have changing areas of benign T2-weighted hyperintensity (unidentified bright spots) and variable degrees of enhancement in the optic pathways. In addition, many children require sedation for their MRI to be completed.

Routine MRI screening in visually asymptomatic children with NF1 has been evaluated in a handful of screening programs and has not been shown to predict visual outcomes. In a review of 32 children with NF1 who were referred but not initially treated for an OPG, there was no difference in radiologic signs between patients who clinically progressed and those that did not. [64] Although contrast-enhancement is cited by some as a factor influencing a clinician's decision to treat OPGs, [8] other large studies have not found an association between enhancement and prognosis. [24] In addition, dynamic changes in MRI may not correspond with clinical changes. Retrospective studies demonstrate examples of children where change in enhancement or tumor size on MRI did not correlate with clinical outcomes. [21, 65]

As a surrogate marker in clinical trials, MRI is unable to differentiate important functional outcomes, and traditional measures of imaging response do not correspond with visual outcomes. Campagna et al. [66] in 32 children with sporadic OPG found a poor correlation (59.4%) between visual acuity and radiologic outcomes. Increased tumor volume did seem to be associated with worse visual acuity (corresponding in 10 of 11 cases with increased tumor volume), but a substantial proportion of patients with worse visual acuity were not identified by tumor progression (9 of 19 patients with worse visual acuity had either stable (n=1) or reduced (n=8) tumor volume). Shofty confirmed this trend in 19 patients who received chemotherapy for OPGs. [67] Overall, MRI changes corresponded to visual changes in 52.6%. Again, children with radiologic progression predominantly had worse visual outcomes (8 of 11 children with tumor progression also had worse visual outcomes), but children with worse visual outcomes did not consistently have tumor progression (57% progression, 21% stable, and 21% improved tumor size). Fisher et al. performed the largest study on visual outcomes in children treated for NF1 associated OPGs. [8] Their study found that visual and imaging outcomes corresponded only in 38% of the 71 children with both radiologic and visual outcome data. In that study, only 3 of 22 children with a decline in visual acuity had radiologic progression. Although it seems that children with progressive disease after therapy may have an increased risk of worse visual outcomes, it is clear that tumor progression on MRI is a poor surrogate for visual acuity as an outcome measure in children treated for OPGs.

Visual Outcomes

Change in visual acuity is one of the most important factors guiding treatment decisions in OPGs. [6, 8] While the impact of childhood-onset vision loss has never been systematically

investigated in children with OPGs, vision loss in adults affects quality of life, educational level, income and self-perception. In a study of adult vision loss from a 1958 British birth cohort, Rahi and colleagues found that all-cause visual impairment was associated with increased odds of unemployment, lower socioeconomic status and worse mental health. [68] Likewise, in a study from the 1995 National Health Interview Survey on Disability, Swanson et al. demonstrated that severe vision impairment and legal blindness were associated with difficulties in activities of daily living (ADLs) and instrumental activities of daily living (IADLs). [69] In a stratified analysis, the effect of vision on ADL/IADL function was modified by age: vision impairment and blindness produced a graded effect, with the greatest impact among the youngest subjects. For instance, the odds ratio for difficulty with eating was 6.18 (95% CI 2.52–15.17) in subjects 18–44 years old with severe visual impairment, but only 1.23 (95% CI 0.75–2.01) among subjects ≥ 65 years of age.

Visual acuity is the most important functional outcome measure in clinical trials of OPG. Unlike radiologic progression which may be clinically asymptomatic, vision loss is a significant and clinically meaningful consequence of OPG progression. Although there are other clinically meaningful complications associated with OPG (i.e., endocrinologic or neurologic), loss of visual acuity is by far the most frequent and therefore the most commonly used functional outcome measure. Among visual parameters, visual acuity is the most easily obtainable, reliable, and reproducible, even in young children, [6, 9] and the intervals for improvement or worsening (lines or logMAR – see below) are easily understandable. Other vision-related abnormalities, such as visual field defects, color vision deficits, strabismus, or nystagmus rarely occur without associated visual acuity loss in OPGs, and optic atrophy can occur without vision loss. For these reasons, and because these other features are not as reliably measured, they are less desirable than visual acuity as outcome measures for OPGs.

Comparing visual acuity longitudinally or between groups can be challenging. Vision testing relies on patient cooperation and attention, and patient factors such as fatigue due to therapy may impact test results and reliability. Learning differences and attention deficit/hyperactivity disorder are common among children with NF1 [70] and might further complicate vision testing. Children with NF1 who are unable or unwilling to complete a comprehensive ophthalmologic examination should return for repeat testing within a week or two. In addition, a review of ophthalmologic screening in 37 children followed at a single center showed that only 43% received the recommended annual ophthalmologic screening from diagnosis until seven years old. [71] Visual acuity assessments were completed successfully in all patients, but 35 of 37 (95%) of subjects were deemed too immature to undergo color vision and visual field testing.

Age-based visual acuity testing can also affect the comparability of visual outcomes. Appropriate visual acuity assessments depend on the age and developmental abilities of each child. While preferential looking tests such as Teller Acuity Cards may be used in very young and pre-verbal children, [72] older children will be able to match symbols (Lea figures) or letters (HOTV) and eventually identify letters (Snellen). While these tests all reflect visual acuity, they may not be entirely comparable. Teller Acuity Cards are a test of preferential looking, while tests in older children require recognition of the symbol and letter. These tests may under- or over-estimate visual acuity in different acuity ranges. [73, 74] The complexity of the test may also affect the results. Lea figure testing and HOTV testing each present four possible choices, giving a 25% chance of guessing each correctly, while Snellen charts use twenty or more different letter options. [9] As children age and develop, they are able to

complete more rigorous visual acuity testing. However, it is unclear how transitions among tests for visual acuity may affect reported outcomes. Further, children with NF1 may be developmentally behind their age-matched peers and, therefore, may not be able to complete the same testing as non-NF1 control subjects at an equivalent age. For this reason, it has been suggested that future clinical trials in NF1-associated OPG mandate the use of identical testing methods (Teller acuity cards or HOTV) throughout longitudinal analysis. [9]

Many older studies use a two-line decline in visual acuity measured on a Snellen chart as a significant change in vision. However, gradations in the traditional Snellen chart are not linear, thus the magnitude of a two-line decline in vision is greater when a patient's baseline is 20/70 (to 20/200) than when it is 20/25 (to 20/40). This can complicate both inter- and intra-testing comparisons. To convert the Snellen gradations to a linear scale, many researchers have used the logarithm of the minimum angle of resolution ($\log\text{MAR}$ = base 10 logarithm of (1/Snellen decimal equivalent)). For example, if the Snellen visual acuity is 20/40, the decimal equivalent is 20 divided by 40 = 0.5, and therefore the $\log\text{MAR}$ equivalent of 20/40 is $\log_{10}(1/0.5) = 0.3$). While the meaning of a single visual acuity reading in $\log\text{MAR}$ may be less intuitively obvious than Snellen readings, the change in vision can be compared across the range of visual deficits. It is generally agreed that a change in visual acuity of 0.2 $\log\text{MAR}$ is clinically significant.

ALTERNATIVE OUTCOME MEASURES

Because of the varied penetrance, clinical course and timing of OPGs and their complications, there has been increasing interest in non-invasive measures as surrogate markers for important endpoints and to predict complications that might influence patient care or quality of life. The development of inexpensive outcome measures to predict vision loss in NF1-associated OPGs is especially important and may help steer potential treatment decisions or improve monitoring of treatment effect.

However, the development of surrogate endpoints faces many challenges. OPG remains a rare tumor, and tumors that merit treatment are even less common. In addition, conventional molecular markers and immunohistochemical stains developed for low-grade gliomas are of limited value in OPGs. These markers require surgical biopsy, which is often avoided due to significant morbidity and is unnecessary in the setting of classical radiologic findings. Instead, interest in OPG surrogate endpoints has turned to alternative imaging and electrophysiologic modalities (Table 2).

Diffusion and Perfusion Imaging

Diffusion Weighted Imaging (DWI) and Apparent Diffusion Coefficient (ADC) maps measure the relative diffusion of water in each voxel of an MR image. Dynamic Contrast-Enhanced (DCE) imaging repeats sequences of images during the infusion of a contrast agent to measure the permeability of a lesion. Both have been useful in the evaluation of brain tumors. Diffusion imaging has been shown to distinguish low- versus high-grade gliomas in adults [75, 76] and differentiate enhancing pilocytic astrocytomas from higher-grade gliomas. [75] DCE

MRI has been useful in identifying glioma grade and assessing their microvasculature and heterogeneity. [77] Recent studies suggest that diffusion imaging may predict brain tumor response to therapy at an early stage, providing an early non-invasive indicator of treatment efficacy. [78, 79]

Table 2. Surrogate Markers of Vision*

Author, Year	Reference	OPGs / NF1-associated OPGs	Results	Comments
Diffusion/Perfusion Imaging				
Jost, 2008	84	27 / 14	Permeability was greater in tumors that had required treatment	No difference in diffusion-weighted imaging
Diffusion Tensor Imaging (DTI)				
Lober, 2012	87	10 / 4	Visual pathway fiber attenuation was seen in all OPGs	No correlation between visual acuity and fiber attenuation
Filippi, 2012	98	53 / 9	DTI measurements were different in NF1 vs normal controls	A correlation between vision and DTI was not measured
de Blank, 2013	138	50 / 50	Visual acuity deficits correlated with changes in white matter integrity	
Positron Emission Tomography (PET)				
Moharir, 2010	108	9 / 9	Sensitivity 62.5% and specificity 87.5% to detect symptomatic OPG	Unable to predict tumors that will need therapy
Molloy, 1999	107	24 / 24	Increased FDG uptake correlated with tumor behavior (stable vs progressive) in 96%.	Only published as an abstract
Visual Evoked Potentials (VEPs)				
Ng, 2001	124	10 / 8	50% correlation between change in VEP and change in visual acuity	VEP changes corresponded to change in MRI in 7 of 10 subjects
Trisciuzzi, 2004	125	18 / 5	VEP amplitude correlated with visual acuity deficits	12 of 18 had partial surgical excision of tumor
Kelly, 2012	130	21 / 6	Change in VEP amplitude did not correlate with change in visual acuity after treatment	
Wolsey, 2006	128	14 / 14	Sensitivity of 86% and specificity of 75% to detect symptomatic or asymptomatic OPGs	Normalization of VEP and improvement in vision after radiation in one subject.
Falsini, 2008	129	14 / 4	Change in flicker VEP testing corresponded to visual acuity changes 58% of the time.	
Optical Coherence Tomography				
Avery, 2011	26	48 / 34	Retinal Nerve Fiber Layer thickness associated with visual acuity	Prospective study ongoing

*includes only reports with more than 5 subjects that compare surrogate measures to visual outcomes.

Studies of these techniques in children and in optic pathway gliomas are limited. Lope et al. [80] examined the utility of DWI in the identification of pediatric orbital tumors. The single OPG in this sample was too small for an adequate reading of diffusivity. Sener studied a case of a one year-old child with NF1 and a chiasmatic OPG. [81] ADC values were compared with 12 control children (aged 9 month to 3 years). The ADC of the chiasmatic OPG was not

different from that of normal white matter in controls ($0.81 \times 10^{-3} \text{ mm}^2/\text{s}$ vs. $0.84 \pm 0.14 \times 10^{-3} \text{ mm}^2/\text{s}$).

Jost et al. provide the only large study of diffusion and perfusion imaging in children with OPGs. [82] They performed a cross-sectional study of 27 children with OPGs (14 with NF1) with diffusion weighted and DCE MRI protocols. Mean diffusivity values on ADC were no different between OPGs requiring and not needing treatment; however, mean permeability measured by DCE MRI was greater in more aggressive ($2.24 \text{ mL/min per } 100 \text{ cm}^3$) compared to clinically stable tumors ($1.38 \text{ mL/min per } 100 \text{ cm}^3$, $p=0.05$). Forty-seven percent of aggressive tumors but no stable tumors had permeability greater than $2 \text{ mL/min per } 100 \text{ cm}^3$, suggesting that this may be a useful threshold to identify a subset of those tumors requiring closer monitoring or treatment. A larger sample of NF1 associated OPGs will need to be studied in a longitudinal framework in order to determine if permeability correlates with progressive disease over time.

Diffusion Tensor Imaging

Diffusion Tensor Imaging (DTI) allows identification and quantitation of white matter tracts including optic nerves, tracts and radiations. DTI measures the diffusion of water molecules as they move preferentially along, as opposed to across, white matter fibers. By measuring the predominant direction of water diffusion in each voxel, a map of white matter tracts can be created. DTI can thus localize and quantify the relative number and density of selected white matter tracts (see Figure 2). By measuring the proportion of diffusion that occurs in a single direction (fractional anisotropy, FA) and the mean squared diffusion of water in three perpendicular directions (mean diffusivity, MD), [83] DTI can also examine the integrity of white matter tracts.

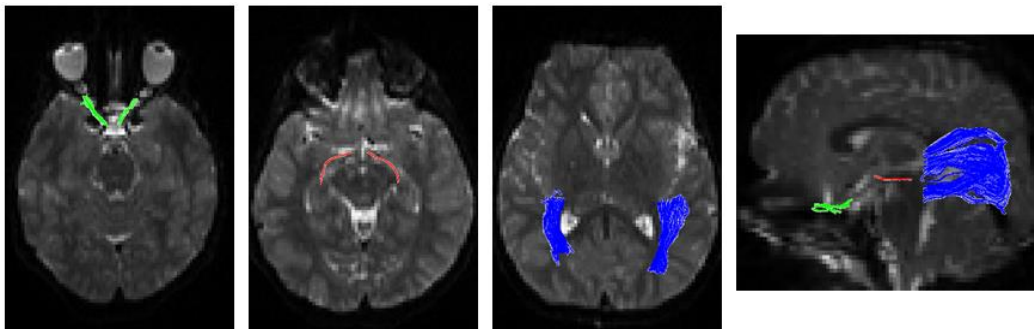


Figure 2. Visual pathway tracts from diffusion tensor imaging superimposed on axial and sagittal views of the brain. Optic nerves (green), optic tracts (red) and optic radiations (blue) are depicted.

Tract localization using DTI has been used to help guide surgical resection to avoid optic radiations, [84, 85] and measures of white matter integrity have predicted histological grade in cerebral gliomas. [86-88] In addition, DTI of the optic nerve was performed in a mouse model of NF1-associated OPG. In that model, DTI measurements showed that white matter integrity was decreased in gliomatous optic nerves, and integrity continued to decrease over time. [89]

However, there are inherent difficulties with measuring a diffusion tensor in the optic nerve. The nerves are small, subject to motion with eye movement and surrounded by intraconal fat, bone and air in the adjacent sinuses that can cause susceptibility artifact. [90-92] Further, measurements of the diffusion tensor in the optic chiasm are prone to error due to the decussation of white matter tracts. [93, 94] Despite these difficulties, Nickerson et al. were able to measure the diffusion tensor of the optic nerve in 5 children with intrinsic optic nerve pathology (septo-optic dysplasia and optic glioma), four children with extrinsic optic nerve compression (from craniopharyngioma, hypothalamic germinoma, and pilomyxoid astrocytoma), and 70 healthy pediatric controls. [95] Intrinsic lesions (including OPGs) had significantly different measures of white matter integrity (FA and MD) than either extrinsic lesions or healthy controls. Unfortunately, OPGs were not studied independent of other intrinsic lesions, and neither NF1 status nor vision was reported. In a follow-up study, the investigators performed DTI of the optic nerve and optic radiations in 9 children with NF1 (5 with OPGs) and 44 healthy age-matched controls. [96] Although OPGs were limited to the optic nerves and chiasm, children with NF1 had significantly different FA and MD measurements than healthy controls in both optic nerves and optic radiations. It is unclear whether the difference in FA and MD found in children with NF1 is due to their NF1 status or the presence of OPGs, and a lack of reported vision testing makes it impossible to relate changes in DTI parameters to visual deficits in these children.

A comparison of visual outcomes with the quantity of visual fibers in 10 children with OPGs (four with NF1) found attenuated or absent visual fibers in all subjects, but no statistically significant association could be found with abnormal visual acuity. [85] Of interest, optic radiation fibers were diminished or absent in nine of 10 children, although none had optic radiation involvement of their OPG. The deficit in posterior fibers may be due to anterograde transsynaptic degeneration from disruption of visual input, [85, 97] or perhaps indicates that DTI is more sensitive to tumor involvement in the posterior radiations than MRI.

To determine if DTI may be an appropriate endpoint for vision loss in NF1-associated OPG, our group investigated 50 children with NF1-OPG who had both a comprehensive ophthalmic examination and DTI of brain and orbits within three months of each other. [138] Multivariate analysis suggests children with abnormal visual acuity had significantly worse measures of white matter integrity (FA and MD) in the optic radiations. This difference persisted after controlling for age and tumor location, as well as in a subanalysis of subjects without tumor involvement of the radiations.

DTI offers many advantages as a potential surrogate marker of visual acuity in OPG. It is acquired with traditional MRI, and so is both non-invasive and convenient. It is also cost-effective as an “add-on” sequence if MRI is already being performed. Unlike many potential surrogate endpoints that are difficult to perform in young children, DTI can be performed at almost any age, although age-based norms for DTI measurements should be determined before wide use can be recommended. DTI can also characterize structural deficits of the posterior optic pathway, and might in the future be investigated in conjunction with optical coherence tomography (which examines the anterior pathway) to determine if these tests may predict vision loss better in combination.

For DTI to become a useful surrogate endpoint for vision loss in OPG there are several limitations to address. First, DTI measurements show a significant variability that may limit their usefulness to individuals rather than populations. Second, measurements of pre-geniculate structures (optic nerve, chiasm and tract) are technically difficult, increasing the variability of

DTI measurements in these regions. Finally, the impact of unidentified bright objects, commonly seen in the brains of NF1 patients, should be investigated to determine if these asymptomatic structures confound the relationship between DTI measurements and visual outcomes. Ultimately, longitudinal studies in larger NF1 populations are needed to determine if DTI changes predict or are provoked by vision loss.

Positron Emission Tomography

Positron emission tomography (PET) has been investigated as an alternate imaging modality for the evaluation of OPGs in NF1. PET imaging depends on a positron-emitting radionuclide tracer attached to a biologically active molecule that is introduced into the subject's body. Tracer concentration can be imaged in three dimensions and provide information on a variety of metabolic processes. The most common tracer used is [18F] fluorodeoxyglucose (FDG), which has a relatively long half-life (approximately 110 minutes) and can be used to detect levels of glucose metabolism throughout the body.

FDG-PET imaging has already proven its usefulness in patients with NF1 to help differentiate benign neurofibromas from malignant peripheral nerve sheath tumors. [98-101] In addition, the utility of PET for the clinical management of adult brain tumors, particularly for tumor detection, defining tumor extent and treatment planning is established. [102] There is relatively less information on PET in the evaluation of pediatric brain tumors. In a survey of physicians who ordered PET imaging for a child with a CNS tumor, PET was deemed to be helpful in 48 of 59 (77%) cases, but changed management in only 9 (15%). [103]

Small studies and case series suggest that PET imaging may be useful in characterizing the growth potential in OPGs. Increased FDG was reported to be useful in clinical decision-making in an adult with an OPG. [104] In a larger study describing 24 patients with NF1 and low grade gliomas of the optic pathway, thalamus or brainstem, there was a 96% correlation between tumor FDG uptake and disease status. [105] In that study, all seven subjects with progressive disease had increased FDG uptake, while 16 of 17 with stable disease had no increased uptake; however, only an abstract of this work has been published.

In a retrospective report of 16 OPGs among nine children with NF1, five of eight symptomatic tumors (four with disc pallor, three with visual impairment and one with proptosis) had abnormal FDG uptake (standardized uptake value (SUV) >3), in contrast to only one of eight asymptomatic OPGs. [106] The diagnostic sensitivity and specificity of FDG PET for identifying symptomatic tumors was 62.5% and 87.5%, respectively. Overall, FDG-PET does not appear to have the sensitivity necessary to be a useful tool in identifying symptomatic OPGs in children too young to sufficiently cooperate with ophthalmologic testing. Further, on longitudinal evaluations, FDG-PET was unable to predict which tumors would subsequently need treatment. However, two subjects with elevated SUV at baseline had an improvement in symptoms and a reduction in SUV uptake following chemotherapy.

Other case reports of PET imaging for OPGs have used alternative tracer molecules, such as alpha-[11C]methyl-L-tryptophan (AMT) [107] and [18F]fluorocholine. [108] These molecules provide better contrast between tumor tissue and surrounding healthy brain. [109, 110] However, the short half-life of AMT requires on-site production with a cyclotron, limiting its general utility. Larger studies of these alternative tracer molecules will be necessary before their potential utility in following OPGs can be determined.

PET imaging can help differentiate malignant from non-neoplastic tissue and orient therapeutic management in pediatric brain tumors, [111] but it has not yet proven its usefulness in OPGs. While new tracers may lead to increased sensitivity in detecting symptomatic OPGs, they have not been able to predict vision loss or aggressive behavior in an OPG. Further, the short half-life of many radionuclides limits their utility in many centers without an on-site cyclotron. Despite anecdotal examples of FDG uptake in OPGs, [112] FDG-PET is not sensitive to symptomatic OPGs and longitudinal evaluation has only been shown to correlate with chemotherapy response in select cases. Further, PET has innate disadvantages in the NF1 population, where unnecessary radiation should be avoided because of significant risks of second malignancies [55, 113] and cerebrovascular complications. [56, 57] While PET imaging has many uses in childhood brain tumors, radionuclide development and further investigations are necessary before PET can be useful as an endpoint in OPGs.

Visual Evoked Potentials

Visual evoked potentials (VEPs) measure cortical activity in response to visual stimuli. Electrodes are placed on the scalp in the occipital region (similar to those used for electroencephalography) and patients are shown a variety of patterns or lights to provoke electrical activity in the visual cortex. Resulting electrical potentials can be plotted against time to give information about the visual circuits from retina to occipital cortex. Decreases in amplitude or delays in evoked responses (latency) may signal damage to those circuits. VEPs are noninvasive and require no anesthesia. Thus, VEPs have been proposed as a method to detect the presence of an OPG and to monitor for vision loss.

Initial investigations with VEPs in OPG in 1983 showed that VEPs were abnormal or absent in seven children with proven or putative OPG. [114] Subsequent studies suggest that VEP can identify OPGs with a sensitivity between 90% [115] and 100% [116, 117] and specificity of 60% [115] to 69%, [117] although other studies of NF1-associated OPG revealed normal latency and amplitude. [118] Traditional VEP tests use an alternating, high-contrast checkerboard pattern. Alternative visual stimuli, including flash VEP (using rapid flashes of LED lights) and sweep VEP (using varying pattern size and contrast), resulted in decreased sensitivity but allowed younger children to be tested because less cooperation and concentration was required. [119, 120]

Although VEP testing has been shown to predict the presence of an OPG, many of the identified tumors were clinically silent. [116] Because the growth of OPGs is so variable, the utility of identifying asymptomatic OPGs is unclear. It is uncertain whether VEP can predict or identify symptomatic OPGs that require treatment. [121] Ng et al. [122] attempted to address this concern in a retrospective cohort of OPGs who had received repeated VEP, MRI and ophthalmology evaluations. In seven of ten subjects, VEP results over time corresponded with MRI results; however, a change in MRI corresponded to a change in VEP in only two of five paired evaluations. In addition, VEP corresponded with visual acuity only 50% of the time. In sum, VEP was not able to sensitively predict MRI or visual acuity changes.

VEP has been associated with visual acuity loss and visual field deficits in small studies of children with OPG. In 18 subjects with OPG and 16 age-matched controls, VEP results were correlated with visual acuity. [123] However, two-thirds of OPG patients had undergone tumor debulking, suggesting that abnormal VEP in this study may be identifying post-surgical tract

damage rather than the vision loss associated with this damage. In 15 subjects with OPG and 12 age-matched controls, Kelly et al. [124] showed that visual field loss was associated with decreased VEP amplitude. However, perimetry was not performed in the control group, and the OPG group had five subjects with NF1, while there were none in the control group. Previous studies [125] have shown that individuals with NF1 without OPG have abnormal VEP testing, suggesting the possibility of a primary abnormality of visual circuits in NF1. Therefore, it is important to control for NF1 status when evaluating VEP results.

To avoid this potential confounder, Wolsey et al. studied a retrospective cohort of 30 children with NF1 (14 with optic nerve glioma and 16 without OPG) who had undergone MRI and VEP evaluation. [126] VEP abnormalities identified optic nerve glioma with a sensitivity of 86% and specificity of 75%. Four individuals had serial VEPs and MRIs. In one subject, VEP abnormalities lagged behind MRI findings. In contrast, in two subjects, VEP abnormalities preceded MRI evidence of OPG. The last individual had normalization in VEP latency after radiation for an OPG that correlated with a decreased size on MRI and improved vision. These serial measurements suggest that VEP may be able to predict changes in MRI and may be useful in following vision after treatment in some patients.

Two recent longitudinal studies attempted to further examine the utility of VEP for monitoring children with OPGs. Falsini et al. [127] followed 14 children with OPG (four with NF1) with serial neuro-ophthalmic, MRI and flicker VEP evaluations roughly annually for a median of 3 years (range 6-76 months). Both serial flicker VEP ($K=0.67$, $p<0.001$) and visual acuity ($K=0.54$, $p<0.001$) correlated with serial MRI results 79% of the time. However, serial flicker VEP evaluations only agreed with serial visual acuity tests 58% of the time. Although both flicker VEP and visual acuity correlate with changes in MRI, visual acuity has the advantage of being a clinically relevant endpoint that might influence treatment decisions. Kelly et al. [128] recently reviewed 21 subjects (6 with NF1) aged 0.7 to nine years with bilateral OPG treated with chemotherapy, radiation or both. All subjects underwent VEP testing prior to treatment, during treatment and one to two years after treatment. Although latency was variable, all subjects had reduced amplitude in both eyes prior to treatment. Serial VEP testing showed no change in VEP amplitude between pre-treatment, during treatment or post-treatment time periods ($p>0.19$), and there was no relationship between VEP amplitude and response to treatment ($p=0.5$).

In summary, although VEP testing has been shown to be able to identify OPGs with some sensitivity, it has not yet been able to identify those tumors that require treatment, correlate with changing visual acuity or predict vision loss over time. Therefore, despite its theoretical potential as a fast and non-invasive method of interrogating visual circuits in OPG, it has not fulfilled this potential as a clinically relevant biomarker. Because of the wide variability in clinical behavior of OPGs, a clinically relevant biomarker must be able to predict or identify visual acuity loss or radiologic progression in order to affect treatment decisions.

Optical Coherence Tomography

Optical Coherence Tomography (OCT) of the retinal nerve fiber layer (RNFL) offers potential as a surrogate endpoint and biomarker of vision loss in NF1 associated OPGs. OCT relies on echo time delay of back-scattered infrared light to create a cross-sectional image of the retinal nerve fiber layer. The modality is analogous to ultrasound, but provides much better

resolution (approximately 3 micrometers). [129] OCT measurements of RNFL correspond with recovery of visual acuity and visual fields following resection of other tumors compressing the chiasm. [130, 131] In addition, OCT has been extensively investigated in multiple sclerosis and optic neuritis. Thinning of the RNFL on OCT is associated with visual field deficits in both cross-sectional [132] and longitudinal [133] analyses of subjects with a history of optic neuritis. Additionally, visual acuity, visual field and color vision deficits are associated with thinning of the RNFL in subjects with incompletely recovered optic neuritis. RNFL thinning is associated with visual acuity loss in patients with multiple sclerosis (without recent optic neuritis), [134] and OCT has been proposed as a surrogate endpoint and biomarker in multiple sclerosis.[135]

As OCT allows a rapid and inexpensive method of correlating structure and function, it may offer many advantages as a potential surrogate endpoint in OPGs. OCT may combine visual deficits in acuity, perimetry and color vision into a single structural endpoint. It requires less patient cooperation and is unaffected by patient fatigue. Although it may be difficult to perform in some young children, [136] portable devices are allowing investigators to measure RNFL during sedation.

Two investigations have examined the use of OCT in subjects with OPG. Chang et al. evaluated the RNFL of 15 subjects with NF1, nine with and six without an OPG. RNFL was significantly thinner (61.1 micrometers vs. 97.9 micrometers, $p=0.0001$) in subjects with OPGs. [137] However, visual deficits were not compared directly to OCT results in this study. Closer analysis reveals that seven of nine children with OPG had visual acuity equal or worse than 20/30, and all children without OPG had vision equal or better than 20/25, [136] suggesting that RNFL may not be simply a marker of OPG, but rather OPG with vision loss.

Avery et al. evaluated the association between RNFL thickness and visual deficits among 48 subjects with OPG (with and without NF1), 14 subjects with NF1 without OPG, and 14 healthy controls at least six years of age. [23] Among subjects with OPG, thin RNFL was associated with decreased high contrast visual acuity and low-contrast letter acuity when accounting for age and tumor location. Although a high degree of variability was seen in RNFL measurements, normal RNFL thickness was strongly associated with normal visual acuity and visual fields. The study did not perform a separate analysis restricted to NF1 subjects, but it did show that there was no significant difference in RNFL thickness between NF1 subjects without OPGs and healthy controls.

OCT measurement of RNFL holds promise as a potential surrogate endpoint for vision loss in children with OPGs. However, the correlation between RNFL thickness and visual acuity will have to be replicated in children less than 6 years, who are at the greatest risk for vision loss. Smaller, faster portable machines may make reliable measurements of RNFL in younger subjects possible, and studies of this age group are underway. Measurements of RNFL demonstrate significant variability, and larger populations should be studied to demonstrate the sensitivity and specificity of identifying vision loss with OCT. In addition, studies of OCT in optic neuritis reveal that significant changes in RNFL thickness lagged 4-6 months behind vision changes. [133] If this holds true for OPGs, OCT may not be able to predict acute vision loss, which is important for treatment decisions. Longitudinal studies currently enrolling children with NF1 as young as 18 months may help answer this question.

CONCLUSION

Currently, visual acuity is the most consistent and comparable measure of visual function aiding the management of children with NF1-associated OPG. However, the development of surrogate endpoints for vision loss offers the potential for reliable non-invasive tests that may work side by side with comprehensive ophthalmologic evaluations. Ideally, these potential biomarkers may also help predict subsequent vision loss and response to chemotherapy in this group that shows such enormous clinical variability. However, further research will be necessary before any of these novel potential surrogate markers are ready to help guide clinical management and decision-making in NF1-associated OPGs.

REFERENCES

- [1] Albers AC, Gutmann DH. Gliomas in patients with neurofibromatosis type 1. *Expert Rev Neurother.* Apr 2009;9(4):535-539.
- [2] Listernick R, Charrow J. Neurofibromatosis type 1 in childhood. *J Pediatr.* Jun 1990;116(6):845-853.
- [3] Listernick R, Charrow J, Greenwald M, Mets M. Natural history of optic pathway tumors in children with neurofibromatosis type 1: a longitudinal study. *J Pediatr.* 1994;125:63-66.
- [4] Listernick R, Louis DN, Packer RJ, Gutmann DH. Optic pathway gliomas in children with neurofibromatosis 1: consensus statement from the NF1 optic pathway glioma task force. *Annals of Neurology.* 1997;41:143-149.
- [5] King A, Listernick R, Charrow J, Piersall L, Gutmann DH. Optic pathway gliomas in neurofibromatosis type 1: the effect of presenting symptoms on outcome. *Am J Med Genet A.* Oct 1 2003;122A(2):95-99.
- [6] Listernick R, Ferner RE, Liu GT, Gutmann DH. Optic pathway gliomas in neurofibromatosis-1: controversies and recommendations. *Ann Neurol.* Mar 2007;61(3):189-198.
- [7] Balcer LJ, Liu GT, Heller G, et al. Visual loss in children with neurofibromatosis type 1 and optic pathway gliomas: relation to tumor location by magnetic resonance imaging. *Am J Ophthalmol.* 2001;131:442-445.
- [8] Fisher MJ, Loguidice M, Gutmann DH, et al. Visual outcomes in children with neurofibromatosis type 1-associated optic pathway glioma following chemotherapy: a multicenter retrospective analysis. *Neuro Oncol.* Jun 2012;14(6):790-797.
- [9] Avery RA, Ferner RE, Listernick R, Fisher MJ, Gutmann DH, Liu GT. Visual acuity in children with low grade gliomas of the visual pathway: implications for patient care and clinical research. *J Neurooncol.* Oct 2012;110(1):1-7.
- [10] Listernick R, Charrow J, Greenwald MJ, Esterly NB. Optic gliomas in children with neurofibromatosis type 1. *J Pediatr.* May 1989;114(5):788-792.
- [11] Listernick R, Ferner RE, Piersall L, Sharif S, Gutmann DH, Charrow J. Late-onset optic pathway tumors in children with neurofibromatosis 1. *Neurology.* Nov 23 2004;63(10):1944-1946.

- [12] Czyzyk E, Jozwiak S, Roszkowski M, Schwartz RA. Optic pathway gliomas in children with and without neurofibromatosis 1. *J Child Neurol.* Jul 2003;18(7):471-478.
- [13] Dutton JJ. Gliomas of the anterior visual pathways. *Surv Ophthalmol.* 1994;38:427-452.
- [14] Janss AJ, Grundy R, Cnaan A, Savino PJ, Packer RJ, Zackai EH. Optic pathway and hypothalamic chiasmatic gliomas in children younger than age 5 years with a 6-year follow-up. *Cancer.* 1995;75:1051-1059.
- [15] Thiagalingam S, Flaherty M, Billson F, North K. Neurofibromatosis type 1 and optic pathway gliomas: follow-up of 54 patients. *Ophthalmology.* Mar 2004;111(3):568-577.
- [16] Stern J, DiGiacinto GV, Housepian EM. Neurofibromatosis and optic glioma: clinical and morphological correlations. *Neurosurgery.* Jun 1979;4(6):524-528.
- [17] Cappelli C, Grill J, Raquin M, et al. Long-term follow up of 69 patients treated for optic pathway tumours before the chemotherapy era. *Archives of Disease in Childhood.* 1998;79(4):334-338.
- [18] Hersh JH. Health supervision for children with neurofibromatosis. *Pediatrics.* Mar 2008;121(3):633-642.
- [19] Guillamo JS, Creange A, Kalifa C, et al. Prognostic factors of CNS tumours in Neurofibromatosis 1 (NF1): a retrospective study of 104 patients. *Brain.* Jan 2003;126(Pt 1):152-160.
- [20] Leonard JR, Perry A, Rubin JB, King AA, Chicoine MR, Gutmann DH. The role of surgical biopsy in the diagnosis of glioma in individuals with neurofibromatosis-1. *Neurology.* Oct 24 2006;67(8):1509-1512.
- [21] Chateil JF, Soussotte C, Pedespan JM, Brun M, Le Manh C, Diard F. MRI and clinical differences between optic pathway tumours in children with and without neurofibromatosis. *Brit J Radiol.* 2001;74(877):24-31.
- [22] Avery RA, Fisher MJ, Liu GT. Optic pathway gliomas [review]. *J Neuro-ophthalmol.* Sep 2011;31(3):269-278.
- [23] Avery RA, Liu GT, Fisher MJ, et al. Retinal nerve fiber layer thickness in children with optic pathway gliomas. *Am J Ophthalmol.* Mar 2011;151(3):542-549 e542.
- [24] Kornreich L, Blaser S, Schwarz M, et al. Optic pathway glioma: correlation of imaging findings with the presence of neurofibromatosis. *AJNR: Am J Neuroradiol.* 2001;22:1963-1969.
- [25] Hoffman HJ, Soloniuk DS, Humphreys RP, et al. Management and outcome of low-grade astrocytomas of the midline in children: a retrospective review. *Neurosurgery.* Dec 1993;33(6):964-971.
- [26] Laithier V, Grill J, Le Deley MC, et al. Progression-free survival in children with optic pathway tumors: dependence on age and the quality of the response to chemotherapy--results of the first French prospective study for the French Society of Pediatric Oncology. *J Clin Oncol.* Dec 15 2003;21(24):4572-4578.
- [27] Sutton LN, Molloy PT, Sernyak H, et al. Long-term outcome of hypothalamic/chiasmatic astrocytomas in children treated with conservative surgery. *Journal of Neurosurgery.* 1995;83:583-589.
- [28] Rodriguez FJ, Perry A, Gutmann DH, et al. Gliomas in neurofibromatosis type 1: a clinicopathologic study of 100 patients. *J Neuropathol Exp Neurol.* Mar 2008;67(3):240-249.

- [29] Cummings TJ, Provenzale JM, Hunter SB, et al. Gliomas of the optic nerve: histological, immunohistochemical (MIB-1 and p53), and MRI analysis. *Acta Neuropathol.* 2000;99(5):563-570.
- [30] Parsa CF, Hoyt CS, Lesser RL, et al. Spontaneous regression of optic gliomas: thirteen cases documented by serial neuroimaging. *Archives of Ophthalmology.* Apr 2001;119(4):516-529.
- [31] Liu GT, Lessell S. Spontaneous visual improvement in chiasmal gliomas. *Am J Ophthalmol.* 1992;114:193-201.
- [32] Perilongo G, Moras P, Carollo C, et al. Spontaneous partial regression of low-grade glioma in children with neurofibromatosis-1: a real possibility. *J Child Neurol.* 1999;14(6):352-356.
- [33] Piccirilli M, Lenzi J, Delfinis C, Trasimeni G, Salvati M, Raco A. Spontaneous regression of optic pathways gliomas in three patients with neurofibromatosis type I and critical review of the literature. *Childs Nerv Syst.* Oct 2006;22(10):1332-1337.
- [34] Zuccoli G, Ferrozzi F, Sigorini M, Virdis R, Bassi P, Bellomi M. Early spontaneous regression of a hypothalamic/chiasmatic mass in neurofibromatosis type 1: MR findings. *Eur Radiol.* 2000;10(7):1076-1078.
- [35] Gottschalk S, Tavakolian R, Buske A, Tinschert S, Lehmann R. Spontaneous remission of chiasmatic/hypothalamic masses in neurofibromatosis type 1: report of two cases. *Neuroradiology.* 1999;41(3):199-201.
- [36] Parazzini C, Triulzi F, Bianchini E, et al. Spontaneous involution of optic pathway lesions in neurofibromatosis type 1: serial contrast MR evaluation. *Ajnr: American Journal of Neuroradiology.* 1995;16(8):1711-1718.
- [37] Tow SL, Chandela S, Miller NR, Avellino AM. Long-term outcome in children with gliomas of the anterior visual pathway. *Pediatr Neurol.* Apr 2003;28(4):262-270.
- [38] Packer RJ, Ater J, Allen J, et al. Carboplatin and vincristine chemotherapy for children with newly diagnosed progressive low-grade gliomas. *J Neurosurg.* May 1997;86(5):747-754.
- [39] Fouladi M, Hunt DL, Pollack IF, et al. Outcome of children with centrally reviewed low-grade gliomas treated with chemotherapy with or without radiotherapy on Children's Cancer Group high-grade glioma study CCG-945. *Cancer.* Sep 15 2003;98(6):1243-1252.
- [40] Ferner RE. Neurofibromatosis 1 and neurofibromatosis 2: a twenty first century perspective. *Lancet Neurol.* Apr 2007;6(4):340-351.
- [41] Fletcher WA, Imes RK, Hoyt WF. Chiasmal gliomas: appearance and long-term changes demonstrated by computerized tomography. *J Neurosurg.* 1986;65:154-159.
- [42] Moreno L, Bautista F, Ashley S, Duncan C, Zacharoulis S. Does chemotherapy affect the visual outcome in children with optic pathway glioma? A systematic review of the evidence. *Eur J Cancer.* Aug 2010;46(12):2253-2259.
- [43] Ater JL, Zhou T, Holmes E, et al. Randomized study of two chemotherapy regimens for treatment of low-grade glioma in young children: a report from the Children's Oncology Group. *J Clin Oncol.* Jul 20 2012;30(21):2641-2647.
- [44] Ater J, Holmes E, Zhou T, et al. Abstracts from the thirteenth international symposium on pediatric neuro-oncology: Results of COG protocol A9952- a randomized phase 3 study of two chemotherapy regimens for incompletely resected low-grade glioma in young children. *Neuro-Oncology.* 2008;10:451.

- [45] Sievert AJ, Fisher MJ. Pediatric low-grade gliomas. *J Child Neurol*. Nov 2009;24(11):1397-1408.
- [46] Massimino M, Spreafico F, Cefalo G, et al. High response rate to cisplatin/etoposide regimen in childhood low-grade glioma. *J Clin Oncol*. Oct 15 2002;20(20):4209-4216.
- [47] Packer RJ, Jakacki R, Horn M, et al. Objective response of multiply recurrent low-grade gliomas to bevacizumab and irinotecan. *Pediatr Blood Cancer*. Jul 2009;52(7):791-795.
- [48] Bouffet E, Jakacki R, Goldman S, et al. Phase II study of weekly vinblastine in recurrent or refractory pediatric low-grade glioma. *J Clin Oncol*. Apr 20 2012;30(12):1358-1363.
- [49] Gururangan S, Fisher MJ, Allen JC, et al. Temozolomide in children with progressive low-grade glioma. *Neuro Oncol*. Apr 2007;9(2):161-168.
- [50] <http://www.clinicaltrials.gov/ct2/results?term=optic+pathway+glioma&Search=Search>. Accessed 10/12/2012, 2012.
- [51] Horwich A, Bloom HJG. Optic gliomas: radiation therapy and prognosis. *Int J Radiat Oncol Biol Phys*. 1985;11:1067-1079.
- [52] Grabenbauer GG, Schuchardt U, Buchfelder M, et al. Radiation therapy of optico-hypothalamic gliomas (OHG)--radiographic response, vision and late toxicity. *Radiother Oncol*. Mar 2000;54(3):239-245.
- [53] Jenkin D, Angyalfi S, Becker L, et al. Optic glioma in children: surveillance, resection, or irradiation? *Int J Radiat Oncol Biol Phys*. 1993;25:215-225.
- [54] Cappelli C, Grill J, Raquin M, et al. Long-term follow up of 69 patients treated for optic pathway tumours before the chemotherapy era. *Arch Dis Child*. 1998;79(4):334-338.
- [55] Sharif S, Ferner R, Birch JM, et al. Second primary tumors in neurofibromatosis 1 patients treated for optic glioma: substantial risks after radiotherapy. *J Clin Oncol*. Jun 1 2006;24(16):2570-2575.
- [56] Grill J, Couanet D, Cappelli C, et al. Radiation-induced cerebral vasculopathy in children with neurofibromatosis and optic pathway glioma. *Annals of Neurology*. 1999;45(3):393-396.
- [57] Kestle JR, Hoffman HJ, Mock AR. Moyamoya phenomenon after radiation for optic glioma. *J Neurosurg*. 1993;79:32-35.
- [58] Ullrich NJ, Robertson R, Kinnamon DD, et al. Moyamoya following cranial irradiation for primary brain tumors in children. *Neurology*. Mar 20 2007;68(12):932-938.
- [59] Pierce SM, Barnes PD, Loeffler JS, McGinn C, Tarbell NJ. Definitive radiation therapy in the management of symptomatic patients with optic glioma. Survival and long-term effects. *Cancer*. Jan 1 1990;65(1):45-52.
- [60] Lacaze E, Kieffer V, Streri A, et al. Neuropsychological outcome in children with optic pathway tumours when first-line treatment is chemotherapy. *Br J Cancer*. Dec 1 2003;89(11):2038-2044.
- [61] Packer RJ, Lange B, Ater J, et al. Carboplatin and vincristine for recurrent and newly diagnosed low-grade gliomas of childhood. *J Clin Oncol*. May 1993;11(5):850-856.
- [62] Prados MD, Edwards MS, Rabbitt J, Lamborn K, Davis RL, Levin VA. Treatment of pediatric low-grade gliomas with a nitrosourea-based multiagent chemotherapy regimen. *J Neurooncol*. May 1997;32(3):235-241.
- [63] Lafay-Cousin L, Holm S, Qaddoumi I, et al. Weekly vinblastine in pediatric low-grade glioma patients with carboplatin allergic reaction. *Cancer*. Jun 15 2005;103(12):2636-2642.

- [64] Grill J, Laithier V, Rodriguez D, Raquin M-A, Pierre-Kahn A, Kalifa C. When do children with optic pathway tumours need treatment? An oncological perspective in 106 patients treated in a single centre. *Eur J Pediatr*. 2000;159(9):692-696.
- [65] Blazo MA, Lewis RA, Chintagumpala MM, Frazier M, McCluggage C, Plon SE. Outcomes of systematic screening for optic pathway tumors in children with Neurofibromatosis Type 1. *Am J Med Genet A*. Jun 15 2004;127A(3):224-229.
- [66] Campagna M, Opocher E, Viscardi E, et al. Optic pathway glioma: long-term visual outcome in children without neurofibromatosis type-1. *Pediatr Blood Cancer*. Dec 1 2010;55(6):1083-1088.
- [67] Shofty B, Ben-Sira L, Freedman S, et al. Visual outcome following chemotherapy for progressive optic pathway gliomas. *Pediatr Blood Cancer*. Sep 2011;57(3):481-485.
- [68] Rahi JS, Cumberland PM, Peckham CS. Visual function in working-age adults: early life influences and associations with health and social outcomes. *Ophthalmology*. Oct 2009;116(10):1866-1871.
- [69] Swanson MW, McGwin G. Visual impairment and functional status from the 1995 National Health Interview Survey on Disability. *Ophthalmic Epidemiol*. Jul 2004;11(3):227-239.
- [70] Hyman SL, Shores A, North KN. The nature and frequency of cognitive deficits in children with neurofibromatosis type 1. *Neurology*. Oct 11 2005;65(7):1037-1044.
- [71] Pilling RF, Lloyd IC, Huson S. Utility of optic pathway glioma screening in young children with neurofibromatosis type I: questions generated by a clinical audit. *Eye (Lond)*. Oct 2010;24(10):1603-1605.
- [72] McDonald MA, Dobson V, Sebris SL, Baitch L, Varner D, Teller DY. The acuity card procedure: a rapid test of infant acuity. *Invest Ophthalmol Vis Sci*. Aug 1985;26(8):1158-1162.
- [73] Dobson V, Quinn GE, Tung B, Palmer EA, Reynolds JD. Comparison of recognition and grating acuities in very-low-birth-weight children with and without retinal residua of retinopathy of prematurity. Cryotherapy for Retinopathy of Prematurity Cooperative Group. *Invest Ophthalmol Vis Sci*. Mar 1995;36(3):692-702.
- [74] Quinn GE, Berlin JA, James M. The Teller acuity card procedure. Three testers in a clinical setting. *Ophthalmology*. 1993;100:488-494.
- [75] Murakami R, Hirai T, Sugahara T, et al. Grading astrocytic tumors by using apparent diffusion coefficient parameters: superiority of a one- versus two-parameter pilot method. *Radiology*. Jun 2009;251(3):838-845.
- [76] Alvarez-Linera J, Benito-Leon J, Escribano J, Rey G. Predicting the histopathological grade of cerebral gliomas using high b value MR DW imaging at 3-tesla. *J Neuroimaging*. Jul 2008;18(3):276-281.
- [77] Pauliah M, Saxena V, Haris M, Husain N, Rathore RK, Gupta RK. Improved T(1)-weighted dynamic contrast-enhanced MRI to probe microvasculature and heterogeneity of human glioma. *Magn Reson Imaging*. Nov 2007;25(9):1292-1299.
- [78] Chenevert TL, McKeever PE, Ross BD. Monitoring early response of experimental brain tumors to therapy using diffusion magnetic resonance imaging. *Clin Cancer Res*. Sep 1997;3(9):1457-1466.
- [79] Moffat BA, Chenevert TL, Lawrence TS, et al. Functional diffusion map: a noninvasive MRI biomarker for early stratification of clinical brain tumor response. *Proc Natl Acad Sci U S A*. Apr 12 2005;102(15):5524-5529.

- [80] Lope LA, Hutcheson KA, Khademian ZP. Magnetic resonance imaging in the analysis of pediatric orbital tumors: utility of diffusion-weighted imaging. *J AAPOS*. Jun 2010;14(3):257-262.
- [81] Sener RN. Diffusion MRI in neurofibromatosis type 1: ADC evaluations of the optic pathways, and a comparison with normal individuals. *Comput Med Imaging Graph*. Mar-Apr 2002;26(2):59-64.
- [82] Jost SC, Ackerman JW, Garbow JR, Manwaring LP, Gutmann DH, McKinstry RC. Diffusion-weighted and dynamic contrast-enhanced imaging as markers of clinical behavior in children with optic pathway glioma. *Pediatr Radiol*. Dec 2008;38(12):1293-1299.
- [83] Le Bihan D, Mangin JF, Poupon C, et al. Diffusion tensor imaging: concepts and applications. *J Magn Reson Imaging*. Apr 2001;13(4):534-546.
- [84] Tummala RP, Chu RM, Liu H, Truwit CL, Hall WA. Application of diffusion tensor imaging to magnetic-resonance-guided brain tumor resection. *Pediatr Neurosurg*. Jul 2003;39(1):39-43.
- [85] Lober RM, Guzman R, Cheshier SH, Fredrick DR, Edwards MS, Yeom KW. Application of diffusion tensor tractography in pediatric optic pathway glioma. *J Neurosurg Pediatr*. Aug 17 2012.
- [86] Lee HY, Na DG, Song IC, et al. Diffusion-tensor imaging for glioma grading at 3-T magnetic resonance imaging: analysis of fractional anisotropy and mean diffusivity. *J Comput Assist Tomogr*. Mar-Apr 2008;32(2):298-303.
- [87] Inoue T, Ogasawara K, Beppu T, Ogawa A, Kabasawa H. Diffusion tensor imaging for preoperative evaluation of tumor grade in gliomas. *Clin Neurol Neurosurg*. Apr 2005;107(3):174-180.
- [88] Goebell E, Paustenbach S, Vaeterlein O, et al. Low-grade and anaplastic gliomas: differences in architecture evaluated with diffusion-tensor MR imaging. *Radiology*. Apr 2006;239(1):217-222.
- [89] Hegedus B, Hughes FW, Garbow JR, et al. Optic nerve dysfunction in a mouse model of neurofibromatosis-1 optic glioma. *J Neuropathol Exp Neurol*. May 2009;68(5):542-551.
- [90] Hickman SJ. Optic nerve imaging in multiple sclerosis. *J Neuroimaging*. Apr 2007;17 Suppl 1:42S-45S.
- [91] Kolbe S, Chapman C, Nguyen T, et al. Optic nerve diffusion changes and atrophy jointly predict visual dysfunction after optic neuritis. *Neuroimage*. Apr 15 2009;45(3):679-686.
- [92] Vinogradov E, Degenhardt A, Smith D, et al. High-resolution anatomic, diffusion tensor, and magnetization transfer magnetic resonance imaging of the optic chiasm at 3T. *J Magn Reson Imaging*. Aug 2005;22(2):302-306.
- [93] Roebroek A, Galuske R, Formisano E, et al. High-resolution diffusion tensor imaging and tractography of the human optic chiasm at 9.4 T. *Neuroimage*. Jan 1 2008;39(1):157-168.
- [94] Yamada M, Momoshima S, Masutani Y, et al. Diffusion-tensor neuronal fiber tractography and manganese-enhanced MR imaging of primate visual pathway in the common marmoset: preliminary results. *Radiology*. Dec 2008;249(3):855-864.
- [95] Nickerson JP, Salmela MB, Koski CJ, Andrews T, Filippi CG. Diffusion tensor imaging of the pediatric optic nerve: intrinsic and extrinsic pathology compared to normal controls. *J Magn Reson Imaging*. Jul 2010;32(1):76-81.

- [96] Filippi CG, Bos A, Nickerson JP, Salmela MB, Koski CJ, Cauley KA. Magnetic resonance diffusion tensor imaging (MRDTI) of the optic nerve and optic radiations at 3T in children with neurofibromatosis type I (NF-1). *Pediatr Radiol*. Feb 2012;42(2):168-174.
- [97] Beatty RM, Sadun AA, Smith L, Vonsattel JP, Richardson EP, Jr. Direct demonstration of transsynaptic degeneration in the human visual system: a comparison of retrograde and anterograde changes. *J Neurol Neurosurg Psychiatry*. Feb 1982;45(2):143-146.
- [98] Ferner RE, Lucas JD, O'Doherty MJ, et al. Evaluation of (18)fluorodeoxyglucose positron emission tomography ((18)FDG PET) in the detection of malignant peripheral nerve sheath tumours arising from within plexiform neurofibromas in neurofibromatosis 1. *J Neurol Neurosurg Psychiatry*. Mar 2000;68(3):353-357.
- [99] Ferner RE, Golding JF, Smith M, et al. [18F]2-fluoro-2-deoxy-D-glucose positron emission tomography (FDG PET) as a diagnostic tool for neurofibromatosis 1 (NF1) associated malignant peripheral nerve sheath tumours (MPNSTs): a long-term clinical study. *Ann Oncol*. Feb 2008;19(2):390-394.
- [100] Cardona S, Schwarzbach M, Hinz U, et al. Evaluation of F18-deoxyglucose positron emission tomography (FDG-PET) to assess the nature of neurogenic tumours. *Eur J Surg Oncol*. Aug 2003;29(6):536-541.
- [101] Bredella MA, Torriani M, Hornicek F, et al. Value of PET in the assessment of patients with neurofibromatosis type 1. *AJR Am J Roentgenol*. Oct 2007;189(4):928-935.
- [102] Fisher MJ, Phillips PC. PET imaging of brain tumors. In: M C, ed. *Practical Pediatric PET Imaging*. New York: Springer Science+Business Media, Inc.; 2006:175-219.
- [103] Wegner EA, Barrington SF, Kingston JE, et al. The impact of PET scanning on management of paediatric oncology patients. *Eur J Nucl Med Mol Imaging*. Jan 2005;32(1):23-30.
- [104] Miyamoto J, Sasajima H, Owada K, Mineura K. Surgical decision for adult optic glioma based on [18F]fluorodeoxyglucose positron emission tomography study. *Neurol Med Chir (Tokyo)*. Oct 2006;46(10):500-503.
- [105] Molloy P, Defeo R, Hunter J, et al. Excellent correlation of FDG-PET imaging with clinical outcome in patients with neurofibromatosis type I and low grade astrocytomas. *J Nucl Med*. 1999;40:129P
- [106] Moharir M, London K, Howman-Giles R, North K. Utility of positron emission tomography for tumour surveillance in children with neurofibromatosis type 1. *Eur J Nucl Med Mol Imaging*. Jul 2010;37(7):1309-1317.
- [107] Peng F, Juhasz C, Bhambhani K, Wu D, Chugani DC, Chugani HT. Assessment of progression and treatment response of optic pathway glioma with positron emission tomography using alpha-[(11)C]methyl-L-tryptophan. *Mol Imaging Biol*. May-Jun 2007;9(3):106-109.
- [108] Roselli F, Pisciotta NM, Aniello MS, et al. Brain F-18 Fluorocholine PET/CT for the assessment of optic pathway glioma in neurofibromatosis-1. *Clin Nucl Med*. Oct 2010;35(10):838-839.
- [109] Juhasz C, Chugani DC, Muzik O, et al. In vivo uptake and metabolism of alpha-[11C]methyl-L-tryptophan in human brain tumors. *J Cereb Blood Flow Metab*. Mar 2006;26(3):345-357.
- [110] Hara T. 18F-fluorocholine: a new oncologic PET tracer. *J Nucl Med*. Dec 2001;42(12):1815-1817.

- [111] Pirotte B, Acerbi F, Lubansu A, Goldman S, Brotchi J, Levivier M. PET imaging in the surgical management of pediatric brain tumors. *Childs Nerv Syst.* Jul 2007;23(7):739-751.
- [112] Fisher M. The Use of PET in the Evaluation of Tumors in Children with Neurofibromatosis Type 1. *PET Clinics JA* -. Vol 32008:531-549 PB -.
- [113] Chao RC, Pyzel U, Fridlyand J, et al. Therapy-induced malignant neoplasms in Nf1 mutant mice. *Cancer Cell.* Oct 2005;8(4):337-348.
- [114] Cohen ME, Duffner PK. Visual-evoked responses in children with optic gliomas, with and without neurofibromatosis. *Childs Brain.* 1983;10(2):99-111.
- [115] North K, Cochineas C, Tang E, Fagan E. Optic gliomas in neurofibromatosis type 1: role of visual evoked potentials. *Pediatr Neurol.* 1994;10:117-123.
- [116] Jabbari B, Maitland CG, Morris LM, Morales J, Gunderson CH. The value of visual evoked potential as a screening test in neurofibromatosis. *Arch Neurol.* 1985;42:1072-1074.
- [117] Lund AM, Skovby F. Optic gliomas in children with neurofibromatosis type 1. *Eur J Pediatr.* Oct 1991;150(12):835-838.
- [118] Rossi LN, Pastorino G, Scotti G, et al. Early diagnosis of optic glioma in children with neurofibromatosis type 1. *Childs Nerv Syst.* 1994;10:426-429.
- [119] Groswasser Z, Kriss A, Halliday AM, McDonald WI. Pattern- and flash-evoked potentials in the assessment and management of optic nerve gliomas. *Journal of Neurology, Neurosurgery, and Psychiatry.* 1985;48:1125-1134.
- [120] Chang BC, Mirabella G, Yagev R, et al. Screening and diagnosis of optic pathway gliomas in children with neurofibromatosis type 1 by using sweep visual evoked potentials. *Invest Ophthalmol Vis Sci.* Jun 2007;48(6):2895-2902.
- [121] Liu GT, Molloy PT, Needle M, Phillips P. Optic gliomas in neurofibromatosis type 1: role of visual evoked potentials [letter]. *Pediatr Neurol.* 1995;12:89-90.
- [122] Ng YT, North KN. Visual-evoked potentials in the assessment of optic gliomas. *Pediatr Neurol.* Jan 2001;24(1):44-48.
- [123] Trisciuzzi MT, Riccardi R, Piccardi M, et al. A fast visual evoked potential method for functional assessment and follow-up of childhood optic gliomas. *Clin Neurophysiol.* Jan 2004;115(1):217-226.
- [124] Kelly JP, Weiss AH. Comparison of pattern visual-evoked potentials to perimetry in the detection of visual loss in children with optic pathway gliomas. *J Aapos.* Aug 2006;10(4):298-306.
- [125] Iannaccone A, McCluney RA, Brewer VR, et al. Visual evoked potentials in children with neurofibromatosis type 1. *Doc Ophthalmol.* Jul 2002;105(1):63-81.
- [126] Wolsey DH, Larson SA, Creel D, Hoffman R. Can screening for optic nerve gliomas in patients with neurofibromatosis type I be performed with visual-evoked potential testing? *J Aapos.* Aug 2006;10(4):307-311.
- [127] Falsini B, Ziccardi L, Lazzareschi I, et al. Longitudinal assessment of childhood optic gliomas: relationship between flicker visual evoked potentials and magnetic resonance imaging findings. *J Neurooncol.* May 2008;88(1):87-96.
- [128] Kelly JP, Leary S, Khanna P, Weiss AH. Longitudinal measures of visual function, tumor volume, and prediction of visual outcomes after treatment of optic pathway gliomas. *Ophthalmology.* Jun 2012;119(6):1231-1237.

- [129] Ko TH, Fujimoto JG, Schuman JS, et al. Comparison of ultrahigh- and standard-resolution optical coherence tomography for imaging macular pathology. *Ophthalmology*. Nov 2005;112(11):1922 e1921-1915.
- [130] Danesh-Meyer HV, Carroll SC, Foroozan R, et al. Relationship between retinal nerve fiber layer and visual field sensitivity as measured by optical coherence tomography in chiasmal compression. *Invest Ophthalmol Vis Sci*. Nov 2006;47(11):4827-4835.
- [131] Danesh-Meyer HV, Papchenko T, Savino PJ, Law A, Evans J, Gamble GD. In vivo retinal nerve fiber layer thickness measured by optical coherence tomography predicts visual recovery after surgery for parachiasmal tumors. *Invest Ophthalmol Vis Sci*. May 2008;49(5):1879-1885.
- [132] Cheng H, Laron M, Schiffman JS, Tang RA, Frishman LJ. The relationship between visual field and retinal nerve fiber layer measurements in patients with multiple sclerosis. *Invest Ophthalmol Vis Sci*. Dec 2007;48(12):5798-5805.
- [133] Costello F, Hodge W, Pan YI, Eggenberger E, Coupland S, Kardon RH. Tracking retinal nerve fiber layer loss after optic neuritis: a prospective study using optical coherence tomography. *Mult Scler*. Aug 2008;14(7):893-905.
- [134] Talman LS, Bisker ER, Sackel DJ, et al. Longitudinal study of vision and retinal nerve fiber layer thickness in multiple sclerosis. *Ann Neurol*. Jun 2010;67(6):749-760.
- [135] Frohman E, Costello F, Zivadinov R, et al. Optical coherence tomography in multiple sclerosis. *Lancet Neurol*. Oct 2006;5(10):853-863.
- [136] Phillips PH. Is optical coherence tomography indicated for the evaluation of patients with neurofibromatosis type 1? *J AAPOS*. Dec 2010;14(6):467-468.
- [137] Chang L, El-Dairi MA, Frempong TA, et al. Optical coherence tomography in the evaluation of neurofibromatosis type-1 subjects with optic pathway gliomas. *J AAPOS*. Dec 2010;14(6):511-517.
- [138] de Blank PMK, Berman JI, Liu GT, et al. Fractional anisotropy of the optic radiations is associated with visual acuity loss in optic pathway gliomas of neurofibromatosis type 1. *Neuro-Oncology*. In press; doi: 10.1093/neuonc/not068.

Chapter 111

NEUROFIBROMATOSIS TYPE 1 BONE DISEASE: DIAGNOSIS AND MANAGEMENT

Elizabeth K. Schorry¹ and Alvin H. Crawford²

¹Division of Human Genetics, Children's

Hospital Medical Center, Cincinnati, OH, US

²Department of Pediatric Orthopaedics, Children's

Hospital Medical Center, Cincinnati, OH, US

ABSTRACT

Osseous abnormalities occur frequently in NF1, and can present in a multitude of ways. Bone disorders in NF1 can be categorized as generalized or focal. Generalized bone involvement occurs frequently to a mild degree, and can include short stature, macrocephaly, and osteopenia. Focal, dystrophic features such as tibial dysplasia/pseudarthrosis and dystrophic scoliosis are challenging to manage effectively. Additional features can include dural ectasia of the spine, sphenoid wing dysplasia, leg length inequality, pectus deformities of the chest, non-ossifying fibromas of long bones, and ossifying subperiosteal hematoma. Early recognition of musculoskeletal complications is important in improving outcome.

INTRODUCTION

There is increasing evidence that a primary skeletal dysplasia exists in NF1. This concept was introduced by Dr. Vincent Riccardi in 1986 [1], and has received revived attention in more recent years. Osseous abnormalities occur in up to 38% of patients with NF1 [2], and can usually be categorized as either focal or generalized. Generalized skeletal manifestations, such as osteopenia, osteoporosis, and short stature are common in NF1 but usually mild [3]. Focal skeletal abnormalities include dystrophic scoliosis, sphenoid wing dysplasia, and tibial dysplasia/pseudarthrosis. These focal deficits are less common than the generalized ones, but can carry significant morbidity and treatment challenges.

Pectus deformities of the chest wall are also well known to be associated with NF1 and may represent an area of overlap between focal and generalized osseous manifestations.

The pathophysiology of bone manifestations in NF1 is not yet fully elucidated. Although some bone abnormalities may be secondary to presence of tumors and their effects on Nf1 haploinsufficient bone, others are more likely the result of a primary osseous dysplasia. The study of animal models and human tissues has provided significant insight into bone biology in NF1. While neurofibromin's role as a tumor suppressor is well known, it is becoming apparent that neurofibromin also plays a significant role in bone homeostasis, remodeling, and fracture healing.

Mouse models have demonstrated that neurofibromin plays a role in growth and differentiation of both osteoblasts and osteoclasts with decreased differentiation of osteoblasts and increased activity of osteoclasts seen with haploinsufficiency of neurofibromin [4]. These changes likely contribute to the bone dysplasia and abnormal healing of bone seen in NF1. Some of the focal complications of NF1, particularly tibial pseudarthrosis, have been found to involve bi-allelic inactivation of Nf1 in affected tissue [5, 6] and it is possible that other focal osseous complications may involve a similar second hit event.

SPINAL MANIFESTATIONS : SCOLIOSIS

Spinal deformity is the most common of the focal osseous complications of NF1, occurring in 10–30% of patients in NF clinic populations [7, 8]. In general NF clinic patients, the prevalence of scoliosis is about 15%; but a prevalence over 25% can be seen in clinics biased towards specialized orthopedic care (Schorry and Crawford, unpublished data). Both males and females with NF1 are equally likely to be affected with scoliosis, in contrast to the non-NF population where idiopathic scoliosis is much more common in females. Two distinct types of scoliosis curves can occur in NF1: a non-dystrophic curve similar to that seen in idiopathic scoliosis (Figure 1); and a dystrophic curve, which is highly likely to progress.



Figure 1. Non-dystrophic scoliosis. This 13 year old child presented with a 45 degree curve. No dystrophic features are present.



Figure 2. Dystrophic scoliosis. AP and lateral views of a 12 year old child with dystrophic scoliosis. This is a 60 degree focal left thoracic scoliosis. The hazy density above the apex of the curve seen on both the AP and lateral views was identified on MRI as a paraspinal plexiform neurofibroma.

Dystrophic curves (Figure 2) are characterized by a short-segmented, sharply angulated deformity, usually involving 6 or fewer vertebrae [9]. Additional dystrophic features on radiographs (Figure 3, 4) can be predictive of progression and need for future surgery and include: posterior and anterior vertebral scalloping, vertebral rotation, vertebral body wedging, rib “pencil” or rotation, spindling of transverse processes, widened interpediculate distance and enlarged intervertebral foramina [10].

Some of the dystrophic bony features are closely related to paraspinal or regional plexiform neurofibromas (Figure 5), and appear to be secondary to erosion by tumor growth; however, others can occur in the total absence of nearby tumors. Spinal deformities in NF1 can involve all areas of the vertebral column with thoracic and lumbar areas most commonly affected.

Cervical spine deformities, although less common, can result in cervical spinal instability and resultant neurologic compromise [2]. Kyphosis is commonly associated with the scoliotic curve. Spondylolisthesis, a pathologic subluxation of the lumbar vertebrae, is a rare complication. In a series of 131 patients with NF1 and scoliosis from Cincinnati Children’s Hospital, mean age at diagnosis of scoliosis was 9 years with 15% of patients having onset before age 7 years. Thirty-five (35)% required surgical correction, primarily anterior and posterior spinal fusion, and 63% had evidence of paraspinal tumors on imaging.

Need for surgery was equally distributed between males and females. Radiographic features highly predictive of subsequent need for surgery included vertebral scalloping, dural ectasia, presence of short-segmented focal curve and paraspinal tumors. However, presence of tumors was not a requirement for development of a severe dystrophic scoliotic curve (Schorry et al., manuscript in review).

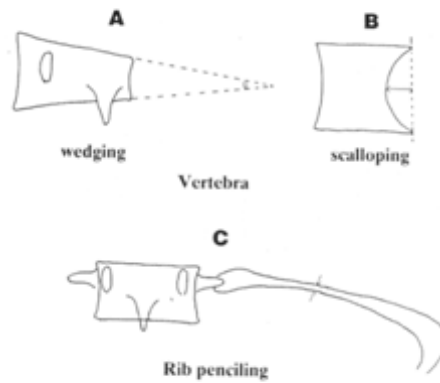


Figure 3. Dystrophic elements. The predominant features of dystrophic vertebral bodies are: wedging in the frontal plane, scalloping or indentation of the cortex of greater than 3mm in the thoracic spine and greater than 4 mm in lumbar spine. Rib penciling is defined as the proximal third of the rib being thinner than that of the 2nd rib.



Figure 4. Proximal thoracic scoliosis. 3D reconstruction of the characteristic focal short segmented, sharply angulated proximal thoracic scoliosis, so often associated with dystrophic scoliosis.

Management of Scoliosis

Management of non-dystrophic curves should be similar to that used for idiopathic scoliosis. Generally, bracing should be considered for patients with curves between 25–40°; posterior spinal fusion for curves greater than 40°, and combined anterior and posterior fusion for curves greater than 55–60°. Dystrophic curves in NF1 represent a challenge for the spine surgeon. They are characterized by a rapidly progressive course, a tendency to evolve to a severe deformity, and a higher rate of pseudarthrosis after surgery.

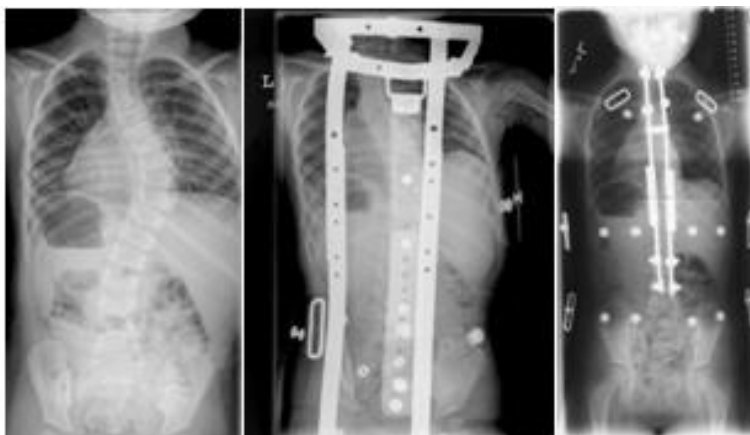


Figure 5. Growing rods. Early onset scoliosis treated by Milwaukee brace and subsequent growing rods. The patient underwent a Risser case correction when the curve reached 50 degrees, and was corrected to 35 degrees and placed into a Milwaukee brace which subsequently failed. She was then treated by insertion of growing rods.

Typically, the goal of surgical management is to arrest the progression of deformity, rather than achieving complete correction. Almost all authors agree on an early and aggressive approach to dystrophic curves in NF1. Bracing has not been effective and is not recommended in any dystrophic deformity [11]. Dystrophic curves less than 20° should be monitored at 6 month intervals. For patients with curves between 20° and 40° , posterior spinal fusion is recommended, with use of autogenous iliac crest graft to minimize the occurrence of pseudarthrosis. For dystrophic curves greater than 40° or with kyphosis of greater than 50° , combined anterior and posterior spinal arthrodesis is recommended.

For very young patients less than 9 years of age with significant curve, placement of a growing rod (Figure 5) is an option. This technique requires fusing the cephalic and caudal anchors of the spinal rods which span the curvature and act as an internal brace. There is a tandem connector which allows for “growing”, lengthening the rods until skeletal maturity begins. Periodic lengthening of the spinal implants is required, usually at about 6 month intervals, followed by eventual definitive spinal fusion. It is a taxing process and although the child does not require a brace, the multiple procedures and potential and real complications cause it to be a less than ideal resolution of this problem. While trunk length may be of concern when performing a spinal fusion in young patients, the risk of plexiform tumors or dural ectasia eroding the bony vertebra justifies anterior and posterior fusion in severe dystrophic curves at a somewhat earlier age.

There is occasional modulation of a non-dystrophic curve to dystrophic, at which time it acquires the previously mentioned dystrophic characteristics. Data has shown that subtle dystrophic characteristics may have been present in this group of patients early on but could not be identified on plain x-rays [10].

Anterior and posterior spinal fusion is recommended for all dystrophic curves and every effort made to achieve this. On occasion the presence of plexiform neurofibromas or dural ectasia has eroded the vertebral body bone mass and pedicles to the point of not having sufficient fixation points for the instrumentation. In general, resection of paraspinal tumors encroaching on the spinal canal is recommended at the same time as spinal fusion, as tumor

resection without fusion can result in an unstable spine [9]. Complete resection of extensive paraspinal neurofibromas is generally neither feasible nor recommended.

After obtaining parental consent, we have occasionally used the growth factor, bone morphogenetic protein (BMP), off-label as an adjunct to achieve fusion. Postoperative brace-assisted immobilization is recommended after spinal fusion until a fusion mass is noted on imaging, in an effort to prevent pseudarthrosis.

Even though the strength of titanium, cobalt chromium and stainless steel is considerable, a failure of fusion of the vertebral bony mass will result in pseudarthrosis and ultimate rod breakage. Perioperative surgical complications include intraoperative hemorrhage, neuromonitoring alerts with or without paralysis.

Pseudarthrosis may occur but is usually noted after one to two years and more often by rod breakage which may be extremely subtle and not apparent to the patient. X-rays may reveal a progression of the curve or a break in the rods. It is important to revise and re-instrument the spinal fusion if pseudarthrosis occurs, to stabilize the deformity and prevent further progression of the scoliosis curve.

DURAL ECTASIA

Dural ectasia, a widening of the dural sac surrounding the spinal cord, can contribute to spinal pathology in NF1 (Figure 6). Dural ectasia also occurs in Marfan syndrome and associated connective tissue disorders. NF1-related dural ectasia may be observed as an independent finding or related to para-axial or intraspinal tumors. In a review of 56 patients with NF1 and dystrophic scoliosis, 38% had evidence of dural ectasia on spinal MRI, and presence of dural ectasia was a strong predictor of need for future spine surgery [Schorry et al, unpublished data]. [3].

Dural ectasia can be progressive, can result in destabilization of the spine, or can be associated with meningoceles (protrusion of the spinal meninges through a neuroforamen) or erosion of the vertebral bodies [9].

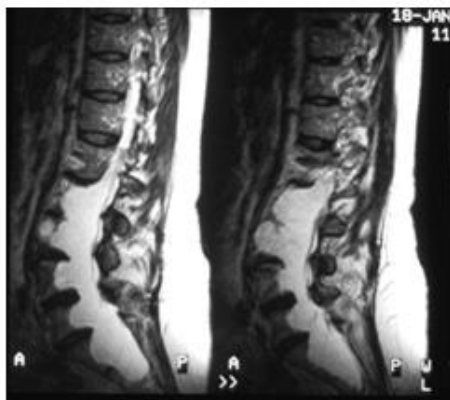


Figure 6. Dural ectasia. Dural ectasia is an expansion of the spinal canal secondary to increased hydrostatic pressure generated within the dural sleeve. It has been shown to cause severe scalloping and erosion of the vertebral bodies. Occasionally it will exit the spinal canal through the neuroforamina and cause thoracic and lumbar meningoceles.

LONG BONE DYSPLASIA

Long bone dysplasia affects between 2–5% of individuals with NF1 [12], with the most common bones affected being the tibia and fibula. Less commonly affected bones have included ulna and clavicle. Despite being a rare complication of NF1, tibial dysplasia (TD) can result in significant morbidity. Tibial dysplasia occurs in a spectrum ranging from anterolateral tibial bowing to frank pseudarthrosis with non-union of bone after fracture (Figure 7). Natural history studies have shown that tibial dysplasia in NF1 is usually evident in the first year of life; affects males more frequently than females; and is almost always unilateral (although rare bilateral cases have been reported) [13]. Typically, a patient will present with anterolateral bowing of the tibia and/or fibula in infancy. Some patients may not progress to fracture, particularly if they are braced continuously from a young age. However, up to 66% of patients with tibial dysplasia will progress to fracture and pseudarthrosis, at an average age of 4.6 years based on a natural history study [13]. Once fractured, successful surgical management is difficult, requiring multiple surgical procedures and prolonged bracing to achieve a good result. Historically, up to one-third of patients have required eventual amputation of the affected limb.

Management of Tibial Dysplasia

Bracing plays a major role in management of tibial dysplasia which has not yet fractured. Young infants with tibial dysplasia should be placed in an ankle-foot orthosis, with conversion to a knee-ankle-foot orthosis as soon as the child begins to bear weight. The brace should be worn as close as possible to 24 hours a day, with one half hour allowance for personal hygiene time, until skeletal maturity is achieved at adolescence [2]. Bracing appears to be effective in preventing fracture and pseudarthrosis in up to 1/3 of patients; however, fractures can occur due to non-compliance with bracing. Corrective osteotomy before fracture has occurred is generally not recommended, due to the risk for development of pseudarthrosis. McFarland developed a bypass fibular graft procedure (Figure 8), which has been successful in some cases in reducing risk of fracture; however, many grafts fail.

For patients who have progressed from tibial bowing to frank pseudarthrosis, various surgical techniques have been attempted. Some surgeons have chosen to use external fixation compression/distraction procedures, such as the Ilizarov frame [14]. More commonly, surgeons have used intramedullary rodding (Figure 9), with an approach including: excision of the pseudarthrosis material; placement of an intramedullary rod; and placement of autogenous bone graft from iliac crest [15, 16].

However, many patients have historically required multiple surgeries to achieve union and the re-fracture rate may be as high as 50%. Many NF1 surgeons have now begun adding recombinant human bone morphogenetic protein (rhBMP-2) to the fracture site at surgery, as an anabolic agent. Results with addition of BMP-2 have shown over 70% healing with one surgical procedure, at a mean time of 6.4 months post-operatively [17]. The addition of bisphosphonates to BMP in NF1 animal models has shown encouraging results with further improvement in bone healing [18]; however, controlled studies have not yet been conducted in human NF1 patients.



Figure 7. Congenital tibial dysplasia and pseudarthrosis. Three of the four types of congenital tibial dysplasia. a) type one with increased density in the lower mid third with no fracture; b) Fracture of the lower third with minimal displacement; c) Fracture displacement of both the tibia and fibula with pointed “sucked candy” appearance of the ends of the bones.



Figure 8. McFarland procedure. This 9 year old child presented with bilateral severe tibial bowing and NF1. She had been treated with a lateral hemiepiphyseal arrest for her ankle bowing. At a subsequent visit a stress fracture was noted at the apex of the tibial bow. She subsequently underwent bilateral medial fibular allograft by-pass procedures with subsequent incorporation of the grafts and correction of the ankle bowing.

Designing controlled clinical trials has been challenging because of the small numbers of patients treated at any one center. Clearly, better management is still needed for this rare, but highly morbid complication of NF1.

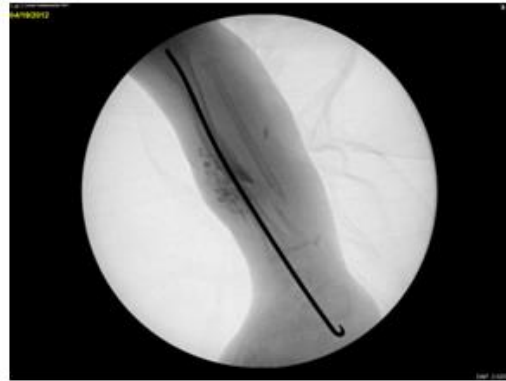


Figure 9. Intramedullary rod. This 2.5 year old female was noted at birth to have severe anterolateral bowing with subsequent fracture of tibia and fibula. She underwent a “Peter Williams” type procedure with open reduction, intramedullary rodding, bone grafting and rhBMP.

LEG LENGTH INEQUALITY/SEGMENTAL HYPERTROPHY

Leg length inequality occurs frequently in NF1; it may be related to disuse atrophy in patients with tibial dysplasia, or may occur as an isolated finding.

Segmental hypertrophy involving an entire limb and soft tissues can occur usually in association with a regional plexiform neurofibroma and may pose significant management challenges. The zones of bone and soft tissue overgrowth are usually unilateral and affect only one extremity. The osseous changes characteristically cause the bone to elongate with a wavy irregularity or thickening of the cortex. The associated plexiform neurofibromas are usually extensive involving multiple peripheral nerves and their roots in a regional plexus or the entire length of nerve. As such, resection is rarely an option.

Attempts to debulk soft tissue along with associated bone resection have not necessarily resulted in cosmetic improvement. For very large lesions, amputation and conversion to prosthetic fitting should be strongly considered. Early epiphyseal arrest of the involved bone or lengthening of the normal side has achieved a fair amount of success.

SPHENOID WING DYSPLASIA

Sphenoid wing dysplasia, one of the diagnostic criteria for NF1, usually manifests as unilateral erosive changes of the wing of the sphenoid bone of the orbit. It can be found on imaging in up to 11% of NF1 patients [12]. Most occurrences are associated with an orbital plexiform neurofibroma [19] and therefore may represent a secondary erosive effect rather than a primary osseous manifestation of NF1 [3]. Although no intervention is required in the vast majority of patients, rare cases with progressive orbital plexiform neurofibromas leading to pulsating enophthalmos have been reported [1].

PECTUS DEFORMITIES

Chest wall deformities, with varying degrees of pectus excavatum or pectus carinatum, have been reported in at least 30% of patients with NF1 when examined carefully [1]. Pectus excavatum of the lower portion of the sternum is most common, although carinatum and combined deformities can also occur.

Most mild deformities do not require surgery. Surgical repair may be chosen for cosmetic reasons or when indicated for cardiac or pulmonary decompensation. The management of pectus deformity has changed considerably from the aggressive thoracotomy procedures in the past. The Nuss procedure is a minimally invasive procedure performed by video-assisted thoracoscopy, where a metallic pre-bent plate is inserted anterior to the mediastinum and the retrodisplaced sternum is forced outward assuming its normal position [20].

GENERALIZED BONE DISORDERS

In addition to the focal, dysplastic osseous complications, more generalized bone disorders occur as well, including short stature, macrocephaly, and osteopenia. Many authors have reported stature in NF1 to be below that of age-matched peers. Studies of large populations with NF1 have shown a unimodal distribution of height shifted to the left, with 13% of patients having short stature defined as more than 2 standard deviations below the mean. Likewise, distribution of head circumference is shifted to the right, with 24% of NF1 individuals having macrocephaly [21].

OSTEOPENIA/OSTEOPOROSIS

Multiple studies have reported osteopenia or osteoporosis in up to 50% of children and adults with NF1, with mean bone mineral density (BMD) about 1 standard deviation below the mean of the general population [22, 23, 24, 25, 26]. However, the severity of osteopenia and osteoporosis in this population and its implications for bone health are still unclear. Although adults with NF1 and osteoporosis can be managed with standard anti-resorptive medications, their response to bisphosphonates may be less than that of the general population based on in vitro studies [27]. At present, there have been no studies on use of anti-resorptive medications in children with NF1. Conservative therapies, such as increased exercise and calcium and vitamin D supplementation, remain the first line of treatment for pediatric patients with NF1 and osteopenia [3].

FRACTURES

Fracture of dysplastic bone resulting in pseudarthrosis is a well-known complication of NF1 in a small subset of patients. However, despite the known reduction of BMD in individuals with NF1, limited information is available on whether this degree of bone mineral deficiency results in an increased rate of fracture of other bones, which do not appear clinically dysplastic.

Recent data have suggested a higher rate of fractures in adults with NF1. A survey study of 72 adults with NF1 in Hamburg, Germany showed the overall lifetime rate of ever having a fracture was 33%, compared to only 8% in sibling and spouse controls [28]. However, the very low fracture rate in controls raised questions of bias in the study methods. Another study involving NF1 patients of all ages reported a fracture rate of 52%, but did not have a comparison group [29].

A pediatric study of 256 children and adolescents with NF1, ages 5–20 years, found no difference in prevalence of patients with any lifetime fractures between the NF1 group without tibial dysplasia (22%) and unaffected sibling controls (25%); median number of fractures also did not differ [30]. Interestingly, there were significant differences in fracture location, with a higher frequency of fractures of the lower extremities in NF1 individuals without tibial dysplasia compared to controls. The authors concluded that children with NF1 do not appear to have a higher risk of fractures compared to non-NF1 children. It is possible that progressive osteopenia and osteoporosis do result in a higher fracture rate in older adults with NF1, but further data are needed.

BONE CYSTS; NON-OSSIFYING FIBROMAS

Benign lytic bone lesions, or non-ossifying fibromas, can be seen in the long bones in patients with NF1 (Figure 10). Riccardi and Eichner [1] noted cortical bone cysts in 10% of a series of 173 patients with NF1 who had long-bone radiographic studies performed. Non-ossifying fibromas have been reported in association with café-au-lait spots in the orthopedic literature since 1958 and called Jaffe-Campanacci syndrome; however, it is now thought that Jaffe-Campanacci syndrome is part of the variable manifestations of NF1 [31]. Non-ossifying fibromas also occur commonly in the pediatric population, and it is unclear if their incidence is increased in NF1. In the large majority of patients, these lesions are asymptomatic.

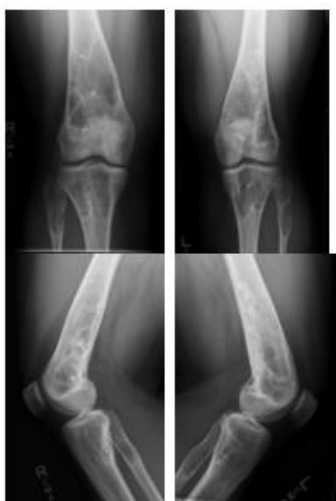


Figure 10. Multiple non-ossifying fibromas. Multiple non-ossifying fibromas in distal femur and proximal tibia in a teenager with NF1. This patient has not been restricted and followed for greater than 5 years without need for surgical intervention.

They are most frequently seen around the knee when x-rays have been obtained for sprains or strains. Most often both femurs and tibiae are involved. In a subset of patients, pathological fractures can occur [31], or microfracture may be a cause of pain.

We have generally been conservative managing these lesions, even when there are non-displaced pathological fractures. Most heal following treatment with a knee immobilizer or cast. In spite of the lucency seen radiographically, they rarely present as a displaced fracture. When seen, x-rays should be taken of both knees to confirm diagnosis.

There have been discussions questioning the benefit of percutaneous injection of bone substitutes, or treatment with bisphosphonates, but no studies have been documented. We recommend non-surgical measures for initial management of non-ossifying fibromas.

OTHERS: OSSIFYING SUBPERIOSTEAL HEMATOMA

Ossifying subperiosteal hematoma has been described as a rare osseous complication of NF1 (Figure 11). Most cases are thought to be initiated by minor fractures with subperiosteal bleeding, followed by osseous dysplasia of the hematoma [2]. The periosteum in NF1 has been described as less adherent to bone, thus predisposing it to periosteal hemorrhage. The ossifying subperiosteal hematoma occurs most frequently in the tibia and may present a clinical problem for the orthopaedic surgeon or primary care giver who does not treat this condition often.

A rapidly growing mass in a patient with minimal or no history of trauma can create concerns of infection or malignancy. The majority of cases have presented with swelling and erythema but without other signs of infection. Clinical and radiographic assessment, including MRI, will aid in differentiation of this condition from a tumor or infection.

It is important to be aware of these different findings in the physical exam and to avoid unnecessary aspiration of the lesion. The subperiosteal elevation noted in this condition is the key to differentiating between these three differential diagnoses. MRI will distinguish the subperiosteal hematoma by the rim reaction at the subperiosteal region and the fluid leveling.



Figure 11. Ossifying subperiosteal hematoma. This patient presented after minor trauma (a) with swelling, redness and pain. There is a density on the periphery of the swelling. The axial MRI (b) revealed the characteristic thick “orange rind” appearance of the periosteum but no evidence of infection. Ossification of the mass (c) and further ossification and trabeculation of the bone (d) were noted with maturity.

MUSCLE INVOLVEMENT IN NF1

Overview of musculoskeletal involvement in NF1 should include consideration of the muscle compartment as well. Motor developmental delays are well known to occur frequently in children with NF1, and have usually been attributed to CNS etiology. However, neurofibromin is strongly expressed in embryonic muscle cells [32] and its haploinsufficiency may also have a direct effect on muscle. Studies using peripheral quantitative CT scanning have shown a decrease in cross-sectional area of muscle in NF1 patients compared to controls [33]. Functional studies of muscle strength have shown children with NF1 to have a decrease in hand grip strength and lower force production in jumping when compared to controls [34; 35]. It is likely that muscle involvement in NF1 is multifactorial in etiology, with both primary and secondary effects.

CONCLUSION

A wide spectrum of osseous abnormalities can occur in individuals with NF1. Generalized abnormalities such as short stature, macrocephaly, and osteopenia, occur frequently but are usually mild. Decreased bone mineral density can lead to increased risk of fractures in adults with NF1, but does not appear to affect the pediatric population. Focal manifestations such as tibial dysplasia/pseudarthrosis and dystrophic scoliosis may be caused in part by bi-allelic inactivation of NF1 in localized tissues. Tibial dysplasia presents a challenge in management due to the risk of fracture and chronic non-union. Management with chronic bracing of non-fractured cases, and standard surgical protocols with intramedullary rodding and anabolic agents can result in improved outcome for those who have developed pseudarthrosis. Scoliosis in NF1 can be either dystrophic or non-dystrophic, presents at an early age, and affects girls and boys equally. Dystrophic curves have a high rate of progression; do not respond to bracing; and frequently require spinal fusion surgery. Current surgical recommendation for significant dystrophic curves is combined anterior and posterior spinal fusion, with immobilization post-operatively. Additional musculoskeletal features of NF1 include sphenoid wing dysplasia, dural ectasia, non-ossifying fibromas of long bones, leg length inequality, pectus deformities, ossifying subperiosteal hematomas, and decreased muscle strength. Early detection of skeletal complications and management in an orthopaedic center familiar with NF can improve outcome. Recent advances in understanding the biology of skeletal disease in NF1 raises the probability of additional biologic-based therapies for NF1 bone disease in the near future.

REFERENCES

- [1] Riccardi, V. M. and Eichner, J. E. *Neurofibromatosis: Phenotype, natural history, and pathogenesis*. Baltimore: Johns Hopkins University Press, 1986.
- [2] Crawford, A. C. and Schorry, E. K. 1999. Neurofibromatosis in children: The role of the orthopaedist. *J. Am. Acad. Orthop. Surg.* 7:217-230.
- [3] Eleftheriou, F., Kolanczyk, M., Schindeler, A., Viskochil, D., Hock, J., Schorry, E., Crawford, A., Friedman, J., Little, D., Peltonen, J., Carey, J., Feldman, D., Yu, X.,

- Armstrong, L., Birch, P., Kendler, D., Mundlos, S., Yang, F. C., Agiostratidou, G., Hunter-Schaedle, K., Stevenson, D. 2009. Skeletal abnormalities in neurofibromatosis type 1: Approaches to therapeutic options. *Am. J. Med. Genet.* 149A:2327–2338.
- [4] Wang, W., Nyman, J. S., Ono, K., Stevenson, D. A., Yang, X., Elefteriou, F. 2011. Mice lacking NF1 in osteochondroprogenitor cells display skeletal dysplasia similar to patients with neurofibromatosis type 1. *Hum. Molec. Gen.* 20(20):3910-3924.
- [5] Stevenson, D. A., Zhou, H., Ashrafi, S., Messiaen, L. M., Carey, J. C., D'Astous, J. L., Santora, S. D., Viskochil, D. H. 2006. Double inactivation of NF1 in tibial pseudarthrosis. *Am. J. Hum. Genet.* 79:143-148.
- [6] Lee, S. M., Choi, I. H., Lee, D. Y., Park, M. S., Yoo, W. J., Chung, C. Y., Cho, T. J. 2012. Is double inactivation of the NF1 gene responsible for the development of congenital pseudarthrosis of the tibia associated with NF1? *J. Orthoped. Res.* Epub. ahead of print.
- [7] Halmai, V., Doman, I., de Jone, T., Illes, T. 2002. Surgical treatment of spinal deformities associated with neurofibromatosis type 1. Report of 12 cases. *J. Neurosurg.* 97(3):310-316.
- [8] Akbarnia, B. A., Gabriel, K. R., Beckman, E., Chalk, D. 1992. Prevalence of scoliosis in neurofibromatosis. *Spine* 17(8 Suppl.):S244-248.
- [9] Crawford, A. H., Parikh, S., Schorry, E. K., VonStein, D. 2007. The immature spine in type-1 neurofibromatosis. *J. Bone Joint Surg.* 89:123-142.
- [10] Durrani, A. A., Crawford, A. H., Chouhdry, S. N., Saifuddin, A., Morley, T. R. 2000. Modulation of spinal deformities in patients with neurofibromatosis type 1. *Spine* 25: 69-75.
- [11] Crawford, A. H., Lykissas, M. G., Schorry, E. K., Gaines, S., Jain, V., Gregg, T., Viskochil, D. 2012. Neurofibromatosis: Etiology, commonly encountered spinal deformities, common complications and pitfalls of surgical treatment. *Spine Deformity Preview Issue Sept.* 2012: 85-94.
- [12] Friedman, J. M., Birch, P. H. 1997. Type 1 Neurofibromatosis: A descriptive analysis of the disorder in 1728 patients. *Am. J. Med. Genet.* 70:138-143.
- [13] Stevenson, D. A., Birch, P. H., Friedman, J. M., Viskochil, D. H., et al. 1999. Descriptive analysis of tibial pseudarthrosis in patients with neurofibromatosis 1. *Am. J. Med. Genet.* 84:413-9.
- [14] Boero, S., Catagni, M., Donzelli, O., et al. 1997. Congenital pseudarthrosis of the tibia associated with neurofibromatosis 1: Treatment with Ilizarov's device. *J. Pediatr. Orthop.* 17:675-684.
- [15] Johnston, C. E. 2002. Congenital pseudarthrosis of the tibia: Results of technical variations in the Charnely-Williams procedure. *J. Bone Joint Surg. Am.* 84A(10):1799-1810.
- [16] Dobbs, M. B., Rich, M. M., Gordon, E., Szymanski, D. A., Schoenecker, P. L. 2004. Use of an intramedullary rod for the treatment of congenital pseudarthrosis of the tibia. A long-term follow-up study. *J. Bone Joint Surg.* 86A:1186-1197.
- [17] Richards, B. S., Oetgen, M. E., Johnston, C. E. 2010. The use of rhBMP-2 for the treatment of congenital pseudarthrosis of the tibia: A case series. *J. Bone Joint Surg.* 92:177-185.
- [18] Schindeler, A., Birke, O., Yu, N. Y. C., Morse, A., Ruys, A., Baldock, P. A., Little, D. G. 2011. Distal tibial fracture repair in a neurofibromatosis type 1-deficient mouse treated

- with recombinant bone morphogenetic protein and a bisphosphonate. *J. Bone Joint Surg.* 93-B:1134-1139.
- [19] Jacquemin, C., Bosley, T. M., Svedberg, H. 2003. Orbit deformities in craniofacial neurofibromatosis type 1. *Am. J. Neuroradiol.* 24:1678-1682.
- [20] Nuss, D., Kelly, R. E. Jr. 2010. Indications and technique of Nuss procedure for pectus excavatum. *Thorac. Surg. Clin.* 20(4):583-597.
- [21] Szudek, J., Birch, P., Friedman, J. M. Growth in North American white children with neurofibromatosis 1 (NF1). 2000. *J. Med. Genet.* 37:933-938.
- [22] Stevenson, D. A., Moyer-Mileur, L. J., Murray, M., Slater, H., Sheng, X., Carey, J. C., Dube, B., Viskochil, D. H. 2007. Bone mineral density in children and adolescents with neurofibromatosis type 1. *J. Pediatr.* 150:83-88.
- [23] Illes, T., Halmai, V., deJone, T., Dubousset, J. 2001. Decreased bone mineral density in neurofibromatosis-1 patients with spinal deformities. *Osteoporos. Int.* 12:823-82.
- [24] Kuorilehto, T., Poyhonen, M., Bloigu, R., Keikkinen, J., Vaanane, K., Peltonen, J. 2004. Decreased bone mineral density and content in neurofibromatosis type 1: Lowest local values are located in the load-carrying parts of the body. *Osteoporos. Int.* 16:928-936.
- [25] Lammert, M., Kappler, M., Mautner, V.-F., Lammert, K., Storkel, S., Friedman, J. M., Atkins, D. 2005. Decreased bone mineral density in patients with neurofibromatosis 1. *Osteoporosis international: a journal established as result of cooperation between the European Foundation for Osteoporosis and the National Osteoporosis Foundation of the US* 16:1161-1166.
- [26] Duman, O., Ozdem, S., Turkkahraman, D., Olgac, N. D., Gungor, F., Haspolat, S. 2008. Bone metabolism markers and bone mineral density in children with neurofibromatosis type-1. *Brain Dev.* 30:584-588.
- [27] Heerva, E., Peltonen, S., Svedstrom, E., Aro, H. T., Vaananen, K., Peltonen, J. 2012. Osteoclasts derived from patients with neurofibromatosis 1 (NF1) display insensitivity to bisphosphonates in vitro. *Bone* 50:798-803.
- [28] Tucker, T., Schnabel, C., Hartmann, M., Friedrich, R. E., Frieling, I., Kruse, H. P., Mautner, V. F., Friedman, J. M. 2009. Bone health and fracture rate in individuals with neurofibromatosis 1 (NF1). *J. Med. Genet.* 46:259-265.
- [29] Brunetti-Pierri, N., Doty, S. B., Hicks, J., Phan, K., Mendoza-Londono, R., Blazo, M., Tran, A., Carter, S., Lewis, R. A., Plon, S. E., Phillips, W. A., O'Brian Smith, E., Ellis, K. J., Lee, B. 2008. Generalized metabolic bone disease in Neurofibromatosis type I. *Mol. Genet. Metab.* 94:105-111.
- [30] George-Abraham, J. K., Martin, L. J., Kalkwarf, H. J., Rieley, M. B., Stevenson, D. A., Viskochil, D. H., Hopkin, R. J., Stevens, A. M., Hanson, H., Schorry, E. K. 2013. Fractures in children with neurofibromatosis type 1 from two NF clinics. *Amer. J. Med. Genet* 161A:921-926.
- [31] Colby, R. S. and Saul, R. A. 2003. Is Jaffe-Campanacci syndrome just a manifestation of neurofibromatosis type 1? *Am. J. Med. Genet.* 123A:60-63.
- [32] Daston, M. M. and Ratner, N. 1992. Neurofibromin, a predominantly neuronal GTPase activating protein in the adult, is ubiquitously expressed during development. *Dev. Dynamics* 195:216-226.
- [33] Stevenson, D. A., Allen, S., Tidyman, W. E., Carey, J. C., Viskochil, D. H., Stevens, A., Hanson, H., Sheng, X., Thompson, B. A., Okumura, M., Reinker, K., Johnson, B., Rauhen, K. A. 2012. Peripheral muscle weakness in RASopathies. *Muscle Nerve* 46(3): 394-399.

- [34] Johnson, B. A., MacWilliams, B., Carey, J. C., Viskochil, D. H., D'Astous, J. L., Stevenson, D. A. 2012. Lower extremity strength and hopping and jumping ground reaction forces in children with neurofibromatosis type 1. *Hum. Mov. Sci.* 31(1):247-254.
- [35] Stevenson, D. A., Allen, S., Tidymen, W. E., Carey, J. C., Viskochil, D. H., Stevens, A., Hanson, H., Sheng, X., Thompson, B. A., Okumura, M., Reinker, K., Johnson, B., Rauen, K. A. 2012. Peripheral muscle weakness in RASopathies. *Muscle Nerve* 46(3): 394-399.

Chapter 112

COGNITION AND BEHAVIOUR IN NEUROFIBROMATOSIS TYPE 1: PATHOGENESIS AND EMERGING THERAPIES

***Jonathan M. Payne^{1,2}, Natalie A. Pride^{1,2}
and Kathryn N. North^{1,2,3}***

¹Institute for Neuroscience and Muscle Research,
The Children's Hospital at Westmead, Australia

²Discipline of Paediatrics and Child Health,
Faculty of Medicine, University of Sydney, Australia

³Murdoch Children's Research Institute,
Royal Children's Hospital, Melbourne, Australia

ABSTRACT

This chapter provides an overview of the neurocognitive, behavioral and developmental aspects of the neurofibromatosis type 1 (NF1) phenotype. Investigations into the pathogenesis of cognitive dysfunction are also presented, with a focus on human neuroimaging studies and mouse models. These studies have provided major insights into the molecular and biochemical abnormalities associated with NF1 that impact on cognitive performance. Preclinical studies suggest that pharmacological correction of these abnormalities has the potential to normalize aspects of the NF1 cognitive phenotype. These are reviewed along with the rationale for ongoing and future human clinical trials.

INTRODUCTION

Neurofibromatosis type 1 (NF1) is one of the most common single-gene disorders to affect the human nervous system, with a birth incidence of at least 1 in 2700. [1] The condition is caused by a mutation of the gene encoding neurofibromin on chromosome 17q11.2 and is characterized by a diverse range of cutaneous, neurological and neoplastic manifestations. [2, 3] Although identification of the *NF1* gene was made in 1990, diagnosis of the disorder is

usually based on the National Institutes of Health (NIH) Diagnostic Criteria, which were established in 1987 and re-evaluated without modification in 1997. [4, 5] Clinical features of NF1 include café-au-lait macules, Lisch nodules, axillary freckling, cutaneous and plexiform neurofibromas and optic gliomas. [3] Skeletal anomalies, including tibial dysplasia/pseudoarthrosis, scoliosis, and sphenoid wing dysplasia have also been well documented. [6] Although not a diagnostic feature of the condition, cognitive impairments are the greatest cause of lifetime morbidity in children with NF1, with up to 80% of children demonstrating moderate-to-severe impairments in one or more areas of cognition. [7] Cognitive deficits contribute to poor quality of life for individuals with NF1 as they result in school failure, poor peer relationships, failure to complete higher education, and limitation of career choice. The following chapter will provide an overview of cognitive, behavioral and developmental aspects of the NF1 phenotype, discuss the neural and biochemical pathologies undermining cognitive performance and will present the rationale for ongoing and future clinical trials.

1. NF1 COGNITIVE PHENOTYPE

Cognitive and behavioral impairment is one of the most common complications of NF1 in childhood and represents a major challenge to all clinicians who care for children with the condition. Children with NF1 experience noticeable difficulties in several areas of academic functioning, without consistently dramatically failing in any one particular area. [8] There is significant phenotypic variation between individuals with NF1, requiring an individualized approach to the clinical care of these children. As discussed below, cognitive and behavioral deficits manifest in a variety of ways, including academic failure due to intellectual impairment, specific learning disabilities, attention deficit-hyperactivity disorder (ADHD) as well as psychosocial maladjustment and poor social skills. The ultimate aim of all research in NF1 is to develop therapies that will improve patient quality of life. Before successful remediation of cognitive impairments and learning disabilities can occur, the NF1 phenotype needs to be carefully profiled. Over the past 20 years, significant advances have been made in describing the cognitive and behavioral phenotype of individuals with NF1 and despite the marked variability between individuals with the disorder, a number of core neuropsychological features have been identified. These in turn, provide a basis for investigations of disease mechanism and targets for therapy.

1.1. Intelligence

The earliest description of intellectual functioning in patients with NF1 was made by von Recklinghausen in 1882, when he described the first two patients with NF1. Of the first, he commented, “Apart from a great attraction to the male sex, she exhibited nothing unusual in her mental sphere”; and of the second, he reported that “His intelligence did not seem exceptional nor, on the other hand, below average.” [9] Since that time, scores of studies have reported on the intellectual capacity of individuals with NF1. Although earlier studies provided a gross overestimation of intellectual impairment due to recruitment biases, [10, 11] the

majority have reported a relative sparing of full-scale IQ, indicating that most children fall within the lower cusp of the normal range. When examining group data, the mean full-scale IQ tends to cluster around 90, rather than the normative mean of 100. [7, 12-18] Full scale IQ scores are normally distributed, indicating a true downward shift, rather than a subset of individuals skewing the group mean. [7] Indeed only around 6-8% of children with NF1 have an IQ lower than 70, compared to 2% of the general population. [7, 19] Despite the robust finding of a subtle lowering of full-scale IQ, there is no consistency between studies that have examined discrepancies between verbal and nonverbal IQ. While some report no difference between verbal and nonverbal abilities, [14, 16, 20, 21] others have reported significantly inferior verbal [22, 23] or nonverbal skills. [24-26]

It is important to take into account the limitations of intellectual measures. Although extremely valuable in providing a general level of cognitive functioning and in elucidating patterns of strengths and weaknesses, composite scales on IQ tests have often been shown to be insensitive to identifying specific cognitive impairments that are commonly described in individuals with NF1. For example, studies have highlighted minimal correlations between executive measures and IQ in children from the general population, suggesting that an IQ within the normal range does not preclude executive impairments in a given individual. [27] In order to obtain a more complete understanding of the NF1 cognitive phenotype, it is thus important to use assessment measures that evaluate narrower aspects of cognition and behavior.

1.2. Visuospatial Function

Visuospatial ability, which is the processing of visual orientation or location in space, has been one of the best described features of the NF1 cognitive phenotype and has been a significant area of focus when examining neuropsychological deficits. Eliason conducted the first study to specifically examine visuospatial abilities in 23 children with NF1. [25] Performance on the Judgment of Line Orientation task (JLO), a test requiring judgments about angulation (the orientation of various lines), was significantly abnormal in children with NF1. Since that study, the JLO has been used in a large number of studies, with the vast majority reporting impaired visuospatial judgment. [7, 16, 28-31] When examining NF1 group means, the degree of impairment has ranged from one [30] to over two [29, 32] standard deviations below the normative mean. Hyman and colleagues have also documented that 56% of children with NF1 performed at least one standard deviation below the normative mean. [7] In a more complex design, Schrimsher and colleagues examined the hypothesis that visuospatial abilities could be used to discriminate and classify children with NF1 from controls beyond parent reported ADHD symptomatology. [33] Discriminant analysis revealed that the JLO, Block Design subtest, the Recognition-Discrimination Test and the Beery Visual-Motor Integration Test were able to distinguish 92% of NF1 children from the control group. Deficits on a range of other visuospatial and visuo-perceptual measures, such as the Rey Complex Figure and the Test of Visual Perceptual Skills, have also been described. [7, 28] There is suggestion that some visuospatial measures, such as the JLO and Rey Complex Figure Test, are sensitive to executive deficits, which are also highly frequent in NF1 (see section 1.6 below). [34] The specific cluster of deficits observed across multiple visuospatial tasks, however, provides strong evidence of a true visuospatial deficit.

1.3. General Language

Language, which includes verbal expression (object naming, vocabulary and discourse), verbal comprehension, and verbal fluency, can be divided into two broad categories; expressive and receptive. Expressive language disorder is characterized by a difficulty in expressing oneself at the level expected for his/her developmental stage. Children with expressive language difficulties will often speak in short, simplified sentences, with omission of some grammatical features, such as past tense and will be able to understand language better than they are able to produce it. Receptive language disorders occur when a child has difficulties comprehending the speech of others. In most cases, a child with a receptive language difficulty will also have an expressive language disorder, which is known as a mixed receptive-expressive language disorder.

Early studies in NF1 focused much of their efforts on understanding nonverbal aspects of the disorder, giving little attention to verbal-based abilities. For example, Eliason (1986) identified a discrepancy between verbal and performance IQ (VIQ mean 99.6 vs PIQ mean 87.1) and 20 of the 23 children were impaired on tests of visuoperceptual function. [25] A follow-up study comparing children with NF1 from a learning disorders clinic to age, sex and IQ matched controls with a learning disability but without a known genetic disorder revealed that the NF1 group had significantly lower PIQ and decreased visuoperceptual performance on cognitive testing. [26] On the basis of this IQ discrepancy and the high incidence of visuoperceptual deficits, it was proposed that nonverbal learning problems were predominant in the NF1 population. Limitations of this early work, however included questionable diagnoses of NF1, high proportions of children with intracranial pathology (30%, [25] 25% [26]) that could have impacted test performance, and biased samples that were derived from learning disorders clinics and cannot be generalized to the NF1 population.

Mazzocco and colleagues were among the first researchers to undertake a detailed investigation of language in children with NF1. [30] Compared to unaffected siblings, children with NF1 ($n = 19$) performed significantly lower on tests of vocabulary, picture naming, written vocabulary, receptive language and verbal reasoning. A deficit in phonological awareness, particularly phoneme segmentation, was also identified in the NF1 group; an important finding given that phonological skills are recognized as an important precursor of literacy skills. Dilts and colleagues compared expressive and receptive language performance in 19 children with NF1 to their siblings on the Clinical Evaluation of Language Fundamentals-Revised (CELF-R) test. [28] They reported 58% of children with NF1 to have failed the screening test compared to a failure rate of 16% in the control group. Pure expressive language impairment was identified in 32% of the NF1 sample while 26% displayed combined expressive and receptive language deficits. Pure receptive language deficits were not identified in the NF1 cohort. In a study by Hyman and colleagues, broad deficits in both receptive and expressive language deficits were identified in 81 children with the condition. [7] Their greatest difficulties were in defining word meanings, comprehending simple passages, describing pictures and higher-level verbal reasoning skills. Performance on language measures was highly correlated with IQ, with only 2.5% of the sample displaying a discrepancy between FSIQ and receptive language scores and no child with NF1 demonstrated a discrepancy between FSIQ and expressive language scores. As a whole, the wide range of language skills implicated in NF1, suggest that a global language deficit is a relatively common phenotypic feature and that expressive skills are more likely to be implicated than receptive. As has been observed previously, however, it is important

to note that NF1-related language deficits nearly always occur with concurrent difficulties in other cognitive domains, such as a visuospatial impairment. [35, 36]

1.4. Learning and Memory

Memory refers to the mental process of acquiring and retaining information for later retrieval. [37] Classic models of memory conceptualize three separate stages of information processing: (1) encoding the initial processing of information and instigating a transient memory trace (2) the storage and consolidation of information contained within the memory trace into a more enduring form and (3) the retrieval of the stored information. [38] The question of whether learning and memory impairments are common in individuals with NF1 is a critical one, given the convincing evidence that hippocampal abnormalities and learning impairments can be corrected in *Nf1*^{+/-} mouse models (see Section 4.2 for detailed discussion). [39, 40] If deficits in learning and memory are identified as a phenotypic feature of children with NF1, then measures of learning and memory would be an obvious choice for a primary outcome in human trials.

Review of the literature up to 2006 by Levine and colleagues [34] indicated that while approximately half the studies investigating memory in NF1 indicated impairment in either verbal or visual memory, [29, 41] the others did not. [7, 28, 42, 43] A limitation with a number of these studies is that the measures used to assess learning and memory also rely on other complex cognitive operations, such as language or visuospatial abilities, which are well documented impairments in children with NF1. It is unclear whether such studies are reporting “true” hippocampal-based learning impairments, or whether they actually reflect a primary language or visuospatial impairment instead. Since that review, several studies have published data on visuospatial learning and memory in NF1, with the purpose of drawing a phenotypic parallel between NF1 murine models and humans with the condition. [44, 45] Ullrich and colleagues examined visuospatial learning in children with NF1 (n = 10) and sibling controls (n = 6) using a computerized Arena Maze task. [44] This virtual environment task, which assesses learning of spatial navigation using relations between distal cues, was developed as a human analogue of the Morris Water Maze; the visuospatial learning task often employed in mice studies. [39, 40] Children were required to navigate to a target hidden in the arena over multiple learning trials. Although children with NF1 demonstrated improvements in navigation over repeated trials, they displayed greater difficulty in earlier search trials compared with unaffected siblings. Performance of the NF1 group on the final trial (a probe trial in which the target was removed) indicated less dwell time in the target quadrant, suggesting recall of the target location was inferior to control participants. In another study, Payne and colleagues used a visuospatial Paired Associate Learning (PAL) test from the Cambridge Neuropsychological Test Automated Battery (CANTAB); a collection neuropsychological tests specifically designed to improve the comparative assessment of cognition from animals to humans. [46] Participants were required to make arbitrary associations between a simple visual pattern and a spatial location. While the task commenced with only one paired association to be recalled, it became progressively more difficult with eight paired associations required to be formed in the final stage. The PAL task has been well validated in human clinical populations with damage to mesial temporal lobe structures [47-49] and in monkey studies which report impaired learning following administration of NMDA antagonists which block long term potentiation

(LTP) in the hippocampus. [50, 51] The authors reported that compared to controls ($n = 29$), children with NF1 ($n = 71$) displayed a reduced capacity to form arbitrary visuospatial associations after an initial presentation of the stimuli and a poorer learning curve across repeated presentations of the same stimuli. Interrelationships between performance on the PAL task and other cognitive abilities that may potentially influence performance on the task, such as intelligence, attention and visuospatial function were also explored. While regression analysis indicated that FSIQ was a significant predictor of visuospatial learning, the NF1 group was found to make errors at twice the rate of the control group after controlling for the influence of FSIQ, providing convincing evidence for a hippocampal-based learning impairment in children with NF1.

1.5. Attention

Attention is the cognitive process of selectively filtering out irrelevant aspects of environmental stimuli in order to benefit processing of behaviorally relevant information. In everyday life, attention is mediated by two partially segregated networks – a cognitive (top-down) network that is under an individual's control, guided by knowledge, expectation and current goals, and a bottom-up network that automatically orients attention to sensory stimuli that are unexpected, abrupt, novel, salient or potentially dangerous. [52] The dynamic interaction between these top-down and bottom-up processes control where and how we allocate attentional resources. This distinction is critical, as evidence suggests that different neurocognitive processes control these two attentional mechanisms [52] and that distractibility and inattentiveness might arise from an increase in bottom-up interference or from a failure to engage top-down control mechanisms. [53]

Attention difficulties are one of the most commonly described impairments within the NF1 literature. The Test of Everyday Attention for Children (TEA-Ch) has been used in a number of studies to assess different attentional demands placed on children. In a large sample of children with NF1 ($n = 199$), Payne and colleagues demonstrated that children with NF1 performed significantly poorer than sibling controls ($n = 55$) and normative data on measures of divided attention (61% of the NF1 group fell more than 1 SD below normative data), sustained attention (56%), attentional control (54%) and selective attention (39%). [21] Other studies using the TEA-Ch have also indicated impairments in divided attention, [15, 54] sustained attention, [7, 15, 54] and attentional control. [7, 15] Continuous performance tests (CPTs), such as Conners CPT-II and the Test of Variables of Attention (TOVA), have also been used to assess sustained attention and vigilance. Increased omission errors, which are recorded when participants fail to respond to a target, have been documented in children with NF1 when compared to unaffected siblings [13, 30] and normative data. [41, 54] Although this finding is not universal, [28, 55] the use of CPTs to identify sustained attention impairments is not necessarily straightforward. By their nature, CPTs tend to be drawn out and monotonous, possibly resulting in children changing the strictness of their response criterion. As such, poor performance on these tasks may be due to children becoming less strict in deciding whether a signal is a target or not, and not because their ability to attend to the target is impaired. Furthermore, like other cognitive measures, they are administered in controlled conditions that lack the environmental challenges faced by the child in the real world.

In an attempt to overcome these limitations, Gilboa and colleagues assessed attention function in children with NF1 using a more ecologically valid, virtual reality system. [56] Their Virtual Classroom consisted of a head-mounted display that simulates a real-world classroom. Children were required to monitor digit sequences presented on a “blackboard” over a 10 minute period while real-world classroom distractions, such as ambient classroom noises, a person walking into the room and a flying paper aeroplane, were occurring. Results showed that the NF1 group ($n = 29$) made more omission errors than controls ($n = 25$), indicating a reduced focus on the target stimuli. There was also a significant relationship between omission errors and a parent rating scale of inattention, supporting the validity of the Virtual Classroom task and the hypothesis that NF1 is characterized by inattention. It would be interesting for future research to examine the associations between outcomes of the Virtual Classroom task and traditional cognitive measures of attention.

1.6. Executive Functions

Executive functions are interrelated cognitive and behavioral skills that are responsible for purposeful, goal-oriented activity, such as planning/organization, cognitive flexibility, and impulse control. [57] A primary purpose of the executive functioning system is to facilitate our adaption to novel situations, particularly when routine actions are insufficient. [58] From a neuropsychological perspective, executive functioning plays a unique “command” role in terms of guiding and regulating thought and behavior and is considered an umbrella term for cognitive processes such as planning/organization, mental flexibility, initiation and monitoring of actions, problem solving, working memory and the ability to inhibit impulses. Neuroimaging and brain lesion studies converge on the view that executive functions are critically dependent on the frontal cortex; indeed the terms ‘executive function’ and ‘frontal lobe’ are often used synonymously. However, recent theories have suggested that this view is overly simplistic and subcortical regions such as the basal ganglia, cingulate gyrus, cerebellum and thalamus are also critically involved. The link between this neural circuitry and executive functioning is demonstrated by a range of neurological and psychiatric disorders with prominent involvement of prefrontal and frontostriatal circuitry, such as attention deficit hyperactivity disorder, [59] autism spectrum disorders [60] and schizophrenia; [61] all of which are characterized by impairments of the executive system.

Early documentation of executive impairments in NF1 largely consisted of anecdotal reports and behavioral descriptions, including increased impulsivity, distractibility, failure to plan, an unstructured learning style and monochromatic problem solving strategies (such that one strategy is excessively used at the expense of more efficient and effective strategies). [20, 25, 62] Initial studies using cognitive-based measures reported that attention, sequencing and verbal working memory, as assessed with the Freedom from Distractibility Index from the Wechsler intelligence scales, were impaired in children with NF1. [25, 63] Since these early studies, a number of empirical studies have explicitly examined executive abilities in children with NF1, confirming executive dysfunction as a core phenotypic feature of the condition. Deficits in cognitive control, [7, 15, 21, 41, 64, 65] inhibitory control, [16, 65, 66] working memory, [7, 15, 16, 66, 67] abstract concept formation, [29] and spatial/action planning [7, 15, 17] have been identified. A number of studies reporting on executive functioning in NF1 have controlled for other cognitive abilities, such as intelligence or visuospatial skills, which on the

whole, indicate that executive impairments persist over and above any lowering of intelligence. Ferner and colleagues reported group differences between individuals with NF1 and normal controls on measures of selective and divided attention after including IQ as a covariate. [41] Impairments on measures of spatial working memory and inhibitory control have also been reported after controlling for verbal intellectual and visuospatial skills. [16] Similarly, significantly reduced planning abilities have also been reported to remain after controlling for IQ [7, 17] and visuospatial abilities. [17]

Perhaps more so than any other area of cognition, the valid identification of impairments of executive function is difficult and often elusive. Indeed, it is generally accepted that traditional standardized neurocognitive measures of executive function may lack sufficient sensitivity to relate directly to the multidimensional nature of situations experienced in daily life, particularly in pediatric populations. [68, 69] That is, they have questionable ecological validity; the degree to which test performance corresponds to real world performance. [70] For example, these measures (a) are administered in a structured clinic setting (b) are dependent of lower-level cognitive skills, such as language, visuospatial skills and memory, which are commonly impaired in children with NF1 (c) have often been designed for use in adult populations (usually to determine the presence/absence of a lesion) and subsequently adapted for children, and (d) are often dissociated with real-world behavior. In a large sample ($n = 199$), Payne and colleagues assessed the frequency of everyday attention and executive impairments in children with NF1 and examined the relationship between these functional impairments and performance on neurocognitive measures. In addition to everyday impairments in attention, children with NF1 were reported to display significantly more executive deficits in their daily activities than controls. Working memory (60% impaired), self-monitoring (59%) and planning and organization (58%) were the most common executive abilities at risk of impairment. Somewhat surprisingly, there were no associations between many neurocognitive measures of executive function (such as the Tower of London, Children's Category Test and Controlled Oral Word Association Test) and the everyday skills they purport to assess. The best neurocognitive predictors of functional executive abilities were measures of attentional control and working/immediate memory. These findings highlight the importance of supplementing neuropsychological assessments with functional assessment tools to provide a more accurate and sensitive encapsulation of a child's strengths and weaknesses to guide remediation programs.

1.7. Academic Achievement and Learning Disability

The neurocognitive impairments experienced by children with NF1 place the majority of them at significant risk of academic underachievement. [12, 19, 71] North and colleagues reported 65% of their sample of children with NF1 ($n = 40$) had impairment in at least one area of academic achievement; 45% performed more than two years below chronological age on tests of reading accuracy; 48% displayed impaired performance on measures of reading comprehension; and 28% were reported to fall more than 1.96 standard deviations below the mean on a standardized mathematics test. [20] More recently, Krab and colleagues calculated a 'learning efficacy' score, based on a discrepancy between academic age-equivalent scores and a child's chronological age. [12] They reported that 75% of children with NF1 performed more than one standard deviation below grade peers in at least one of the domains of reading,

spelling or mathematics. Only 10% of children did not demonstrate any school-functioning problems in this cohort and those without apparent learning disabilities still frequently displayed neurocognitive deficits. A current issue in the NF1 literature is whether a learning disability should be identified in children who are performing significantly worse than a comparison group or normative data, or whether a significant discrepancy between academic achievement scores and the individual's IQ is additionally required. [8] Hyman and colleagues examined this distinction by examining frequencies of specific learning disability (SLD; academic difficulties in the presence of normal IQ) and general learning disability (GLD; academic difficulties in the presence of low IQ) in children with NF1. [19] Using the Wechsler Individual Achievement Test, the authors identified SLD in 20% of their NF1 sample, GLD in 32% and normal academic achievement in 48%. Interestingly, 15 of the 16 children identified as having a SLD were male, suggesting that females with NF1 may not be at any greater risk of SLD than those in the general population. [19] This gender effect will ideally be confirmed in future studies.

Research examining the reading performance of children with NF1 has typically employed academic achievement measures to identify reading disabilities based on significant discrepancies between intellectual functioning and reading achievement. [30, 72] For example, Cutting and colleagues found children with NF1 had single word reading and reading comprehension deficits when compared with normal controls. [72] Other skills associated with reading, including rapid naming and phoneme segmentation, were also delayed. Similarly, Mazzocco and colleagues reported a higher incidence of reading deficits among school-aged children with NF1, based on Cluster Scores on the Woodcock-Johnson-Revised, when compared to unaffected siblings. [30] Children with NF1 were found to have verbal weaknesses, including deficits in verbal reasoning, vocabulary, picture naming and receptive language. More basic linguistic deficits were also identified on measures of phonological memory and phoneme segmentation. A recent study compared the reading profile of children with NF1 to those with idiopathic reading disability. [73] Results indicated that children with NF1 and reading disability performed similarly to children with idiopathic reading disability on measures of phonological awareness, rapid naming and reading comprehension. However, the NF1 + reading disability group displayed pronounced visuospatial deficits when compared to children with idiopathic reading disability and typically developing readers. While visuospatial deficits are considered a hallmark feature of NF1, it is interesting to note that children with NF1 without reading disability did not have a distinct visuospatial weakness. In addition, visuospatial skills only correlated with basic reading ability in the NF1 + reading disability group. In a further study, Watt and colleagues reported that despite normal levels of intellectual functioning, 67% of children with NF1 ($n = 30$) demonstrated impaired non-word reading. [74] Non-words are pronounceable nonsense words that require the utilization of common rule-based grapheme-to-phoneme correspondences to correctly sound out the word (e.g., "peef"). These findings demonstrate that a significant proportion of children with NF1 experience difficulty employing spelling-to-sound rules to assemble a pronunciation when reading.

In addition to examining literacy skills and linguistic ability of children with NF1, a number of studies have also explored mathematic ability. As is pointed out by Moore, the performance of children with NF1 on mathematics tasks covers a broad range of abilities, but the distribution of scores is shifted somewhat lower than the population mean. [75] Most studies examining math ability have reported deficits on measures of math calculations [15, 19, 72, 76] and math

word problems. [15, 30, 77] Although the mechanisms underlying mathematical learning disability in children with NF1 are not clear, evidence from the idiopathic mathematical learning disability literature suggests that cognitive impairments in working memory, visuospatial skills and ADHD can all affect math ability. [75] Recent research in a large NF1 cohort has started to elucidate some of the cognitive mechanisms underlying math ability using backward linear regression models. [15] Attentional control (Creature Counting Subtest from the TEA-Ch) was the best predictor of math ability in children with a comorbid diagnosis of NF1 and ADHD and attentional control and visuospatial skills (JLO) were the best predictors of math achievement in children with NF1 only. Given that the Creature Counting subtest places demands on inhibition, working memory and set shifting, these findings suggest that poor performance on math tasks can be partially attributed to executive sub-skills that direct other higher-order processes.

2. THE SOCIAL AND BEHAVIORAL PHENOTYPE

In addition to cognitive impairments, NF1 patients also demonstrate a number of social and behavioral deficits. One of the most prevalent behavioral phenotypes is ADHD. Characterized by persistent and pervasive symptoms of inattention and/or hyperactivity/impulsivity, comorbid ADHD has been reported to occur with frequencies varying from 17% [78] to 50% of pediatric NF1 samples. [29] The extent of this range is most likely due to variations in clinic demographics, sampling biases and inconsistencies in diagnostic methodology. Prevalence rates in larger samples of children ($n = 80$ -199) assessed on a consecutive basis from Neurofibromatosis Clinics go some way to diminishing some of these biases and estimate the frequency at approximately 30-40%; [7, 15, 16] a marked increase compared to estimates in the general population (3-5% of school aged children). [79]

A burgeoning area of interest in the NF1 literature is the relationship and impact of ADHD on the cognitive phenotype of individuals with the condition, with studies for the most part, examining phenotypic differences of children with NF1 to those who also meet ADHD criteria. [15-17, 64, 80] Results to date lack consistency. For example, while Lidzba and colleagues [80] reported that symptoms of ADHD (with or without hyperactivity) have a negative impact on intellectual performance of individuals with NF1, other retrospective and prospective studies report that ADHD has no impact on the intellectual abilities of children with NF1. [15-17] In a large retrospective study ($n = 192$), Pride and colleagues reported that while children with NF1 and comorbid ADHD exhibited poorer academic achievement, and sustained attention abilities than children with NF1 only; other aspects of attention, visuospatial skills, memory and intellectual function were no more impaired in children with NF1 and ADHD than those without the additional comorbidity. [15] With respect to executive functions, a recent prospective study examined whether spatial working memory and response inhibition – executive abilities commonly impaired in ADHD cohorts [81] – were impaired in children with NF1 only ($n = 49$), and whether a comorbid diagnosis of ADHD ($n = 35$) exacerbated any executive deficits. [16] Results revealed that children with NF1, with or without comorbid ADHD, displayed spatial working memory and inhibitory control impairments. Interestingly, there was no difference in the degree of impairment between the NF1 only and NF1+ADHD groups, suggesting that ADHD comorbidity did not exacerbate the degree of executive

dysfunction. Similarly, in a study investigating cognitive and motor control in NF1, Huijbregts and colleagues reported cognitive control deficits in their NF1 cohort ($n = 30$), seven of whom were diagnosed with ADHD. After excluding participants with ADHD, cognitive control deficits in the NF1 only participants remained. [64] In another study, children with NF1 and ADHD have been reported to display poorer spatial planning than an NF1-only group on one measure (Tower test), but not on others (Mazes, Rey Complex Figure Test). [17] Taken together, these studies suggest that executive impairments are not uniquely associated with ADHD in NF1; it is clear that other cognitive and/or psychological factors underlie the ADHD phenotype. Furthermore, a diagnosis of ADHD does not explain the cognitive problems in NF1 – the cognitive, behavioral and motor impairments observed in individuals with NF1 are consequences of NF1-related pathology rather than an independent ADHD condition. This distinction has important implications for future clinical trials investigating pharmacological and cognitive-based interventions in NF1, such that trials should not only include children that satisfy ADHD diagnostic criteria, but also those with cognitive and/or functional impairments without an ADHD phenotype.

A currently unresolved issue in the literature is whether ADHD in children with NF1 is a cognitive-behavioral phenotype of the disorder or whether it is a comorbid condition in a subsample of children. There is evidence for quite distinct phenotypic differences between children with NF1 and comorbid ADHD and those with ADHD from the general population. The majority of studies report that children with NF1 are more likely to meet the criteria of the inattentive or combined subtype of ADHD rather than the hyperactive/impulsive type, which is the most common subtype in children with ADHD alone. [15] In the general population, the frequency of ADHD is up to nine times higher in males than females, whereas both genders are equally affected in NF1. [82] As is discussed by Huijbregts, structural brain abnormalities observed in individuals with NF1 and ADHD also appear to differ quite markedly from observations in those with ADHD from the general population. [64, 83] While ADHD is associated with smaller volumes in a number of anatomic regions, such as the right prefrontal cortex, corpus callosum, caudate and cerebellum, [84] NF1 is often associated with macrocephaly, increased grey matter volume, abnormal grey-to-white matter ratios and an enlarged corpus callosum (see Section 5.1.1. below). [42, 64, 85] These differences add weight to the argument that ADHD diagnoses in NF1 are based on a more severe behavioral phenotype (NF1-related pathology) rather than an independent comorbid condition such as ADHD. In order to further understand the behavioral mechanisms underlying the ADHD phenotype in NF1, it is important for future studies to include ADHD control groups from the general population to compare cognitive, psychological and neuroanatomic constructs to help our understanding of the degree of cognitive and biological convergence between these two disorders.

Social dysfunction is also a prominent feature of NF1. Parents of children with NF1 commonly describe their child as shy or awkward with peers and indicate they are often the victims of teasing and bullying. As a result, increasing effort over the last decade has been devoted to the analysis of social skills in NF1. Responses on parent and teacher behavioral questionnaires have indicated poorer social skills, social outcomes and increased social problems in children with NF1 compared to unaffected children. [28, 86-88] Adults with NF1 also demonstrate a lower frequency of pro-social behaviours (such as eye-contact when speaking, smiling when they see someone) than unaffected matched controls, suggesting that social dysfunction continues into adulthood. [89]

There is currently no established theory to explain social dysfunction in NF1 and investigations into the processes that may influence the link between NF1, central nervous system (CNS) functioning and social behaviour have rarely occurred. The evidence for cognitive dysfunction underlying social and behavioral problems in NF1 is mixed. While some have reported no relationship between cognitive measures and social behaviour, [65] others have reported a significant difference between social processing abilities of children with NF1 and a healthy control group that disappeared once cognitive processing abilities were accounted for. [67] Others studies have found a link between social problems and greater disease severity in terms of neurological involvement (CNS involvement, cognition, and attention deficit hyperactivity disorder). [86, 88, 90]

One area that has received insufficient attention is social cognitive abilities in NF1. This is particularly relevant given that an increased incidence of autistic traits have been reported in children with NF1. [90] Social cognition encompasses a broad set of processes that allow a person to recognize, understand and behave appropriately with respect to socially relevant stimuli; processes that are critical for adequate social functioning. [91] Huijbregts and colleagues examined social information processing in 32 children with NF1 and 32 controls and found that children with NF1 displayed a weakness with facial recognition and processing fear and anger. [67] A more recent study using voxel-based morphometry examined the relationship between grey matter (GM) brain volume and social perception in 16 adults with NF1 and compared region of interest volumes to 16 matched controls. [92] Although the sample size was limited, results indicated that adults with NF1 experience difficulty processing anger and interpreting social exchanges involving subtle sarcasm. Structural abnormalities in a number of brain regions associated with social cognition were also documented, including the posterior cingulum, ventromedial prefrontal cortex and the superior temporal cortex; the latter correlating with social perception performance. These two studies raise important questions regarding the underlying causes of social dysfunction in NF1 and open several avenues for further investigation. Future work should focus on the relationship between social functioning, social cognitive processes and cognitive dysfunction to help delineate the underlying causes for social dysfunction in NF1.

3. NEURODEVELOPMENTAL ASPECTS

Parents of children with NF1 often raise concern about their child's development around their second or third birthday. Despite this, the majority of studies have limited their investigations to children of school age (6 to 16 years of age). Very few studies have addressed issues of cognitive development in young toddlers and infants; a developmental period which is crucial, as interventions are more likely to have positive outcomes when commenced during this period of increased neural and behavioral plasticity. Early studies that have examined the cognitive profile of young children with NF1 revealed delays in cognition, motor skills and language. [24, 62] Interpretation of these results, however, is limited by either a small sample size [24] or the lack of an appropriate control cohort. [62] A more recent study examining cognition in 39 toddlers with NF1 (aged 21-30 months) and 42 age-matched controls reported that NF1 participants displayed significantly poorer mental and motor development than controls. [93] Parental responses also indicated that most of the children with NF1 experienced

language delays; however group differences in behavior and executive functioning were not evident during this developmental period. Language delays have also been reported in a study which used the preschool version of the CELF to assess receptive and expressive language development in 19 young children with NF1. [94] Results revealed that 37% of the sample demonstrated delays in expressive language and 37% demonstrated delays in receptive language. The characteristic downward shift in IQ, poor visuospatial skills and inattention have also been identified in another cohort of preschool-aged children. [13] Taken together, these findings indicate that cognitive and language delays manifest early in life, can be measured before the child starts school, and appear to mirror the difficulties identified in older children. As such, early assessment and intervention should be considered as it may help circumvent the negative impact of cognitive and motor dysfunction associated with NF1.

Only a handful of studies have explored the cognitive correlates of NF1 in adults. [14, 95-97] Ferner and colleagues examined cognitive function in a large sample of adults and children with NF1. [14] The NF1 group displayed lower IQ scores, reading ability, verbal memory and attention than controls. Difficulties in developing strategies for complex, novel tasks were also reported. Zöller and colleagues examined cognitive performance in 30 adults with NF1. [95] Compared to controls, the NF1 group exhibited deficits in inductive reasoning, visuoconstructive ability, visual and tactual memory, logic abstraction, coordination and mental flexibility. Depressive symptoms were also associated with reduced psychomotor speed. Pavol and colleagues examined the hypothesis that performance on measures of visuospatial and attention abilities could be used to differentiate adults with NF1 ($n = 20$) and unaffected controls. [97] Two tests of visuospatial ability and a language measure were found to be the best predictors of group membership, with a measure of visuomotor integration the best discriminator between groups. Reading difficulties have also been reported to occur in adults with NF1, with 13% of patients experiencing difficulties reading a passage aloud and an additional 63% demonstrating paraphrase misreadings (e.g., reading “the one who could” as “the one of them who could”). [96] Difficulties with descriptive writing (40%) and spelling (50%) were also prominent.

Little is known about the natural history of cognitive functioning from childhood to adulthood in patients with NF1. Two hypotheses appear to emerge from the limited literature that exists. First, some cross-sectional studies suggest improvements in general cognitive functioning from childhood to adulthood [98] as well as specific improvements in language and motor skills. [95] The second hypothesis, which asserts a static cognitive profile, is based on evidence from cross-sectional studies that report a stable IQ throughout childhood and adolescence [99] and between childhood and adulthood. [14]

Longitudinal studies are beginning to provide insight into the natural history of cognitive impairment in NF1. Cutting and colleagues conducted growth curve analyses on six cognitive measures over repeated time points in 19 children and adolescents with NF1. [100] Performance of the NF1 group did not improve relative to sibling controls on impaired measures (JLO, Vocabulary and Block Design), indicating stable function over time. Hyman and colleagues prospectively followed 32 patients with NF1 and 11 unaffected sibling controls, examining cognition during childhood (mean age 12.6 years) and again 8 years later (mean age 20.1 years). [43] Although cognitive performance was reported to be stable across this eight year period, reassessment of a subset ($n = 18$) of this cohort 10 years later indicated a significant improvement in intellectual function. [101] As discussed below (see Section 4.1.1), cognitive improvement was limited to individuals with NF1-specific brain lesions in childhood. A large

prospective longitudinal study, spanning childhood to adulthood, utilizing structural neuroimaging and a comprehensive cognitive battery at each time-point, is needed to confirm and extend these findings.

4. PATHOGENESIS OF COGNITIVE IMPAIRMENTS

Both human data and animal models are beginning to provide crucial insights into the underlying pathological correlates of cognitive dysfunction in NF1. It is critical to understand these mechanisms before directed clinical trials can be implemented. There is already compelling evidence for a relationship between cognitive dysfunction and a number of structural and functional brain abnormalities. [102] As discussed below (see Section 4.2), mouse models have also proven essential in revealing the molecular and biochemical abnormalities that occur due to a loss of neurofibromin and the impact of these abnormalities on cognition.

4.1. Human Neuroimaging Studies

4.1.1. Structural

Macrocephaly and increased brain size have been reported in up to 50% of children with NF1. [103-105] Several studies suggest that increased brain volume in children with NF1 appears to be largely driven by an increase in white matter; predominately in frontal regions and the corpus callosum. There is, however, also indication of increased grey matter volume in posterior regions. [105] Although the majority of studies have repeatedly demonstrated a lack of any association between macrocephaly and neurocognitive performance, [7, 71, 106, 107] some have reported a number of associations between grey matter properties and cognition. Moore and colleagues reported a larger discrepancy between intelligence and academic achievement in children with increased grey matter volume. [42] In contrast, Said and colleagues found a relationship between decreased grey matter volumes and reduced visuospatial functioning. [108] Others have reported absence of the normal positive relationship between grey matter volume and intelligence. [109] The diverse nature of these associations, however, makes it difficult to make firm conclusions regarding the role of grey matter anomalies on cognitive performance.

A number of studies have reported on the influence of focal neuroanatomical abnormalities on cognition. Billingsley and colleagues reported a loss of the normal asymmetry of the planum temporale in children with NF1; a structure within the sylvian fissure located on the on the superior surface of the temporal lobe. [110] Greater symmetry between left and right hemispheres was associated with reduced reading ability and math performance in relation to IQ scores, suggesting that abnormal development of the planum temporale could be a risk factor for specific learning deficits in children with NF1. A relationship between language and an atypical right inferior frontal gyrus has also been documented, with atypical morphology associated with increased language and academic performance. [76] An enlarged corpus callosum, the primary white matter tract connecting the left and right hemispheres, has also been reported in a number of studies, [42, 85, 111, 112] some of which controlled for total brain

size. [85, 112] Correlations between corpus callosum cross sectional area and cognition have also been documented. A larger corpus callosum has been associated with academic underachievement, lower intelligence, executive functioning, and visual spatial impairment, [85, 113] while a smaller corpus callosum has been associated with reduced attention abilities. [111]

NF1 is also associated with focal areas of high intensity observed on T2-weighted MRI (T2H) which are seen in 60-80% of children with the disorder. [114-116] T2H commonly occur in the basal ganglia, cerebellum, brainstem, thalamus and subcortical white matter and are thought to represent areas of increased fluid content within the myelin sheath or dysplastic areas of white matter formation. [115] While the number, size and intensity of T2H have been reported to decrease with age, this is limited to those occurring within the basal ganglia, thalamus and brainstem. Lesions located within the cerebral hemispheres, hippocampus and deep white matter appear to remain stable over time, possibly indicating a different pathological basis. [43, 101, 115] The question of whether T2H represent biological markers for cognitive dysfunction is controversial. While some moderate-to-large studies using clinic-based samples and quantitative neuropsychological measures have identified a significant association between the presence of T2H and a lowering of IQ, [117, 118] others have not. [18, 78] Although evidence for an association between cognitive impairment and either the presence or number of T2H remains contentious, studies drawing parallels between cognition and T2H location have consistently identified a relationship between thalamic lesions and generalized cognitive impairment, including borderline IQ. [114, 116, 119, 120] Preliminary evidence from a recent longitudinal study following a cohort of individuals with NF1 (n = 18) over an 18 year period provides evidence of a relationship between T2H status in childhood and cognitive changes in later life. [101] The authors reported a significant increase in general cognitive function in patients with discrete T2H in childhood; patients without lesions in childhood exhibited a stable profile over the duration of the study. These results raise the possibility that resolution of T2H over time is accompanied by an improvement in general cognitive performance. This needs to be confirmed, however, in a larger prospective study that includes longitudinal examination of the relationship between T2H location (particularly in the thalamus) and cognitive performance.

4.1.2. Functional

When a brain region is performing a task, it consumes more oxygen. To meet this increased demand, blood flow is increased to the active area. Functional MRI (fMRI) detects these changes in blood oxygenation and flow that occur in response to neural activity via the blood oxygen level-dependent (BOLD) response. [121] By comparing activation maps of children with NF1 to comparison groups, it is possible to identify abnormal neural systems and how they relate to cognitive performance. Despite fMRI providing non-invasive neural activation maps with excellent spatial and good temporal resolution, there has been surprisingly little published within the NF1 literature. An area that has received some attention in the NF1 literature is the neural correlates of visuospatial impairment. [31, 77] In one study, 13 children with NF1 and 13 unaffected controls performed an analogue of the JLO task in the scanner. [31] Comparison of activation patterns revealed that the NF1 group demonstrated significantly greater left hemisphere activation as opposed to the dominant right hemisphere activation in controls. These results suggest a functional correlate of visuospatial impairment in NF1 may be a dysfunctional right hemisphere network. Billingsley and colleagues have also reported

greater activity in posterior relative to anterior cortical regions in children with NF1 during a letter/number rotation task. [77] The authors hypothesized that this finding may reflect an alternative resource allocation due to functional inefficiencies in anterior regions.

Functional MRI has also been used to investigate the neural bases for phonologic processing in NF1. [122] Compared to controls, children with NF1 demonstrated less inferior frontal activation relative to temporoparietal activation on a written rhyming decision task. Behavioral performance on the fMRI tasks was related to a number of academic literacy measures for the NF1 group, including letter-word identification and a spelling task. The authors speculated that this pattern could reflect recruitment of more posterior than frontal regions as a result of functional inefficiency of the frontal lobe. A further study, investigating the functional correlates of spatial working memory, has also implicated frontal lobe inefficiencies in young adults with NF1. [123] Compared to controls, the NF1 group displayed less task-related activation in the dorsal lateral prefrontal cortex, parietal cortex, and striatum while completing a spatial working memory task. The degree of dorsolateral PFC activation correlated with behavioral performance on the task.

As clinical trials continue in NF1 cohorts (See Section 5.0), neuroimaging techniques will begin to provide valuable surrogate outcomes by comparing neural activity pre- versus post-treatment. A number of neuroimaging techniques could be used in such a role, the most obvious being task-based fMRI, in which neural activation is measured while the patient performs a cognitive task in the scanner. Resting state fMRI (R-fMRI) is a very promising and powerful neuroimaging tool that could also prove a valuable biomarker. It examines regional connectivity by recording and analysing spontaneous modulations in the BOLD response in the absence of any explicit input or output. Resting state fMRI has been shown to predict the task-response properties of brain regions [124, 125] identify participants' aptitude for different cognitive tasks, [126, 127] and is well suited to monitoring treatment effects by studying participants before and after treatment. Recently published data from a Phase I treatment trial of lovastatin in a subset of seven patients with NF1 suggest that treatment increased functional connectivity within the medial prefrontal lobe and posterior cingulate cortex, highlighting the sensitivity of the technology. [128]

Taken together, the neuroimaging literature to date highlights both structural and functional brain abnormalities in individuals with NF1; abnormalities that contribute to the cognitive impairments commonly observed in the condition. Due to the relatively small number of studies, variable methodologies, different ascertainment procedures, often small sample sizes and heterogeneous nature of the disorder, clear consistent findings are still emerging. As has been noted elsewhere, it is important that future studies investigate specific hypotheses in well-defined subsamples using multiple control groups where appropriate. [102] Assimilating knowledge of the molecular mechanisms underlying cognitive and behavioral impairments in NF1 (see Section 4.2 below) and data on the functional pathways in NF1 and ADHD will significantly further knowledge in these areas and guide the design of future clinical and preclinical trials.

4.2. Insights from Animal Models

Mouse models have been critical experimental tools in which to explore the etiology of cognitive deficits in NF1. The NF1 gene encodes neurofibromin, a 250kDa protein that plays

important roles in several biochemical processes, including modulation of cAMP levels, [129] microtubule binding, [130] and mTOR signaling. [131, 132] It is the effect of neurofibromin on RAS, a protein implicated in cell proliferation and differentiation that has been most extensively investigated. [133-135]

Mice with a heterozygous null mutation of the *Nf1* gene have been the predominant model used to study the result of aberrant RAS signaling on cognition and the cognitive and behavioral consequences of mutations at *Nf1* have been characterized in these mice across a series of detailed studies. [39, 40, 136, 137] Spatial learning deficits in the hidden version of the Morris water maze have been replicated multiple times, with *Nf1*^{+/-} mice requiring more training trials than wild-type littermates to locate the position of the hidden platform; a performance suggestive of impaired hippocampal-based learning. While *Nf1*^{+/-} mice also show deficits in contextual conditioning, other forms of learning, such as classical conditioning and simple associative learning remain intact. [137] In addition to hippocampal-based learning deficits, *Nf1*^{+/-} mice have also been reported to experience attention impairments in a lateralized reaction time task (a measure designed to assess attention processes) and significantly reduced prepulse inhibition; a measure of sensory gating whereby a startle reflex is inhibited if preceded by a weak prestimulus. [40]

Hyperactivation of the RAS-MAPK signaling cascade has been proposed as the core mechanism underlying cognitive impairment in *Nf1*^{+/-} mice. An increase in RAS signaling causes an increase in activity-dependent GABA release, resulting in an imbalance between inhibitory and excitatory processes. [138] In the hippocampus, this has been shown to significantly reduce long term potentiation (LTP), a form of synaptic plasticity resulting in an increased strength of synaptic transmission that is required for effective learning and memory. It is a deficit in LTP that is thought to underlie the learning and memory impairments in *Nf1*^{+/-} mice. [40, 137, 139] This assertion has been reinforced by preclinical studies that have manipulated levels of active RAS and GABA inhibition. Costa and colleagues bred *Nf1*^{+/-} mice and mice deficient in active RAS (K-RAS^{+/-} mice) and tested both groups on a hidden water maze task. [136] Genetic manipulation to reduce RAS resulted in equivalent performance between the wild-type mice and *Nf1*^{+/-}/K-RAS^{+/-} mice, suggesting learning impairments in *Nf1*^{+/-} mice may be caused by excessive RAS activity. Pharmacological manipulation has also been used to decrease RAS hyperactivation. *Nf1*^{+/-} mice were given lovastatin; an agent commonly used to treat hyperlipidemia in children and adults which also decreases p21RAS isoprenylation. [40] After several days of treatment, *Nf1*^{+/-} mice medicated with lovastatin demonstrated improved performance relative to the placebo group on tasks of hippocampal-based learning (hidden version of the Morris water maze) and attention (lateralized reaction time task). Importantly, lovastatin also reversed deficits in LTP and normalized elevated p21RAS/MAPK activity, providing a physiological basis for the cognitive improvement. Learning deficits associated with a mutation at *Nf1* have also been rescued by a subthreshold dose of GABA_A antagonist. [137]

While cognitive impairments caused by dysregulation of the RAS-MAPK pathway have been demonstrated as one potentially reversible aspect of the NF1 cognitive phenotype, a series of recent experiments have identified the dopaminergic pathway as an alternative target for cognitive treatment by demonstrating that reduced dopamine (DA) in the striatum may underlie attention system dysfunction in *Nf1* mice (*Nf1*^{+/-} strain with GFAP⁺ cell bi-allelic *Nf1* gene inactivation; *Nf1* OPG mice). [140, 141] *Nf1* OPG mice exhibited reduced exploratory behaviours and deficits on measures of selective and non-selective attention within the context

of intact sensorimotor abilities. [140] Consistent with these deficits, high-performance liquid chromatography revealed significantly reduced dopamine levels in the striatum of these animals. Based on parallels between attention dysfunction in the *Nf1* OPG mice and attention deficits in children with NF1, the mice were treated with 20mg/kg of methylphenidate (MPH), a commonly used intervention for the behavioral symptoms of ADHD, which blocks striatal dopamine transporters [142] and significantly increases extracellular dopamine in the striatum. [143] MPH treatment rescued the non-selective attention deficit in these mice, as did L-dopa, a precursor to dopamine that acts on the dopaminergic but not serotonergic pathway, highlighting reduced DA as the primary etiology underlying the observed attention deficits in these mice. Neurotransmitter imaging analyses using [11C]-raclopride positron-emission tomography (PET) confirmed the striatal DA defect *in vivo* and demonstrated that improved behavioral functioning was accompanied by a significant increase in striatal DA following treatment with MPH. In contrast, treatments that target neurofibromin-regulated RAS and cAMP defects, such as lovastatin, did not improve the attention deficits or PET abnormalities in these *Nf1* OPG mice.

5.0. NEUROCOGNITIVE THERAPIES IN NF1

Notwithstanding the clear cross-species differences between mice and humans, results from murine studies have provided strong evidence that cognitive deficits, at least in part, are not solely attributable to developmental abnormalities of brain structure that may be more resistant to interventions and therapy. Rather, pharmacological “correction” of these molecular abnormalities, such as aberrant RAS signaling or reduced DA, has the potential to normalize aspects of the human NF1 cognitive phenotype.

To date, only one study has evaluated the effect of stimulant medication in children with NF1. Mautner and colleagues examined the effect of MPH on 20 children with NF1 with comorbid ADHD (NF1+ADHD) in a non-randomized open-label non-controlled trial.⁵⁵ Children with ADHD alone (n = 20) were included for comparison. While on medication, parents and teachers reported an improvement in the attention span of children with NF1+ADHD. From baseline to follow-up 12 months later, parent and teacher ratings indicated significant improvements in attention, anxiety/depression, and social competence scores of children with NF1+ADHD. There was also a significant improvement on a computerized task of attention/inhibition for the NF1+ADHD group, with fewer omission (inattention) and commission (impulsivity) errors, and faster reaction times. While providing some promising results, the limited design of this study, lack of appropriate control group and the small sample size limit its interpretation and generalizability to the NF1 population. Additionally, this study exclusively investigated the effect of MPH in children with ADHD (who also had NF1) and did not investigate the efficacy of MPH as a treatment for attention and executive deficits in NF1 (regardless of ADHD status). We are currently examining the efficacy of MPH in treating sustained attention and spatial working memory impairments in children and adolescents with NF1 in a placebo-controlled, two period crossover, double-blind, Phase II clinical trial (Australian New Zealand Clinical Trials Registry Number: ACTRN12611000765921). In addition to cognitive outcome measures, a combination of functional neuroimaging techniques

(fMRI and R-fMRI) is being employed to establish whether MPH modulates functional connectivity and task-based neural activity within specific regions of interest.

To date, two clinical trials have been published on statin treatment for cognitive impairment in NF1, with variable results. [144, 145] The aim of these studies has been to improve cognitive performance by normalizing RAS activation. In the first of these studies, 62 children with NF1 were randomized to either simvastatin or placebo using a permuted block 1:1 randomization. Randomization was double-blind and the treatment period was 12 weeks. Results revealed no significant differences between the simvastatin and placebo groups on any primary outcome measure, including delayed recall of the Rey Complex Figure, a speeded cancellation test, a prism adaptation measure and the mean brain apparent diffusion coefficient (based on MRI). There were a number of weaknesses with this study, however, that limits interpretation of the results. First, the treatment period was short, with the full dose (40mg/day) only given to half the treated cohort ($n = 15$) for four weeks. The dose was lower in the other treated patients (20mg/day for 12 weeks). Second, the study was not limited to participants with a specific type of cognitive impairment. As such, children were enrolled for treatment even though they did not necessarily demonstrate impairment on a primary outcome. Third, children with very low IQ scores (as low as 48) were also included in the study, making it very difficult to compare their results to patients with normal IQ. In the other study, Acosta and colleagues conducted a Phase I trial examining the safety and tolerability of lovastatin in children with NF1. [145] Neurocognitive assessments were also performed pre-and post- treatment. Twenty-four children with NF1 were recruited for the three month study; three were treated with 20mg/day; three commenced on 20mg/day and escalated to 30mg/day after two weeks; and 18 commenced on 20mg/day and escalated to 40mg/day after two weeks. Results revealed minimal side effects and the absence of any dose-limiting toxicity. Reliable change statistics indicated significant improvements in verbal and nonverbal memory exceeding those expected from test-retest or practice effects. The authors concluded that these observations could be analogous to those observed in NF1 mice treated with lovastatin. [40] As this was a Phase I trial, however, the primary aim of the authors was safety and tolerability, not neurocognitive outcome. Study limitations included open label design, small sample size, lack of an appropriate control group and inclusion of patients without defined cognitive impairment. As such, the conclusions that can be drawn from pre-post treatment cognitive assessment are limited. There are currently a number of statin trials nearing completion that are addressing a number of the limitations of these published studies (e.g., [clinicaltrials.gov](https://clinicaltrials.gov/ct2/show/study?term=NCT00853580) identifiers NCT00853580, NCT00352599).

In addition to pharmacological treatments, there are a number of non-pharmacological interventions that could be explored in children with NF1. To the best of our knowledge, no studies examining non-pharmacological treatments for cognitive impairment have been published in the NF1 literature. Two of the most promising areas of research in populations other than NF1 (e.g., learning disability, ADHD) have been computer-based training programs for working memory and reading impairment; both of which are core features of the NF1 phenotype. Over the past few years, several computer-based working memory training programs have been developed. The most well-known of these is Cogmed; a program that includes both visuospatial and verbal working memory training tasks in which the level of difficulty adapts throughout the treatment course. Cogmed has been used in several controlled trials in children and adolescents with ADHD and has been reported to result in significant improvements in visuospatial and verbal working memory, response inhibition as well as parent ratings of working memory and attention. [146-148] Importantly, treatment effects appear to

be maintained overtime. [147, 149] A recent meta-analysis, however, has suggested that although training programs such as Cogmed have positive effects tasks that are close to those trained; there is currently only limited evidence of the generalization of working memory training on skills such as verbal and nonverbal ability, word decoding and arithmetic. [150] A feasibility study of Cogmed in NF1 is currently underway at Children's National Medical Center, Washington DC.

There is also compelling evidence that systematic and explicit instruction in phonological awareness and phonics (knowledge of letter-to-sound relationships) results in significant gains for most children with reading impairment. [151, 152] Both phonological awareness and phonics are predictors of reading ability, and importantly, there are strong data to suggest that an emphasis of phonological awareness positively influences word reading outcomes. [153] Although the benefits of phonics intervention have been recognized in the general population, the unique cognitive profile of patients with NF1 may influence the efficacy of remediation programs in children with the condition. [73] To the best of our knowledge, there are currently two studies examining reading interventions in children with NF1. The goals of the first trial are to examine the efficacy of two different training programs in poor readers with NF1 and to compare the treatment response of the NF1 group to children with a reading disability from the general population who also complete the training programs (clinicaltrials.gov identifier NCT00624234). In the second study, the efficacy of a commercially-available computer-based phonics training program is being examined in children with NF1 and phonological reading impairment (Australian New Zealand Clinical Trials Registry Number: ACTRN12611000779976). The high frequency of reading impairment in children with NF1 and the negative influence reduced reading ability has on quality of life, school achievement and peer relations, makes treatment studies such as these critical. Poor reading skills have also been associated with poorer health outcomes, decreased responsiveness to health education, less health knowledge and poorer health status. [154] As such, the importance of improving reading and literacy skills in individuals with NF1 – a progressive disorder associated with significant physical manifestations throughout the lifespan – cannot be overestimated.

CONCLUSION

Cognitive and learning impairments are the greatest cause of lifetime morbidity in individuals with NF1. The literature presented in this chapter outlines common neurocognitive and behavioral impairments, academic difficulties and developmental considerations. The goal of research into cognitive deficits in NF1 is the development of successful remediation programs to prevent the pattern of school failure that is common in children with the disorder. Currently, therapeutic intervention is likely to rely on symptomatic educational intervention. This will often involve the provision of a range of strategies to be implemented at home and in the classroom which tailor the child's learning environment to their cognitive and behavioral strengths and weaknesses. Most of these techniques, however, are not evidence-based and it is often unclear how effectively they are being implemented.

Human neuroimaging and murine studies are beginning to provide critical evidence for the pathobiological basis of NF1-related cognitive impairments. While structural abnormalities, such as an enlarged corpus callosum, have been shown to impact on cognitive performance,

they are developmental in origin and given the fundamental relationship between brain structure and function, may be difficult to reverse in childhood with pharmacological intervention. Murine studies, however, provide compelling evidence that mutation at *Nf1* results in molecular anomalies, such as RAS hyperactivation and DA abnormalities which result in specific cognitive deficits. Taken together, these preclinical studies suggest that children with NF1-associated cognitive impairments are likely to comprise a heterogeneous population of individuals with distinct molecular etiologies — but importantly, that pharmacological correction of these molecular abnormalities has the potential normalize aspects of the NF1 cognitive phenotype. Although very few treatment studies have been published in humans with NF1, there are a number of pharmacological and non-pharmacological clinical trials currently underway that aim to overcome the limitations of previous studies.

REFERENCES

- [1] Evans DG, Howard E, Giblin C, et al. Birth incidence and prevalence of tumor-prone syndromes: Estimates from a UK family genetic register service. *American Journal of Medical Genetics* 2010;152A:327-32.
- [2] Wallace MR, Marchuk DA, Andersen LB, et al. Type 1 neurofibromatosis gene: identification of a large transcript disrupted in three NF1 patients. *Science* 1990; 249:181-86.
- [3] Williams VC, Lucas J, Babcock MA, Gutmann DH, Korf B, Maria BL. Neurofibromatosis type 1 revisited. *Pediatrics* 2009;123:124-33.
- [4] Friedman JM. Neurofibromatosis 1: clinical manifestations and diagnostic criteria. *Journal of Child Neurology* 2002;17:548-54.
- [5] Neurofibromatosis. Conference Statement. National Institutes of Health Consensus Development Conference. *Archives of Neurology* 1988;45:575-8.
- [6] Riccardi V. Skeletal system. In: *Neurofibromatosis: Phenotype, Natural History, and Pathogenesis*. 3rd ed. Baltimore, MD: Johns Hopkins University Press; 1999:250-73.
- [7] Hyman SL, Shores EA, North KN. The nature and frequency of cognitive deficits in children with neurofibromatosis type 1. *Neurology* 2005;65:1037-44.
- [8] Lehtonen A, Howie E, Trump D, Huson SM. Behaviour in children with neurofibromatosis type 1: cognition, executive function, attention, emotion, and social competence. *Developmental Medicine and Child Neurology* 2012; DOI: 10.1111/j.1469-8749.2012.04399.x.
- [9] North KN. Neurofibromatosis 1 in childhood. *Seminars in Pediatric Neurology* 1998; 5:231-42.
- [10] Cole WG, Myers NA. Neurofibromatosis in childhood. *Australian and New Zealand Journal of Surgery* 1978;48:360-5.
- [11] Crowe FW, Schull WJ, Neel JV. *A clinical pathological and genetic study of multiple neurofibromatosis*. Springfield: Charles Thomas; 1956.
- [12] Krab LC, Aarsen FK, Goede-Bolder A, et al. Impact of neurofibromatosis type 1 on school performance. *Journal of Child Neurology* 2008;23:1002-10.

- [13] Sangster J, Shores EA, Watt S, North KN. The cognitive profile of preschool-aged children with neurofibromatosis type 1. *Child Neuropsychology* 2010;1:1-16.
- [14] Ferner RE, Hughes RAC, Weinman J. Intellectual impairment in neurofibromatosis 1. *Journal of the Neurological Sciences* 1996;138:125-33.
- [15] Pride N, Payne JM, North KN. The impact of ADHD on cognitive and academic functioning in children with NF1. *Developmental Neuropsychology* 2012;37:590-600.
- [16] Payne JM, Arnold SS, Pride NA, North KN. Does attention-deficit-hyperactivity disorder exacerbate executive dysfunction in children with neurofibromatosis type 1? *Developmental Medicine and Child Neurology* 2012;34:898-904.
- [17] Roy A, Roulin J-L, Charbonnier V, et al. Executive dysfunction in children with neurofibromatosis type 1: a study of action planning. *Journal of the International Neuropsychological Society* 2010;16:1056-3.
- [18] Legius E, Descheemaeker MJ, Steyaert J, et al. Neurofibromatosis type 1 in childhood: correlation of MRI findings with intelligence. *Journal of Neurology, Neurosurgery and Psychiatry* 1995;59:638-40.
- [19] Hyman SL, Shores AE, North KN. Learning disabilities in children with neurofibromatosis type 1: subtypes, cognitive profile, and attention-deficit-hyperactivity disorder. *Developmental Medicine and Child Neurology* 2006;48:973-7.
- [20] North K, Joy P, Yuille D, Cocks N, Hutchins P. Cognitive function and academic performance in children with neurofibromatosis type 1. *Developmental Medicine and Child Neurology* 1995;37:427-36.
- [21] Payne JM, Hyman SL, Shores EA, North KN. Assessment of executive function and attention in children with neurofibromatosis type 1: Relationships between cognitive measures and real-world behavior. *Child Neuropsychology* 2011;17:313-29.
- [22] Eldridge R, Denckla MB, Bien E, et al. Neurofibromatosis type 1 [Recklinghausen's disease]: neurologic and cognitive assessment with sibling controls. *American Journal of Disabled Children* 1989;143.
- [23] Moore BD, 3rd, Ater JL, Needle MN, Slopis J, Copeland DR. Neuropsychological profile of children with neurofibromatosis, brain tumor, or both. *Journal of Child Neurology* 1994;9:368-77.
- [24] Legius E, Descheemaeker MJ, Spaepen A, Casaer P, Fryns JP. Neurofibromatosis type 1 in childhood: a study of the neuropsychological profile in 45 children. *Genetic Counselling* 1994;5:51-60.
- [25] Eliason MJ. Neurofibromatosis: implications for learning and behavior. *Developmental and Behavioral Pediatrics* 1986;7:175-9.
- [26] Eliason MJ. Neuropsychological patterns: neurofibromatosis compared to developmental learning disorder. *Neurofibromatosis* 1988;1:17-25.
- [27] Ardila A, Pineda D, Rosselli M. Correlation between intelligence test scores and executive function measures. *Archives of Clinical Neuropsychology* 2000;15:31-6.
- [28] Dilts CV, Carey JC, Kircher JC, et al. Children and adolescents with neurofibromatosis 1: a behavioral phenotype. *Journal of Developmental and Behavioral Pediatrics* 1996; 17:229-39.
- [29] Hofman KJ, Harris EL, Bryan RN, Denckla MB. Neurofibromatosis type 1: the cognitive phenotype. *Journal of Pediatrics* 1994;124:S1-8.
- [30] Mazzocco MMM, Turner JE, Denckla MB, Hofman KJ, Scanlon DC, Vellutino FR. Language and reading deficits associated with Neurofibromatosis Type 1: evidence for

- a not-so-nonverbal learning disability. *Developmental Neuropsychology* 1995;11:503-22.
- [31] Clements-Stephens AM, Rimrodt SL, Gaur P, Cutting LE. Visuospatial processing in children with neurofibromatosis type 1. *Neuropsychologia* 2008;46:690-7.
 - [32] Joy P, Roberts C, North K, de Silva M. Neuropsychological function and MRI abnormalities in neurofibromatosis type 1. *Dev. Med. Child Neurol.* 1995;37:906-14.
 - [33] Schrimsher GW, Billingsley RL, Slopis JM, Moore BD. Visual-spatial performance deficits in children with neurofibromatosis type-1. *American Journal of Medical Genetics Part A* 2003;120A:326-30.
 - [34] Levine TM, Materek A, Abel J, O'Donnell M, Cutting LE. Cognitive profile of neurofibromatosis type 1. *Seminars in Pediatric Neurology* 2006;13:8-20.
 - [35] Ozonoff S. Cognitive impairment in neurofibromatosis type 1. *American Journal of Medical Genetics* 1999;89:45-52.
 - [36] North K. Neurofibromatosis type 1. *American Journal of Medical Genetics* 2000;97:119-27.
 - [37] Ashcraft MH. *Human Memory and Cognition*. New York: Harper Collins; 1994.
 - [38] Tulving E. *Elements of episodic memory*. New York: Oxford University Press; 1983.
 - [39] Silva AJ, Frankland PW, Marowitz Z, et al. A mouse model for the learning and memory deficits associated with neurofibromatosis type I. *Nature Genetics* 1997; 15:281-4.
 - [40] Li W, Cui Y, Kushner SA, et al. The HMG-CoA reductase inhibitor lovastatin reverses the learning and attention deficits in a mouse model of neurofibromatosis type 1. *Current Biology* 2005;15:1961-7.
 - [41] Ferner RE, Hughes RA, Weinman J. Intellectual impairment in neurofibromatosis 1. *Journal of the Neurological Sciences* 1996;138:125-33.
 - [42] Moore BD, 3rd, Slopis JM, Jackson EF, De Winter AE, Leeds NE. Brain volume in children with neurofibromatosis type 1: relation to neuropsychological status. *Neurology* 2000;54:914-20.
 - [43] Hyman SL, Gill DS, Shores EA, et al. Natural history of cognitive deficits and their relationship to MRI T2-hyperintensities in NF1. *Neurology* 2003;60:1139-45.
 - [44] Ullrich NJ, Ayr L, Leafner E, Irons MB, Rey-Casserly C. Pilot study of a novel computerised task to assess spatial learning in children and adolescents with neurofibromatosis type 1. *Journal of Child Neurology* 2010;2010:1195-202.
 - [45] Payne JM, Barton B, Shores EA, North KN. Paired associate learning in children with neurofibromatosis type 1: implications for clinical trials. *Journal of Neurology* 2012; DOI 10.1007/s00415-012-6620-5.
 - [46] Fray PJ, Robbins TW, Sahakian BJ. Neuropsychiatric applications of CANTAB. *International Journal of Geriatric Psychiatry* 1996;11:329-36.
 - [47] Fowler KS, Saling MM, Conway EL, Semple JM, Louis WJ. Paired associate performance in the early detection of DAT. *Journal of the International Neuropsychological Society* 2002;8:58-71.
 - [48] Fowler KS, Saling MM, Conway EL, Semple JM, Louis WJ. Computerized neuropsychological tests in the early detection of dementia: prospective findings. *Journal of the International Neuropsychological Society* 1997;3:139-46.
 - [49] Owen AM, Sahakian BJ, Semple J, Polkey CE, Robbins TW. Visuo-spatial short-term recognition memory and learning after temporal lobe excisions, frontal lobe excisions or amygdalo-hippocampectomy in man. *Neuropsychologia* 1995;33:1-24.

- [50] Von Huben SN, Davis SA, Lay CC, Katner SN, Crean RD, Taffe MA. Differential contributions of dopaminergic D1- and D2-like receptors to cognitive function in rhesus monkeys. *Psychopharmacology* 2006;188:586-96.
- [51] Nicoll RA, Malenka RC. Expression mechanisms underlying NMDA receptor-dependent long-term potentiation. *Annals of the New York Academy of Sciences* 1999; 868:515-25.
- [52] Corbetta M, Shulman GL. Control of goal-directed and stimulus-driven attention in the brain. *Nature Reviews Neuroscience* 2002;3:201-15.
- [53] Friedman-Hill SR, Wagman MR, Gex SE, Pine DS, Leibenluft E, Ungerleider LG. What does distractibility in ADHD reveal about mechanisms for top-down attentional control? *Cognition* 2010;115:93-103.
- [54] Isenberg JC, Templer A, Gao F, Titus JB, Gutmann DH. Attention skills in children with neurofibromatosis type 1. *Journal of Child Neurology* 2012: DOI: 10.1177/0883073812439435.
- [55] Mautner VF, Kluwe L, Thakker SD, Lark RA. Treatment of ADHD in neurofibromatosis type 1. *Developmental Medicine and Child Neurology* 2002;44:164-70.
- [56] Gilboa Y, Rosenblum S, Fattal-Valevski A, Toledano-Alhadeef H, Rizzo A, Josman N. Using a Virtual Classroom environment to describe the attention deficits profile of children with Neurofibromatosis type 1. *Research in Developmental Disabilities* 2011; 32:2608-13.
- [57] Denckla MB, Reiss AL. Prefrontal-subcortical circuits in developmental disorders. In: Krasnegor NA, Lyon GR, Goldman-Rakic PS, eds. *Development of the Prefrontal Cortex: Evolution, Neurobiology, and Behavior*. Baltimore, MD: Brookes; 1997.
- [58] Stuss DT, Benson DF. New York: Raven Press; 1986.
- [59] Barkley RA. Behavioral inhibition, sustained attention, and executive functions: constructing a unifying theory of ADHD. *Psychological Bulletin* 1997;121:65-94.
- [60] Hill EL. Executive dysfunction in autism. *Trends in Cognitive Sciences* 2004;8:26-32.
- [61] Kerns JG, Nuechterlein KH, Braver TS, Barch DM. Executive functioning component mechanisms and schizophrenia. *Biological Psychiatry* 2008;64:26-33.
- [62] Samango-Sprouse C. Frontal lobe development in childhood. In: Miller BL, Cummings JL, eds. *The Human Frontal Lobes*. New York: The Guilford Press; 1999:576-96.
- [63] Stine SB, Adams WV. Learning problems in neurofibromatosis patients. *Clinical Orthopaedics and Related Research* 1989;245:43-8.
- [64] Huijbregts S, Swaab H, de Sonnevile L. Cognitive and motor control in neurofibromatosis type I: influence of maturation and hyperactivity-inattention. *Developmental Neuropsychology* 2010;35:737-51.
- [65] Descheemaeker MJ, Ghesquière P, Symons H, Fryns JP, Legius E. Behavioural, academic and neuropsychological profile of normally gifted Neurofibromatosis type 1 children. *Journal of Intellectual Disability Research* 2005;49:33-46.
- [66] Rowbotham I, Pit-ten Cate IM, Sonuga-Barke EJ, Huijbregts SC. Cognitive control in adolescents with neurofibromatosis type 1. *Neuropsychology* 2009;23:50-60.
- [67] Huijbregts S, Jahja R, De Sonnevile LEO, De Breij S, Swaab-Barneveld H. Social information processing in children and adolescents with neurofibromatosis type 1. *Developmental Medicine and Child Neurology* 2010;52:620-5.

- [68] Anderson VA, Anderson P, Northam E, Jacobs R, Mikiewicz O. Relationship between cognitive and behavioural measures of executive function in children with brain disease. *Child Neuropsychology* 2002;8:231-40.
- [69] Chan RCK, Shum D, Touloupoulou T, Chen EYH. Assessment of executive functions: review of instruments and identification of critical issues. *Archives of Clinical Neuropsychology* 2008;23:201-16.
- [70] Chaytor N, Schmitter-Edgecombe M. The ecological validity of neuropsychological tests: a review of the literature of everyday cognitive skills. *Neuropsychology Review* 2003;13.
- [71] Brewer VR, Moore BD, 3rd, Hiscock M. Learning disability subtypes in children with neurofibromatosis. *Journal of Learning Disabilities* 1997;30:521-33.
- [72] Cutting LE, Koth CW, Denckla MB. How children with neurofibromatosis type 1 differ from "typical" learning disabled clinic attenders: nonverbal learning disabilities revisited. *Developmental Neuropsychology* 2000;17:29-47.
- [73] Cutting LE, Levine TM. Cognitive profile of children with neurofibromatosis type 1 and reading disabilities. *Child Neuropsychology* 2010;16:417-32.
- [74] Watt SE, Shores A, North KN. An examination of lexical and sublexical reading skills in children with neurofibromatosis type 1. *Child Neuropsychology* 2008;14:401-18.
- [75] Moore B. Potential influences on mathematical difficulties in children and adolescents with neurofibromatosis, type 1. *Developmental Disabilities Research Reviews* 2009;15:45-51.
- [76] Billingsley RL, Slopis JM, Swank PR, Jackson EF, Moore BD. Cortical morphology associated with language function in neurofibromatosis, type 1. *Brain and Language* 2003;85:125-39.
- [77] Billingsley RL, Jackson EF, Slopis JM, Swank PR, Mahankali S, Moore BD. Functional MRI of visual-spatial processing in neurofibromatosis, type I. *Neuropsychologia* 2004;42:395-404.
- [78] Bawden H, Dooley J, Buckley D, et al. MRI and nonverbal cognitive deficits in children with neurofibromatosis 1. *Journal of Clinical and Experimental Neuropsychology* 1996;18:784-92.
- [79] Polanczyk G, de Lima MS, Horta BL, Biederman J, Rohde LA. The worldwide prevalence of ADHD: a systematic review and metaregression analysis. *American Journal of Psychiatry* 2007;164:942-48.
- [80] Lidzba K, Granstrom S, Lindenau J, Mautner VF. The adverse influence of attention-deficit disorder with or without hyperactivity on cognition in neurofibromatosis type 1. *Developmental Medicine and Child Neurology* 2012; doi:10.1111/j.1469-8749.2012.04396.x.
- [81] Willcutt EG, Doyle AE, Nigg JT, Faraone SV, Pennington BF. Validity of the executive function theory of attention-deficit/hyperactivity disorder: a meta-analytic review. *Biological Psychiatry* 2005;57:1336-46.
- [82] Durston S. A review of the biological bases of ADHD: what have we learned from imaging studies? *Mental Retardation and Developmental Disabilities Research Reviews* 2003;9:184-95.
- [83] Huijbregts S. Cognitive-behavioral phenotype of comorbid disorder? The case of attention-deficit-hyperactivity disorder in neurofibromatosis type 1. *Developmental Medicine and Child Neurology* 2012;doi:10.1111/j.1469-8749.2012.04396.x.

- [84] Durston S, Hulshoff-Pol HE, Schnack HG, et al. Magnetic resonance imaging of boys with attention-deficit/hyperactivity disorder and their unaffected siblings. *Journal of the American Academy of Child and Adolescent Psychiatry* 2004;43:332-40.
- [85] Pride N, Payne JM, Webster R, Shores EA, Rae C, North KN. Corpus callosum morphology and its relationship to cognitive function in neurofibromatosis type 1. *Journal of Child Neurology* 2010;25:834-41.
- [86] Barton B, North K. Social skills of children with neurofibromatosis type 1. *Developmental Medicine and Child Neurology* 2004;46:553-63.
- [87] Johnson NS, Saal HM, Lovell AM, Schorry EK. Social and emotional problems in children with neurofibromatosis type 1: evidence and proposed interventions. *Journal of Pediatrics* 1999;134:767-72.
- [88] Noll RB, Reiter-Purtill J, Moore BD, et al. Social, emotional, and behavioral functioning of children with NF1. *American Journal of Medical Genetics Part A* 2007;143A:2261-73.
- [89] Pride NA, Payne JM, Barton B, Crawford H, North KN. Social competence in adults with neurofibromatosis type 1. in preparation.
- [90] Huijbregts SCJ, de Sonnevile LMJ. Does cognitive impairment explain behavioral and social problems of children with neurofibromatosis type 1? *Behavior Genetics* 2011;41:430-6.
- [91] Adolphs R. The neurobiology of social cognition. *Current Opinion in Neurobiology* 2001;11:231-9.
- [92] Pride NA, Korgaonkar M, Barton B, Payne JM, North KN. The genetic and neuroanatomical basis of social dysfunction - lessons from neurofibromatosis type 1. submitted.
- [93] Lorenzo J, Barton B, Acosta MT, North K. Mental, Motor, and Language Development of Toddlers with Neurofibromatosis Type 1. *Journal of Pediatrics* 2010.
- [94] Thompson HL, Viskochil DH, Stevenson DA, Chapman KL. Speech-language characteristics of children with neurofibromatosis type 1. *American Journal of Medical Genetics Part A* 2010;152A:284-90.
- [95] Zöller MET, Rembeck B, Bäckman L. Neuropsychological deficits in adults with Neurofibromatosis type 1. *Acta Neurologica Scandinavica* 1997;95:225-32.
- [96] Lorch M, Ferner R, Golding J, Whurr R. The nature of speech and language impairment in adults with neurofibromatosis 1. *Journal of Neurolinguistics* 1999;12:157-65.
- [97] Pavol M, Hiscock M, Massman P, Moore B, Foorman B, Meyers C. Neuropsychological function in adults with Von Recklinghausen's neurofibromatosis. *Developmental Neuropsychology* 2006;29:509 - 26.
- [98] Riccardi VM, Eichner JE. *Neurofibromatosis. Phenotype, Natural History and Pathogenesis*. Baltimore: John Hopkins University Press; 1986.
- [99] Fisch GS, Carpenter N, Howard-Peebles PN, et al. Studies of age-correlated features of cognitive-behavioral development in children and adolescents with genetic disorders. *American Journal of Medical Genetics Part A* 2007;143A:2478-89.
- [100] Cutting LE, Huang GH, Zeger S, Koth CW, Thompson RE, Denckla MB. Growth curve analyses of neuropsychological profiles in children with neurofibromatosis type 1: specific cognitive tests remain "spared" and "impaired" over time. *Journal of the International Neuropsychological Society* 2002;8:838-46.

- [101] Payne JM, Pickering T, Porter M, et al. Longitudinal assessment of cognition and T2-hyperintensities in NF1: an 18 year study. submitted.
- [102] Payne JM, Moharir MD, Webster R, North KN. Brain structure and function in neurofibromatosis type 1: current concepts and future directions. *Journal of Neurology, Neurosurgery and Psychiatry* 2010;81:304-9.
- [103] Said SM, Yeh TL, Greenwood RS, Whitt JK, Tupler LA, Krishnan KR. MRI morphometric analysis and neuropsychological function in patients with neurofibromatosis. *Neuroreport* 1996;7.
- [104] Moore BD, 3rd, Slopis JM, Schomer D, Jackson EF, Levy BM. Neuropsychological significance of areas of high signal intensity on brain MRIs of children with neurofibromatosis *Neurology* 1996;46:1660-8.
- [105] Greenwood RS, Tupler LA, Whitt JK, et al. Brain morphometry, T2-weighted hyperintensities, and IQ in children with neurofibromatosis type 1. *Archives of Neurology* 2005;62:1904-8.
- [106] Steen RG, Taylor JS, Langston JW, et al. Prospective evaluation of the brain in asymptomatic children with neurofibromatosis type 1: relationship of macrocephaly to T1 relaxation changes and structural brain abnormalities. *American Journal of Neuroradiology* 2001;22:810-7.
- [107] Cutting LE, Koth CW, Burnette CP, Abrams MT, Kaufmann WE, Denckla MB. Relationship of cognitive functioning, whole brain volumes, and T2-weighted hyperintensities in neurofibromatosis-1. *Journal of Child Neurology* 2000;15:157-60.
- [108] Said SM, Yeh TL, Greenwood RS, Whitt JK, Tupler LA, Krishnan KR. MRI morphometric analysis and neuropsychological function in patients with neurofibromatosis. *Neuroreport* 1996;7:1941-4.
- [109] Greenwood RS, Tupler LA, Whitt JK, et al. Brain morphometry, T2-weighted hyperintensities, and IQ in children with neurofibromatosis type 1. *Archives of Neurology* 2005;62:1904-8.
- [110] Billingsley RL, Schrimsher GW, Jackson EF, Slopis JM, Moore BD, 3rd. Significance of planum temporale and planum parietale morphologic features in neurofibromatosis type 1. *Archives of Neurology* 2002;59:616-22.
- [111] Kayl AE, Moore BD, Slopis JM, F. JE, Leeds NE. Quantitative morphology of the corpus callosum in children with neurofibromatosis and attention-deficit hyperactivity disorder. *Journal of Child Neurology* 2000;15:90-6.
- [112] Dubovsky EC, Booth TN, Vezina G, Samango-Sprouse CA, Palmer KM, Brasseux CO. MR imaging of the corpus callosum in pediatric patients with neurofibromatosis type 1. *American Journal of Neuroradiology* 2001;22:190-5.
- [113] Moore BD, 3rd, Slopis JM, Jackson EF, De Winter AE, Leeds NE. Brain volume in children with neurofibromatosis type 1: relation to neuropsychological status. *Neurology* 2000;54:914-20.
- [114] Moore BD, 3rd, Slopis JM, Schomer D, Jackson EF, Levy BM. Neuropsychological significance of areas of high signal intensity on brain MRIs of children with neurofibromatosis. *Neurology* 1996;46:1660-8.
- [115] Gill DS, Hyman SL, Steinberg A, North KN. Age-related findings on MRI in neurofibromatosis type 1. *Pediatric Radiology* 2006;36:1048-56.

- [116] Hyman SL, Gill DS, Shores EA, Steinberg A, North KN. T2 hyperintensities in children with neurofibromatosis type 1 and their relationship to cognitive functioning. *Journal of Neurology Neurosurgery and Psychiatry* 2007;78:1088-91.
- [117] Denckla MB, Hofman KJ, Mazzocco MM, et al. Relationship between T2-weighted hyperintensities (unidentified bright objects) and lower IQs in children with neurofibromatosis-1. *American Journal of Medical Genetics* 1996;67:98-102.
- [118] North K, Joy P, Yuille D, et al. Specific learning disability in children with neurofibromatosis type 1: significance of MRI abnormalities. *Neurology* 1994;44:878-83.
- [119] Goh WH, Khong PL, Leung CSY, Wong VCN. T2-weighted hyperintensities (unidentified bright objects) in children with neurofibromatosis type 1: their impact on cognitive function. *Journal of Child Neurology* 2004;19:853-8.
- [120] Chabernaud C, Sirinelli D, Barbier C, et al. Thalamo-striatal T2-weighted hyperintensities (unidentified bright objects) correlate with cognitive impairments in neurofibromatosis type 1. *Developmental Neuropsychology* 2009;34:736-48.
- [121] Huettel SA, Song AW, McCarthy G. *Functional Magnetic Resonance Imaging*. 2nd ed. Massachusetts: Sinauer; 2009.
- [122] Billingsley RL, Jackson EF, Slopis JM, Swank PR, Mahankali S, Moore BD. Functional magnetic resonance imaging of phonologic processing in neurofibromatosis 1. *Journal of Child Neurology* 2003;18:731-40.
- [123] Shilyansky C, Karlsgodt KH, Cummings DM, et al. Neurofibromin regulates corticostriatal inhibitory networks during working memory performance. *Proceedings of the National Academy of Sciences of the United States of America* 2010;107:13141-6.
- [124] De Luca M, Smith S, De Stefano N, Federico A, Matthews PM. Blood oxygenation level dependent contrast resting state networks are relevant to functional activity in the neocortical sensorimotor system. *Experimental Brain Research* 2005;167:587-94.
- [125] Vincent JL, Z. SA, Fox MD, et al. Coherent spontaneous activity identifies a hippocampal-parietal memory network. *Journal of Neurophysiology* 2006;96:3517-31.
- [126] Hampson M, Driesen NR, Skudlarski P, Gore JC, Constable RT. Brain Connectivity Related to Working Memory Performance. *Journal of Neuroscience* 2006;26:13338-43.
- [127] Seeley WW, Menon V, Schatzberg AF, et al. Dissociable intrinsic connectivity networks for salience processing and executive control. *Journal of Neuroscience* 2007;27:2349-56.
- [128] Chabernaud C, Mennes M, Kardel PG, et al. Lovastatin regulates brain spontaneous low-frequency brain activity in neurofibromatosis type 1. *Neuroscience Letters* 2012;25:28-33.
- [129] Brown JA, Diggs-Andrews KA, Gianino SM, Gutmann DH. Neurofibromatosis-1 heterozygosity impairs CNS neuronal morphology in a cAMP/PKA/ROCK-dependent manner. *Molecular and Cellular Neuroscience* 2012;49:13-22.
- [130] Xu H, Gutmann DH. Mutations in the GAP-related domain impair ability of neurofibromin to associate with microtubules. *Brain Research* 1997;759:149-52.
- [131] Dasgupta B, Yi Y, Weber JD, Gutmann DH. Proteomic analysis reveals hyperactivation of the mammalian target of rapamycin pathway in neurofibromatosis 1- associated human and mouse brain tumors. *Cancer Research* 2005;65:2755-60.

- [132] Johannessen CM, Reczek EE, James MF, Brems H, Legius E, Cichowski K. The NF1 tumor suppressor critically regulates TSC2 and mTOR. *Proceedings of the National Academy of Sciences of the United States of America* 1995;102:8573-8.
- [133] Xu GF, Lin B, Tanaka K, et al. The catalytic domain of the neurofibromatosis type 1 gene product stimulates ras GTPase and complements ira mutants of *S. cerevisiae*. *Cell* 1990;63:835-41.
- [134] Martin GA, Viskochil D, Bollag G, et al. The GAP-related domain of the neurofibromatosis type 1 gene product interacts with ras p21. *Cell* 1990;63:843-9.
- [135] Weiss B, Bollag G, Shannon K. Hyperactive Ras as a therapeutic target in neurofibromatosis type 1. *American Journal of Medical Genetics* 1999;89:14-22.
- [136] Costa RM, Federov NB, Kogan JH, et al. Mechanism for the learning deficits in a mouse model of neurofibromatosis type 1. *Nature* 2002;415:526-30.
- [137] Cui Y, Costa RM, Murphy GG, et al. Neurofibromin regulation of ERK signaling modulates GABA release and learning. *Cell* 2008;135:549-60.
- [138] Shilyansky C, Lee YS, Silva AJ. Molecular and cellular mechanisms of learning disabilities: a focus on NF1. *Annual Review of Neuroscience* 2010;33:221-43.
- [139] Neves G, Cooke SF, Bliss TV. Synaptic plasticity, memory and the hippocampus: a neural network approach to causality. *Nature Reviews Neuroscience* 2008;9:65-75.
- [140] Brown JA, Emmett RJ, White CR, et al. Reduced striatal dopamine underlies the attention system dysfunction in neurofibromatosis-1 mutant mice. *Human Molecular Genetics* 2010;19:4515-28.
- [141] Brown JA, Xu J, Diggs-Andrews KA, Wozniak DF, Mach RH, Gutmann DH. PET imaging for attention deficit preclinical drug testing in neurofibromatosis- 1 mice. *Experimental Neurology* 2011;232:333-8.
- [142] Volkow ND, Wang GJ, Fowler JS, et al. Dopamine transporter occupancies in the human brain induced by the therapeutic doses of oral methylphenidate. *American Journal of Psychiatry* 1998;155:1325-31.
- [143] Volkow ND, Wang GJ, Fowler JS, et al. Therapeutic doses of oral methylphenidate significantly increase extracellular dopamine in the human brain. *Journal of Neuroscience* 2001;21:RC121.
- [144] Krab LC, de Goede-Bolder A, Aarsen FK, et al. Effect of simvastatin on cognitive functioning in children with neurofibromatosis type 1: a randomized controlled trial. *Journal of the American Medical Association* 2008;300:287-94.
- [145] Acosta MT, Kardel PG, Walsh KS, Rosenbaum KN, Gioia GA, Packer RJ. Lovastatin as treatment for neurocognitive deficits in neurofibromatosis type 1: Phase I Study. *Pediatric Neurology* 2011;45:241-5.
- [146] Klingberg T, Forssberg H, Westerberg H. Training of working memory in children with ADHD. *Journal of Clinical and Experimental Neuropsychology* 2002;24:781-91.
- [147] Klingberg T, Fernell E, Olesen PJ, et al. Computerized training of working memory in children with ADHD - A randomized, controlled trial. *Journal of the American Academy of Child and Adolescent Psychiatry* 2005;44:177-86.
- [148] Westerberg H, Jacobaeus H, Hirvikoski T, et al. Computerized working memory training after stroke—A pilot study. *Brain Injury* 2007;21:21-9.
- [149] Holmes J, Gathercole SE, Dunning DL. Adaptive training leads to sustained enhancement of poor working memory in children. *Developmental Science* 2009;12:F9-F15.

- [150] Melby-Lervåg M, Hulme C. Is working memory training effective? A meta-analytic review. *Developmental Psychology* 2012;DOI:10.1037/a0028228.
- [151] Ball E. Phonological awareness: implications for whole language and emergent literacy programs. *Topics in Language Disorders* 1997;17:14-26.
- [152] Ehri LC, Nunes SR, Willows DM, Schuster BV, Yaghoub-Zadeh Z, Shanahan T. Phonemic awareness instruction helps children learn to read: evidence from the National Reading Panel's meta-analysis. *Reading Research Quarterly* 2001;36:250-87.
- [153] Cavanaugh C, Kim A, Wanzek J, Vaughn S. Kindergarten reading interventions for at-risk students: twenty years of research. *Learning Disabilities: A Contemporary Journal* 2004;2:9-21.
- [154] DeWalt DA, Berkman NA, Sheridan S, Lohr KN, Pignone MP. Literacy and health outcomes: a systematic review of the literature. *Journal of General Internal Medicine* 2004;19:1228-39.

Chapter 113

COMMUNICATION DISORDERS ASSOCIATED WITH NEUROFIBROMATOSIS TYPE 1

***Heather L. Thompson¹, David A. Stevenson², Sean M. Redmond¹,
Bruce L. Smith¹ and David H. Viskochil²***

¹University of Utah, College of Health,
Department of Communication Sciences and Disorders,
Salt Lake City, UT, US

²University of Utah, Department of Pediatrics,
Division of Medical Genetics, Salt Lake City, UT, US

ABSTRACT

The purpose of this chapter is to provide an overview of communication disorders observed among children and adults with Neurofibromatosis type 1 (NF1). In each section, terminology, prevalence and information regarding assessment for individuals in the general population will be presented. The literature describing communication disorders observed in the population of individuals with NF1 will be reviewed. Following a review of the literature, options for speech-language assessments for individuals with NF1 will be discussed.

INTRODUCTION

The purpose of this chapter is to provide an overview of communication disorders among children and adults with neurofibromatosis type 1 (NF1). A communication disorder occurs when an individual is unable to comprehend or produce verbal, nonverbal, or graphic messages [1]. About 12 to 13% of school-aged children exhibit a communication disorder [2]. Approximately 1.5 million school-aged children receive special services for speech/language disorders annually [3]. Children with NF1 often exhibit communication disorders in the areas of voice and *resonance* [4], *articulation/phonology* [5, 6] and expressive and receptive language [4], negatively impacting communication throughout development. While some

individuals with NF1 may be unaffected, others exhibit a delay or disorder in one or multiple areas of communication. In light of this variability, all collateral areas of communicative functioning should be assessed by a speech-language pathologist.

NF1 AND MANIFESTATIONS OF THE HEAD AND NECK

NF1 is an autosomal dominant neurocutaneous disorder with a prevalence of approximately 1 in 3000 [7, 8]. NF1 is caused by mutations in the *NF1* gene on chromosome 17q11.2 [9]. Key clinical manifestations of NF1 include café-au-lait macules, axillary freckling, Lisch nodules, optic nerve pathway tumors and distinctive bone abnormalities [10]. The prognosis for individuals with NF1 is highly variable. While approximately 59% of individuals will not develop significant medical complications as a result of NF1, the remainder will develop one or more complications [11].

Studies of patients with NF1 have shown that approximately 87% of individuals experience head and neck symptoms including tumors of the eyelids/orbit and face/cheek [12]. Optic nerve gliomas are the most frequently observed intracranial tumor [13]. The impact of head and neck manifestations can be relatively benign (i.e., macrocephaly), or significant, resulting in problems with hearing, vision, eyelid motility, and facial expression [12].

Approximately 72% of individuals with NF1 experience oral manifestations including oral neurofibromas, enlarged fungiform papillae, and wide inferior alveolar canals [14]. Dysphagia may also occur [15]. Cases of patients with symptoms of NF1 and speech problems [13] or hoarseness [16] have been reported. While the etiology of speech sound disorder in this population is unknown, it has been suggested that speech problems may be secondary to neurofibromas of the tongue, pharynx or larynx [13], decreased oromotor tone [17], or motoric dysfunction [5]. It has been reported that mechanical disruptions caused by tumors cannot always account for the speech impairment observed [15].

NF1 AND VOICE DISORDERS

A voice disorder is defined as an abnormal production or absence of vocal quality, pitch, resonance or duration [1]. Voice disorders occur in <1-23% of school aged children [18-22] and 3-7% of adults [23, 24] in the general population. An assessment of voice typically follows an evaluation by an otolaryngologist to rule out or confirm the presence of laryngeal pathology. Videostroboscopy of the larynx and an assessment of respiratory functioning and flexible nasolaryngoscopy may take place [see 25]. Observation of the type of breathing employed, and measurements of *fundamental frequency*, loudness, and quality of voice may also be completed [26].

Differences in voice have been noted between populations with and without NF1. The growth of a neurofibroma can impact movement of the mechanism [27]. While neurogenic tumors of the larynx are rare, neurofibromas occurring in the aryepiglottic folds can cause hoarseness and stridor [13]. For example, Yucel and colleagues [28] found dyspnoea and coughing in a six-year-old with a plexiform neurofibroma originating in the left aryepiglottic fold.

In describing the voice of individuals with NF1, a number of terms have been used in the literature, including: *breathy* [5, 29], hoarse [4, 5, 27, 29], strained [5], creaky [5], monotone, [29], fluttering [5], having a tremor [29], or unvarying loudness [29]. Among patients with NF1, voice concerns have been reported in children [4, 27, 30] and adults [31-33]. Solot and colleagues [27] found that approximately 22% (5/23) of their group of children with NF1 exhibited hoarseness compared to none of the ten unaffected sibling controls. One child presented with a large plexiform neurofibroma under the left collarbone and exhibited impaired vocal quality and *hypernasality*. In a later study, three of nineteen preschool children (16%) were characterized as having a voice disorder [4].

The prevalence of voice disorder in individuals with NF1 is significantly higher than in the general population [18, 23, 24]. Alivuotila et al. [5] evaluated the voice quality of 62 children and adults with NF1 and 24 unaffected control participants; abnormal phonation was observed in 26 (42%) of the individuals with NF1 compared to 3 unaffected individuals (13%). Twenty-four percent of the individuals with NF1 exhibited a strained, creaky, hoarse, or breathy voice quality and were noted to have a narrower range and difficulty regulating pitch compared to unaffected control participants. It was hypothesized that individuals with NF1 exhibited problems with motor coordination of the vocal folds, resulting in concerns with pitch regulation, phonation, and harmonic structure. It was also suggested that problems with a central mechanism may account for the differences between typical and affected populations.

Other investigations of adults with NF1 found the presence of voice disorders [5, 31] characterized by an atypically loud volume (40%) [31], or a harsh or creaky voice quality [31]. To examine quality of voice, Cosyns et al. [32] studied a group of adults with NF1 and unaffected participants. All participants were nonsmokers without a history of voice disorder, and were native Dutch speakers who did not have laryngeal or pharyngeal neurofibroma. Group differences showed significantly lower vital capacity values for subjects with NF1 versus the control group. Patients with NF1 exhibited narrower frequency and intensity ranges compared to participants of the same gender in the unaffected group. No significant between-group differences were observed for *jitter*, *shimmer*, mean fundamental frequency, and pitch variation. However, both males and females with NF1 exhibited significantly lower *dysphonia* severity index scores versus unaffected individuals, suggesting that overall voice quality was poorer in patients with NF1. Results of this study corroborated the report of Alivuotila et al. [5] that there is a narrower pitch range in patients with NF1 compared to control participants.

To characterize the impact of a voice disorder on an individual's functioning, Cosyns et al. [33] analyzed the results of the Flemish Dutch version of the Voice Handicap Index (VHI) [34] completed by 30 Flemish adults with NF1 and 30 healthy subjects. The VHI asks questions designed to assess the impact of a voice disorder on a person's functional, emotional and physical functioning (e.g., "I tend to avoid groups of people because of my voice" [35]). VHI total scores were significantly higher for the individuals with NF1 compared to the control group. The authors hypothesized that rather than the elevated VHI scores for individuals with NF1 being associated with the voice disorder specifically, scores may have been related to patients living with a progressive disease throughout their lifespan.

NF1 AND RESONANCE/VELOPHARYNGEAL FUNCTIONING

Among children with communication disorders, 3-4% manifest resonance disorders [20]. A resonance disorder can be related to changes in oral or nasal cavity size or configuration, muscle contraction/relaxation, positioning of the oral structures [26], or an imbalance of oral to nasal airflow. Resonance is evaluated by perceptual or instrumental methods during examinees' production of sustained vowels, oral or nasal sounds, or sentences heavily loaded with *oral* or *nasal phonemes* (to detect the presence of hypernasality or hyponasality). Indirect assessment of velopharyngeal functioning involves the use of *nasometry* and the Simplified Nasometric Assessment Procedure (SNAP) [36]. Participants repeat sentences loaded with oral or nasal phonemes. Results yield a ratio of oral to nasal airflow (nasalance score) for each stimulus set. Nasalance scores are compared with test norms from a sample population [36]. The Zoo Passage (a passage composed of sentences containing no nasal phonemes) allows the examiner to assess if a patient can establish and maintain velopharyngeal closure during connected speech [36]. Other assessment techniques, such as the use of nasendoscopy or videofluoroscopy, allow observation of velopharyngeal functioning more directly.

Studies that have examined resonance and/or velopharyngeal functioning among individuals with NF1 indicate the presence of hypernasality and/or *velopharyngeal insufficiency* [VPI; 4, 5, 17, 31, 37-40]. Pollack and Shprintzen [39] observed mild hypernasality in four of seven children (4-13 years of age) with NF1 using multi-view videofluoroscopy and nasoendoscopy. Three exhibited mild-moderate, moderate, or severe hypernasality. VPI did not occur with structural problems of the palate, pharynx or cervical spine, suggesting that a factor other than mechanical deviations likely contributed to the hypernasality observed.

Resonance problems have been reported in 1-52% of individuals with NF1, depending on the study. Reasons for discrepant results are unknown as studies employed large age ranges of participants and did not always describe the methods used to determine the presence of a resonance disorder. Solot and colleagues [27] examined pilot data of 23 children with NF1. Five of them (ca. 22%) exhibited hypernasality, in contrast to none of the unaffected sibling controls (n=10). A large study showed that approximately 1% of 200 children with NF1 (mean age 17.4 years) exhibited velopharyngeal incompetence [38]. Ferner, Hughes and Weinman [41] examined 103 individuals with NF1 and 105 control participants between the ages of 6 and 75 years. Nineteen (19%) of the individuals with NF1 exhibited hypernasal speech compared to none of the control participants. In another report, twelve patients with NF1 (without cleft palate) were identified as having hypernasal speech [17]. Adenoidectomy (n=3), the presence of brainstem pathology (n=2), or the presence of a Chiari I malformation (n=1) were stated as possible causes of velopharyngeal insufficiency. Alivuotila and colleagues [5] found abnormal nasality in 29% (18/62) of individuals with NF1 who were 7-66 years of age. The value obtained by Alivuotila et al. [5] was lower than the values obtained by Zhang and colleagues' [52.3% (11/21)] [42] investigation involving children and adults and the study of Thompson et al. [4] showing 42% of preschool children with NF1 having abnormal resonance. Finally, spontaneous speech and oral reading samples from 33 adults with NF1 [31] showed 37% of the group (n=11) was rated as hypernasal. While some studies were limited by small sample sizes, the literature indicates that velopharyngeal insufficiency occurs among individuals with NF1.

Cosyns and colleagues [37] examined a group of 24 Flemish adults with NF1 and 16 unaffected control participants using nasometric procedures with a variety of stimulus sets. Data from the individuals with NF1 were compared with normative values and control participant data to show male participants with NF1 had significantly higher nasalance scores for oronasal and oral texts than control male participants. When individual scores were compared with normative data, 9 of 12 males and 4 of 12 females with NF1 had scores outside of normal limits (at least one standard deviation above or below the mean) on some speech tasks. Results of these studies indicate that while there is variability across investigations, most individuals with NF1 do not exhibit resonance disorders.

NF1 AND FLUENCY

With a prevalence of approximately 1% in preschool children [43, 44], and 0-1% in school-aged children [18, 43, 44], a fluency disorder is characterized by an interruption in the flow of speaking. Specifically, it is noted by an atypical rate and/or rhythm, with repetitions or *prolongations* of sounds, syllables, words and/or phrases [1] and is often accompanied by secondary behaviors such as excessive tension or struggle [45].

Fluency is typically assessed through a case history interview to obtain information concerning the history, development, and current patterns of *dysfluency*. As part of the assessment, reading and conversational speech samples are obtained in different speaking contexts (e.g., at home, work, or school) to determine speech rate, pattern and frequency of the fluency concern [45].

In a literature review by Van Borsel and Tetnowski [46], fluency was reported as a concern for many individuals with NF1 [46]. Problems with rate [27, 31] and rhythm [13, 27] have been reported as well. Cosyns et al. [30] administered a questionnaire and found that adults with NF1 experienced fluency problems either at the time of the evaluation (5.5%) or historically (3.6%). From the group of individuals who reported that they stuttered, *blocking* was reported as the most frequently-occurring fluency concern [30]. As this initial investigation did not include a formal evaluation, the exact nature of dysfluency in individuals with NF1 is unclear. Following the study using surveys, Cosyns et al. [47] examined the speech of a 49-year old male with a diagnosis of NF1 and no positive family history of dysfluency. To assess dysfluencies, several speech samples were collected including spontaneous speech, monologue, repetition of words and sentences, automatic series, and reading. Dysfluencies were most frequently-occurring in monologue and automatic series samples, with prolongations noted as the most frequently-occurring dysfluency type. The authors concluded that the nature of the prolongations observed was somewhat consistent with dysfluencies in other populations, as they occurred more frequently in content and multisyllabic words.

Finally, a larger examination of speech samples of 21 Dutch speaking adults with NF1 aged 17 to 64 years [48] showed that all participants exhibited dysfluencies during spontaneous speech and monologue conditions. Two individuals (10%) exhibited dysfluencies typical of individuals who stutter. Frequently-occurring types of dysfluencies included interjections, revisions, prolongations, and incomplete phrases. It was concluded that in general, dysfluencies observed in individuals with NF1 are not identical to stuttering.

SPEECH PRODUCTION PROBLEMS ASSOCIATED WITH NF1

A *speech sound* disorder occurs when a child exhibits impaired production of phonemes (i.e., consonants and vowels) past the age that would be expected [1]. One type of speech sound disorder is an articulation disorder, characterized by the substitution, omission, distortion, or addition of speech sounds, contributing to reduced speech intelligibility. Speech sound disorders can have phonological pattern, in which sounds in the same class are affected systematically due to an underlying rule [49]. For example, the phonological pattern of “fronting” occurs when children produce front sounds for back sounds (i.e., “cup” is produced “tup” and “go” is produced “dough”). Approximately 16% of three year olds [50] and 4% of 6-year olds [51] exhibit a speech sound disorder. In the general population, prevalence rates decrease with age [18, 51, 52], affecting approximately 1% of school-aged children and young adults [18, 53].

An assessment of speech typically takes place once children begin to combine words, at approximately 18-24 months of age. A speech evaluation is part of an overall assessment of communication functioning and includes an analysis of children’s sound inventory, and the syllable and word shapes employed [49]. Beginning at three years of age, children’s production of speech sounds in isolation, single words, and connected speech is assessed using standardized tests and informal measures. Testing is required to determine if performance is consistent with age expectations (e.g., standard scores, correctly-produced sounds, and the number, type and consistency of errors).

Estimates of the prevalence of speech sound disorders in the population of individuals with NF1 are difficult to obtain. Presently, there are only a few studies documenting the prevalence of speech sound disorders in this population. Studies are limited by small sample sizes, and/or a lack of information regarding the age of participants or the measures used to diagnose the speech sound disorder. These limitations suggest that prevalence estimates should be interpreted with caution.

Data collected through surveys [30, 54, 55] indicate that speech sound disorders occur in the population of individuals with NF1. Johnson and colleagues [54] found that compared to 5% of unaffected siblings, 60% of children with NF1 exhibited speech concerns. Cosyns and colleagues [30] found that 65% (39/60) of participants 4-61 years of age reported at least one speech problem (of any type), with 82% reporting that they were “less intelligible” to others from results of a questionnaire. Most participants experienced problems with the pronunciation of speech sounds. Stine and Adams [55] found that 55/217 (25.4%) children with NF1 (in grades 1-12) were reported to be enrolled in speech therapy, indicating the need for speech services in this population. These initial studies employed the results of self-report measures without formal testing, thus making the exact nature of speech sound disorders in this population unclear. Nevertheless, these studies underscore the idea that speech is a concern for many individuals with NF1.

Other studies have estimated the frequency of speech sound disorders for individuals with NF1 to be anywhere from 12 to 52%, depending on the population studied or the measure used to diagnose the speech problem. White and colleagues [13] found impaired speech in 19% of a group of 257 patients with NF1. Wong [56] reported that 12% of a group of 50 Chinese children with NF1 (ranging in age from 21 months to 18 years) exhibited speech problems. Noble and colleagues [57] found higher rates, with 24% (9/37) of a group of pre-school children with NF1

exhibiting a speech delay as their only developmental disability. As none of the studies described the assessments used to determine speech delay status and the last two included children who were under two years of age, these initial prevalence statistics may be an inaccurate estimate.

Riccardi [29] reported that 30-40% of individuals with NF1 exhibited speech problems [see Glass, L and Riccardi VM, unpublished data cited in 29]. Hypernasality, reduced rate, imprecise consonants, breathiness, hoarseness, monotone, tremor, and unvarying loudness were observed in their examination of 17 patients [29]. Rather than being attributed to factors related to the larynx, the etiology of the speech problems was thought to have a neurological origin. Solot and colleagues [27] reported that 48% (11/23) of children with NF1 exhibited articulation disorders, compared to only 20% (2/10) of unaffected sibling controls, suggesting that speech sound disorders are more frequent in patients with NF1 than in typically-developing children. However, as the control group exhibited higher rates of impairment than would be expected from the general population, further study is warranted. In another study, approximately 32% (6/19) of four- and five-year-old children with NF1 exhibited speech sound disorders as defined by a standard score of more than one standard deviation below the mean on the Goldman-Fristoe Test of Articulation-2 [4].

Speech sound disorders are present in adults with NF1 as well. For example, Riccardi and Eichner [58] reported that 57 (26%) of their group exhibited “definite” speech impairment, and an additional 57 (26%) were suspected to have a speech concern in an NF1 population followed in an academic clinic setting. Although no data were provided regarding the age of participants at the time of the evaluation, a significant number of individuals with NF1 (including adults) exhibited articulation concerns. Lorch and colleagues [31] found that 17% (5/30) of a group of adults with NF1 exhibited mild articulation difficulties and 40% (12/30) exhibited a fast rate of speech. Some of these studies were limited by small sample sizes, and not all of the studies described the testing methods employed. Nevertheless, results suggest that speech sound disorders occur in both children and adults with NF1 and are more frequently-occurring than would be observed in the general population.

Many early studies examining speech production problems in patients with NF1 reported the prevalence of speech sound disorders. However, more recently, researchers have been interested in describing the nature of speech impairment among individuals with NF1. Alivuotila et al. [5] found significantly more deviations in articulation by individuals affected with NF1 ($n = 62$) compared to a control group ($n = 24$). With the exception of abnormal /l/ and /d/ production, results were consistent with what patients indicated on self-report measures [30]. They demonstrate that patients experience speech problems well into adulthood, and many sound classes including stops, fricatives, nasals, liquids, and vowels are affected. To obtain more detail about the nature of speech sound errors, Cosyns et al. [6] completed a phonetic and a *phonological analysis* on the speech of 43 individuals with NF1 representing a wide age-range (7-53 years) including 14 children and 29 adults. All participants were monolingual speakers of Flemish and fulfilled NF1 diagnostic criteria. Results showed that 2 of the 14 children exhibited an incomplete *phonetic inventory*. Percent Consonants Correct (PCC) values were also reported, with children and adults exhibiting PCC values of 78-99% and 87-100%, respectively. The majority of children with NF1 exhibited a mild phonological disorder, with 4 exhibiting mild-moderate phonological disorders. Both adults and children with NF1 exhibited substitution, distortion, omission, and addition errors. Phonological patterns employed by children with NF1 included *final consonant deletion*, *epenthesis*, cluster simplification, and

devoicing. These results are consistent with cluster reduction, substitution and omission errors also observed in preschool children with NF1[4]. Cosyns et al. [6] noted that while adults were more proficient in their speech production abilities than children with NF1, it was apparent that adults with NF1 do not “grow out” of their speech sound disorders, and they continue to exhibit residual articulation errors well into adulthood. Taken together, results of the literature examining speech production skills of children with adults with NF1 suggest that a significant minority exhibit life-long difficulties with speech sound production.

EXPRESSIVE AND RECEPTIVE LANGUAGE AND INTELLIGENCE IN NF1

Language disorders occur when there are problems in comprehending (*receptive language*) and/or using (*expressive language*) spoken, written or other symbol systems [1]. Language disorders are characterized by problems with the form (i.e., morphology or syntax), content (i.e., semantics) or function (i.e., pragmatics) of verbal or written communication [1]. Approximately 13-19% [59-61] of toddlers/preschool children are considered “late-talkers”. At three years of age, late-talkers exhibit lower scores on language measures when compared to typically-developing peers [62]. The delays of most late-talkers are temporary and resolve prior to children entering school. However, for some late-talking children, the delays persist. Approximately 7.4% of kindergarten children [63] and 5% [64, 65] of school-aged children exhibit a language disorder.

A language assessment for young children (under three years of age) consists of an evaluation of overall communication functioning. Children’s symbolic play skills, comprehension, and production of language are evaluated through formal or informal methods (i.e., observation of children’s language use during play). A language sample may be analyzed to determine vocabulary size (*number of different words*) and *mean length of utterance* (MLU) in morphemes (see [66, 67] for rules for computing). MLU is used as a way of indexing the expansion of clauses [68] and average sentence length [69], and emerging language/grammatical abilities [66, 67]. Formal assessment methods involve the use of standardized tests to assess children’s general communication development (e.g., Preschool Language Scales-5th edition [70], Rossetti Infant and Toddler Language Scale [71]). As children age, standardized language tests are employed. A language disorder is diagnosed when a child obtains a score of at least one standard deviation below the mean on standardized testing or demonstrates limitations in either language comprehension or language production during spontaneous conversation. Analyzed language samples of 175-utterances in length [72] provide information about the child’s expressive morphosyntactic abilities, mean length of utterance in morphemes, and number of different words.

While articulation and voice disorders are common manifestations and recognized, most clinicians still do not associate early language impairment with NF1. In the 1980s and 1990s, much of the research examining language outcomes of individuals with NF1 was developed from the medical field, specifically neurology. Studies that suggested the presence of language delays (or verbal ability/disability) in this population originated from a body of research that was primarily interested in describing the nature of learning disabilities [e.g., 73] or the cognitive phenotype [e.g., 74]. In a few of the studies, the diagnosis of a language problem was

made following participants' poor performance on language-related subtests of neuropsychological measures, where describing language performance was not the primary focus of the study. Speech-language pathologists were frequently not included in the studies. A few studies included measures that are typically used for the identification of language problems in populations with aphasia [e.g., 31, 76]. Although the primary focus of this section concerns language outcomes for individuals with NF1, this literature was initially focused on describing the cognitive skills of individuals with NF1 (see Chapter 7). Thus, prior to the discussion of language outcomes, a brief description of the literature examining results of intelligence testing and learning disabilities is provided.

In general, individuals with NF1 exhibit lower IQ scores than those who are unaffected [41, 74, 76-80]. Scores tend to fall in the below-average to low-average range. When group means of verbal IQ and performance IQ subtests are evaluated, results are variable across studies [see 81 for a review]. Some studies have reported higher scores in verbal IQ than performance IQ [73, 82, 83], or higher scores in performance IQ than verbal IQ [79, 84]. The majority of papers, however, have shown no significant difference between verbal and performance indices [41, 55, 78, 85-87].

Learning difficulties/disabilities have been reported with a frequency of 30-65% in children with NF1 [55, 58, 73, 76, 88-91]. In general, a definition of "learning disability" refers to a discrepancy between intelligence quotients (IQ) and standard academic achievement scores, where IQ scores are higher [e.g., 55]. There is little consensus regarding the nature of learning disabilities in this population. Reasons for the discrepant results across studies include variable comparison groups, differences in the methodologies used, failure of studies to control for IQ [92], or differences in what authors consider a "discrepancy" in a learning disability [93].

Early studies proposed that nonverbal learning disabilities were prevalent in NF1 [73] and more frequent than language-based/verbal learning disability [55]. However, more recently investigations have indicated that language-based learning disabilities occur as well [see 81]. Language problems have been identified in the NF1 population with specific deficits in the areas of reading [31, 41, 76, 82, 94-96], spelling [31, 89, 94, 95], decoding and phonological awareness skills [96], vocabulary [96], naming [74, 79], and written language [31, 77, 82].

Eliason [73] was one of the earliest authors to report language deficits in the population of children with NF1, reporting that 30% (8/23) of children with NF1 referred to a learning disorders clinic exhibited language problems. Eldridge et al. [79] found almost identical rates of spoken language disorder between children and adults with and without NF1, with 2/13 children and adults with NF1 and 3/13 unaffected siblings exhibiting a possible spoken language disorder. Eldridge et al. [79] indicated that perhaps developmental disabilities occurred more frequently than would be expected in the unaffected group of participants, which may account for the elevated rates of language impairment in those without NF1. Others have reported lower expressive vocabulary abilities in children with NF1 [aged 6;6 to 13;7 (years; months)] compared to unaffected siblings [aged 7;11 to 16;4] [74]. Mazzocco et al. [76] assessed verbal and nonverbal performance of 19 pairs of children (with and without NF1, 6-16 years of age) using several neuropsychological measures. Children with NF1 obtained significantly lower scores than the unaffected group in the areas of word definitions, expressive vocabulary, total written language and contextual vocabulary, receptive syntactic language, reasoning, phonological memory, and phoneme segmentation. No significant between-group differences emerged from several neuropsychological measures. Dilts et al. [77] examined the receptive and expressive language of nineteen children with NF1 (aged 6-17 years) and a group

of unaffected siblings as part of a study designed to establish a behavioral phenotype of NF1. Approximately 58% of the group of children with NF1 failed the CELF screening test. Of the children who failed screening, it was noted that 26% of the sample exhibited delays in both expressive and receptive language. Approximately 32% of the group exhibited delays in expressive language alone. Consistent with the findings of Mazzocco and colleagues [76], Dilts et al. [77] concluded that children with NF1 had significant verbal deficits. In the adult population, 13% and 40% of individuals with NF1 exhibit delays in reading and writing, respectively [31]. Results of the studies suggest that even though IQ may be close to equivalent in populations with and without NF1, language disorders are a significant, although not always recognized, concern in childhood.

Many of the previous investigations examined children who were 6-16 years of age. In an effort to better identify children with NF1 who are at-risk for later problems, researchers became interested in examining the language outcomes of a younger group of children with NF1 from 7 months to 8 years of age [97, 98]. Results of these studies varied depending on the measures used to assess the presence or absence of a language delay. Nevertheless, studies reported concerns with language early in life. At 21-30 months of age, longest sentences (as measured by parent report on the MacArthur Communicative Development Inventory) of children with NF1 are not significantly different from age-matched controls, although children with NF1 exhibit poorer global language ability [98]. At 3-5 years of age, language delays occur in 53% of children, with receptive language delays in approximately 37%, and expressive language delay in 37% [4]. Leguis et al. [83] examined children in three different age ranges, including children from 17 months-4 years of age, 4-6 years of age, and 6-16 years of age. In the youngest group, they found that 6 out of 7 children (86%) exhibited language delays, specifically in the area of morphology. At 4-6 years of age, most children exhibited higher verbal intelligence scores than performance intelligence scores. In the oldest age group, while many obtained below-average scores, more than half of the group obtained total intelligence quotients in the average range. The authors reported that because many children exhibited language delays early in development, it would be useful to follow a group of children to determine which ones develop learning disabilities later in life.

Soucy et al. [97] examined a much smaller age range of children with NF1 (aged 7-months to 8-years) using the Parents' Evaluation of Developmental Status: Developmental Milestones (PEDS). The PEDS is a screening tool used to determine the presence of a range of developmental disorders. Children with NF1 were noted to present with delays in 24% and 22%, in receptive and expressive language, respectively. These results indicate that language delays occur very early in development for a significant minority of children with NF1, supporting the referral of children with NF1 to speech-language pathologists prior to their second year [98]. As the sensitivity of the PEDS in predicting language problems in children two years later is poor [99], it is possible that many children with language delays were not correctly identified following test screening, and the values obtained from this study could be an underrepresentation of the language delays in children with NF1. Nevertheless, results of these studies suggest that language impairment frequently occurs in the NF1 population.

While the exact etiology of language impairment in NF1 is unknown, it has been postulated that differences in brain morphology may account for the delays observed [95, 100]. One study found an association between planum temporale surface area asymmetry and IQ-based reading comprehension for individuals with NF1 [100]. Another study found that an extra right hemisphere inferior frontal gyrus in patients with NF1 was associated with better performance

on language-related measures (e.g., phonological fluency, verbal knowledge, reading, spelling, and verbal memory) [95]. Finally, the presence of minor brain malformations (e.g., hypothalamic hamartomas, malformations of cortical development) among individuals with NF1 has been associated with lower scores in intelligence, reading, mathematics and spelling [101]. Researchers sought to determine if reading difficulties in children and adolescents with NF1 were related to the pattern and extent of activation during phonological processing [102]. Individuals with NF1 exhibited more activation in the inferior frontal cortex during auditory phonological processing compared to control participants, suggesting that activity in the inferior frontal cortex may play a role in language deficits in this population. Large corpus callosum index scores have also been associated with poorer intelligence, reading, verbal memory, and executive functioning [103]. Future research may provide additional information on the relationship between anatomy, function and behavior. For example, the use of resting state functional connectivity [see 104] has been proposed as viable outcome measure for evaluating the use of pharmacological agents in the treatment of developmental processes in NF1 [e.g., 105].

Results of these accumulated studies over the last 25 years suggest that while many individuals with NF1 may exhibit language concerns, many individuals are also unaffected. As a result of this research, careful screening and monitoring of the language skills of this population should take place, with follow-up intervention as needed. Potential assessment procedures will be discussed in detail below.

GAPS AND OPPORTUNITIES

The studies presented in this chapter have a number of limitations. The majority of the studies employed large age ranges of participants, limiting the generalization of study results. As younger children with language delays may be slow to develop, they may have the potential to acquire language normally without intervention (i.e., “late-talkers”). In contrast, older children with language delays may have language-based learning disabilities, and display residual language concerns following intervention. Combining children with these two different etiologies into one group prevents researchers from obtaining a clear picture of the language disorder in this population. As the profile of a language disorder changes over time, it is important that studies group children of a similar age range and use measures that are appropriate for examining language performance of that age range.

Second, it was noted that the majority of studies employed language subtests of IQ measures rather than language-specific measures. In speech-language pathology, language subtests of neuropsychological measures are generally not employed to diagnose the presence or absence of a language disorder. Current research on specific language impairment and other developmental language disorders do not use verbal IQ in the development of their behavioral phenotypes. Neuropsychological measures may have an insufficient number of items to assess a given area. Children may exhibit a delay in an area of language that is not captured through neuropsychological testing, but can be identified through language sample analysis or formal language testing. Therefore, while it is important that speech-language pathologists use the information from neuropsychological testing to guide clinical practice, it should be used as a foundation for other testing in the field of speech-language pathology.

Finally, rather than including age-matched control children, unaffected siblings were frequently included in the studies. As communication skills change over time, age-matched, unaffected controls would make for a more informative comparison group. As previous authors have suggested that outcomes may be influenced by the patient's experience of living with NF1, another appropriate comparison group may be individuals living with a different progressive disease.

To date, the etiology of communication disorders in NF1 remains unclear. Some studies have suggested that functional or physiological characteristics limit speech production [13, 17]. Others have suggested that differences in brain morphology can account for the language delays [95, 100]. Two studies that examined the nature of speech sound errors in individuals with NF1 noted the presence of phonological patterns [4, 6]. As phonological disorders are cognitive-linguistic in nature, mechanical factors alone cannot explain the speech sound disorder in this group. From a review of the literature, at this stage there does not appear to be a coherent linguistic phenotype of speech in NF1 [106]. Important moderators of NF1 manifestations that have yet to be determined may place children at risk for language delays. Additional research is needed to determine the etiology of language impairment in this group, and it is imperative that language studies incorporate specific measures rather than subtests of other psychological assessments.

ASSESSMENT

In light of the previous literature review, it is apparent that children and adults with NF1 have an elevated risk for difficulties in a number of areas of communication functioning including voice, resonance, fluency, articulation, and language. Currently, there is a paucity of research describing appropriate assessment protocols specifically addressing speech and language concerns for individuals with NF1. This section provides guidelines regarding speech-language assessment for this population.

From 21-months of age, children with NF1 exhibit problems with language development [98]. By three to five years of age, children with NF1 are more likely to exhibit speech sound disorders and language delays than the general population [4]. As resources may be limited for families with children with disabilities and attendance at speech-language evaluations may place an additional burden on families who already have many medical appointments, the use of screening measures is highly recommended. Options for screening include the MacArthur-Bates Communicative Development Inventories [107] (for younger children), or the Clinical Evaluation of Language Fundamentals-Screening Test [108] (for older children). Follow-up testing is warranted following a failed screening. A referral to a speech-language pathologist is also warranted at 24 months of age if the child produces fewer than 50 words or no two-word combinations, *and* if parents are concerned about their child's speech, language or hearing or if the child has had six or more episodes of otitis media [109].

At two years of age, children's speech and language should be examined to assess vocabulary (Number of Different Words), MLU and production of early-developing sounds. Measures should be compared with normative data to determine if the child is within normal limits or significantly different from typically developing children. While the child's speech

and language status may be quite low given the child's chronological age, voice and fluency should also be monitored.

As children enter preschool, an assessment of general communication skills should take place. Speech sound production should be assessed. Standardized tests of articulation or phonology (such as the Goldman-Fristoe Test of Articulation-2 [110] or the Bankson-Bernthal Test of Phonology [111]) can be employed. To assess expressive and receptive language, omnibus language testing should be administered (using standardized measures), and a language sample should be collected. As children are at risk for hearing loss (as are any unaffected children), they should undergo a hearing screening to ensure that their hearing is within normal limits on the day of testing. An oral mechanism exam should also take place to determine if oral-motor structure and function is within normal limits for the purpose of speech and to assess any differences in oral anatomy. Fluency and voice should be informally assessed during conversation.

For school-aged children with NF1, fluency, voice/resonance, speech, and expressive and receptive language should be assessed. As in the preschool years, standardized tests can be administered to assess speech sound development, expressive and receptive language. A language sample analysis should be completed, as some children with NF1 may be able to perform adequately on omnibus language tests but still present with deficiencies in their expressive language that are uncovered in their spontaneous conversations [112]. Results of subtests of neuropsychological testing can provide information regarding how children use language for problem-solving, as well as inform clinicians as to what additional (and more specific) testing should be completed to obtain a clear picture of a child's language skills.

Language samples can identify strengths and weaknesses in the areas of sentence length and complexity (including MLU), vocabulary/number of different words, morphology, and syntax. In addition to *morphosyntax*, other areas that are not assessed by neuropsychological testing include narrative comprehension and narrative production. These areas can be evaluated through measures such as the Test of Narrative Language [113]. Additionally, as children with NF1 age, other areas of language functioning, such as sentence and paragraph comprehension, nonliteral language use, inferencing, and the ability to abstract meaning from context should be assessed.

CONCLUSION

The purpose of this chapter was to present a literature review of the communication disorders observed in the population of children and adults with NF1. It should be recognized that the speech and language characteristics within this population are extremely variable. Some individuals will not present with any significant limitations and others will exhibit difficulties in one or multiple areas. From a review of the literature, individuals with NF1 often exhibit concerns with voice, resonance, fluency, articulation, and expressive and receptive language more frequently than the general population. A summary of the findings is presented in Appendix 1. The reason for the variability is unclear; the severity of NF1 symptoms does not necessarily predict severity of the communication difficulties as a straightforward association. Communication disorders are a component of the cognitive phenotype for NF1.

As children with NF1 are at-risk for communication disorders, age-appropriate screening measures should be implemented in routine management. Minimally, communication development should be assessed when children are two and then again at least once in the preschool years to prevent negative academic and social outcomes. Children with central nervous system anomalies (i.e., optic pathway tumors, hydrocephalus, posterior fossa tumors, or intracranial vascular abnormalities) require closer surveillance for communication disorders. Upon the identification of a communication disorder, follow-up intervention and re-assessments should take place, as needed.

REFERENCES

- [1] American Speech Language Hearing Association. *Definitions of communication disorders and variations* 1993 [Retrieved on 06/04/12]; Available from: www.asha.org/policy.
- [2] S. McLeod and D. McKinnon, *International Journal of Language and Communication Disorders*. 42, S1 (2007).
- [3] US Department of Education, *27th annual report to congress on the implementation of the individuals with disabilities education act 2005*, 2005, ED Pubs, Education Publications Center: Jessup: MD. p. 1-487.
- [4] H. Thompson, D. Viskochil, D. Stevenson, and K. Chapman, *American Journal of Medical Genetics Part A*. 152A, (2010).
- [5] L. Alivuotila, J. Hakokari, V. Visnapuu, A. M. Korpjaakko-Huuhka, O. Aaltonen, R. P. Happonen, S. Peltonen, and J. Peltonen, *American Journal of Medical Genetics. Part A*. 152A, 1 (2010).
- [6] M. Cosyns, G. Mortier, S. Janssens, F. Bogaert, S. D'Hondt, and J. Van Borsel, *Journal of Communication Disorders*. 45, 2 (2012).
- [7] M. Lammert, J. Friedman, L. Kluwe, and V. Mautner, *Arch Dermatol*. 141, (2005).
- [8] J. M. Friedman, *American Journal Of Medical Genetics*. 89, 1 (1999).
- [9] P. O'Connell, R. Cawthon, D. Viskochil, R. White, J. Carey, and AM. Buchberg. *Annals Of The New York Academy Of Sciences*. 615, (1991).
- [10] J. M. Friedman, *Journal of Child Neurology*. 17, 8 (2002).
- [11] M. Cnossen, A. de Goede-Bolder, K. van den Broek, C. Waasdorp, A. Oranje, H. Stroink, H. Simonsz, A. van den Ouweland, D. Halley, and M. Niermeijer, *Arch. Dis. Child*. 78, (1998).
- [12] E. Adekeye, A. Abiose, and R. Ord, *Journal of Maxillofacial Surgery*. 12, (1984).
- [13] A. White, R. Smith, C. Bigler, W. Brooke, and P. Schauer, *Laryngoscope*. 96, (1986).
- [14] S. Shapiro, K. Abramovitch, M. Van Dis, L. Skoczylas, R. Langlais, R. Jorgensen, R. Young, and V. Riccardi, *Oral Surgery, Oral Medicine, Oral Pathology*. 58, 1 (1984).
- [15] S. Sobel, T. Tewfik, and J. Ortenberg, *The Journal of Otolaryngology*. 26, 1 (1997).
- [16] J. Czigner and G. Fekete-Szabó, *The Journal of Laryngology and Otology*. 108, (1994).
- [17] H. Saal, E. Schorry, and R. Hopkin, *Proc Greenwood Genetic Cent*. 23, 94-95 (2004).
- [18] D. McKinnon, S. McLeod, and S. Reilly, *Language, Speech, and Hearing Services in Schools*. 38, (2007).
- [19] E. Silverman and C. Zimmer, *Journal of Speech and Hearing Disorders*. 40, 2 (1975).

- [20] G. A. Warr-Leeper, R. S. McShea, and H. A. Leeper, *Language, Speech, and Hearing Services in Schools*. X, (1979).
- [21] S. Dawson and M. Pannbacker, *Texas Journal of Audiology and Speech-Language Pathology*. XXXI, (2008).
- [22] P. Carding, S. Roulstone, K. Northstone, and A. S. Team, *Journal of Voice*. 20, 4 (2005).
- [23] N. Roy, R. Merrill, S. Gray, and E. Smith, *The Laryngoscope*. 115, (2005).
- [24] D. Boone, S. McFarlane, and S. Von Berg, *The voice and voice therapy*; seventh edition, Pearson, Boston: MA (2005).
- [25] R. Faust, *Clinical Pediatrics*. 42, (2003).
- [26] D. Boone and S. McFarlane, *The voice and voice therapy*, in *The Voice and Voice Therapy*, 6th edition, Allyn and Bacon: Needham Heights: MA. p. 1-17. (2000).
- [27] C. Solot, E. Zackai, A. Obringer, D. Konkle, S. Handler, and A. Meadows, Communication disorders in children with neurofibromatosis type 1, in *Neurofibromatosis: A handbook for patients, families, and health-care professionals*, A. Rubenstein and B. Korf, Thieme Medical Publishers: New York: NY. p. 59-70. (1990).
- [28] E. Yucel, Y. Guldiken, M. Ozdemir, and A. Ozturk, *The Journal of Otolaryngology*. 116, (2002).
- [29] V. M. Riccardi, *The New England Journal of Medicine*. 305, 27 (1981).
- [30] M. Cosyns, L. Vandeweghe, G. Mortier, S. Janssens, and J. Van Borsel, *International Journal of Language and Communication Disorders*. 45, 5 (2010).
- [31] M. Lorch, R. Ferner, J. Golding, and R. Whurr, *Journal of Neurolinguistics*. 12, 3-4 (1999).
- [32] M. Cosyns, G. Mortier, P. Corthals, S. Janssens, and J. Van Borsel, *Journal of Voice*. 25, 6 (2011).
- [33] M. Cosyns, G. Mortier, S. Janssens, and J. Van Borsel, *Journal of Voice*. 26, 2 (2012).
- [34] M. De Bodt, B. Jacobsen, S. Musschoot, S. Zaman, L. Heylen, F. Mertens, P. Van de Heyning, and F. Wuyts, *Logopedie en Foniatrie*. 13, (2000).
- [35] B. Jacobson, A. Johnson, C. Grywalski, A. Silbergleit, G. Jacobson, M. Benninger, and C. Newman, *American Journal of Speech-Language Pathology*. 6, 3 (1997).
- [36] A. W. Kummer, D. A. Billmire, R. Campbell, J. Corcoran, M. Dock, R. J. Hopkin, C. K. Miller, H. Saal, J. R. Schultz, J. P. Willging, and D. Zajac, *Nasometry*, in *Cleft Palate and Craniofacial Anomalies: The Effects of Speech and Resonance*., Singular: San Diego, California. p. 311-330. (2001).
- [37] M. Cosyns, G. Mortier, S. Janssens, K. Claes, and J. Van Borsel, *Am. J. Med. Genet. Part A*. 155, (2011).
- [38] K. North, *Journal of Child Neurology*. 8, 4 (1993).
- [39] M. Pollack and R. Shprintzen, *International Journal of Pediatric Otorhinolaryngology*. 3, (1981).
- [40] Y. Berman, D. Fitzsimons, D. Gill, K. Prelog, M. Wilson, and K. North, Velopharyngeal insufficiency in neurofibromatosis type 1: A clinical sign of prognostic significance?, in *American Cleft Palate-Craniofacial Association 66th Annual Meeting 2009*: Scottsdale, AZ.
- [41] R. Ferner, R. Hughes, and J. Weinman, *Journal of the Neurological Sciences*. 138, (1996).
- [42] I. Zhang, M. Husein, A. Dworschak-Stokan, J. Jung, D. Matic, V. Siu, C. Prasad, and P. Doyle, *Journal of Otolaryngology-Head and Neck Surgery*. 41, 1 (2012).

- [43] A. Craig, K. Hancock, Y. Tran, M. Craig, and K. Peters, *Journal of Speech, Language, and Hearing Research*. 45, (2002).
- [44] G. Andrews, *Epidemiology of stuttering*, in *Nature and Treatment of Stuttering: New Directions*, R. Curlee and W. Perkins, College-Hill Press: San Diego: CA. (1984).
- [45] B. Guitar, *Stuttering: An Integrated Approach to Its Nature and Treatment*. 2nd ed, Williams and Wilkins, Baltimore: MD (1998).
- [46] J. Van Borsel and J. Tetnowski, *Journal Of Fluency Disorders*. 32, (2007).
- [47] M. Cosyns, G. Mortier, P. Corthals, S. Janssens, and J. Van Borsel, *Journal Of Fluency Disorders*. 35, 3 (2010).
- [48] M. Cosyns, G. Mortier, S. Janssens, N. Saharan, E. Stevens, and J. Van Borsel, *Journal Of Fluency Disorders*. 35, 1 (2010).
- [49] J. Bernthal, N. Bankson, and P. Flipsen Jr, *Articulation and Phonological Disorders: Speech Sound Disorders in Children*, 6th edition, Pearson, Boston: MA (2009).
- [50] T. Campbell, C. Dollaghan, H. Rockette, J. Paradise, H. Feldman, L. Shriberg, D. Sabo, and M. Kurs-Lasky, *Child Development*. 74, 2 (2003).
- [51] L. Shriberg, J. Tomblin, and J. McSweeney, *Journal of Speech, Language, and Hearing Research*. 42, (1999).
- [52] J. Beitchman, R. Nair, M. Clegg, and P. Patel, *Journal of Speech and Hearing Disorders*. 51, (1986).
- [53] G. Culton, *Journal of Speech and Hearing Disorders*. 51, (1986).
- [54] N. S. Johnson, H. M. Saal, A. M. Lovell, and E. K. Schorry, *The Journal Of Pediatrics*. 134, 6 (1999).
- [55] S. Stine and W. Adams, *Clinical Orthopaedics and Related Research*. 245, (1989).
- [56] V. Wong, *Pediatr. Neurol*. 11, (1994).
- [57] F. Noble, A. Kornberg, J. Elder, and M. Delatycki, *Journal of Paediatrics and Child Health*. 43, (2007).
- [58] V. Riccardi and J. Eichner, eds. *Neurofibromatosis, Natural History, and Pathogenesis*. Johns Hopkins University Press: Baltimore: MD. (1986). p. 305.
- [59] S. Zubrick, C. Taylor, M. Rice, and D. Slegers, *Journal of Speech, Language, and Hearing Research*. 50, (2007).
- [60] S. Horwitz, J. Irwin, M. Briggs-Gowan, J. Bosson Heenan, J. Mendoza, and A. Carter, *Journal of the American Academy of Child and Adolescent Psychiatry*. 42, 8 (2003).
- [61] L. Rescorla, M. Hadicke-Wiley, and E. Escarce, *First Language*. 13, 5 (1993).
- [62] L. Rescorla, B. Coll, J. Roberts, and K. Dahlsgaard, *Journal of Speech and Hearing Research*. 40, 3 (1997).
- [63] J. Tomblin, N. Records, X. Zhang, E. Smith, and M. O'Brien, *Journal of Speech, Language, and Hearing Research*. 40, (1997).
- [64] M. Nippold and I. Schwartz, *Advances in Speech-Language Pathology*. 4, 1 (2002).
- [65] G. Rost and K. McGregor, *American Journal of Speech-Language Pathology*. 21, (2012).
- [66] R. Paul, Assessment and intervention for emerging language, in *Language Disorders from Infancy Through Adolescence*, second edition, Mosby: St. Louis: MO. (2001).
- [67] R. Brown, *A First Language: The Early Stages*, Harvard University Press, Cambridge: MA (1973).
- [68] M. Rice, S. Redmond, and L. Hoffman, *Journal of Speech, Language, and Hearing Research*. 49, (2006).

- [69] S. Eisenberg, T. McGovern Fersko, and C. Lundgren, *American Journal of Speech-Language Pathology*. 10, 4 (2001).
- [70] I. Zimmerman, V. Steiner, and R. Pond, *Preschool Language Scales* 5th edition, Pearson, Bloomington: MN (2011).
- [71] L. Rossetti, *The Rossetti Infant-Toddler Language Scale: A Measure of Communication and Interaction*, LinguiSystems, East Moline: IL (1990).
- [72] W. Gavin and L. Giles, *Journal of Speech and Hearing Research*. 39, (1996).
- [73] M. Eliason, *Developmental and Behavioral Pediatrics*. 7, 3 (1986).
- [74] K. Hofman, E. Harris, N. Bryan, and M. Denckla, *The Journal of Pediatrics*. 124, 4 (1994).
- [75] B. D. Moore, 3rd, J. Ater, M. Needle, J. M. Slopis, and D. Copeland, *Journal of Child Neurology*. 9, 4 (1994).
- [76] M. M. M. Mazzocco, J. E. Turner, M. B. Denckla, K. J. Hofman, D. C. Scanlon and F.R. Vellutino *Developmental Neuropsychology*. 11, 4 (1995).
- [77] C. Dilts, J. Carey, R. Hoffman, D. Creel, K. Ward, E. Clark, and C. Leonard, *Developmental and Behavioral Pediatrics*. 17, 4 (1996).
- [78] C. Varnhagen, S. Lewin, J. Das, P. Bowen, K. Ma, and M. Klimek, *Developmental and Behavioral Pediatrics*. 9, 5 (1988).
- [79] R. Eldridge, M. Denckla, E. Bien, S. Myers, M. Kaiser-Kupfer, A. Pikus, S. Schlesinger, D. Parry, J. Dambrosia, M. Zasloff, and J. Mulvihill, *Am. J. Dis. Child*. 143, (1989).
- [80] J. Sangster, E. A. Shores, S. Watt, and K. North, *Child Neuropsychology*. 17, 1 (2010).
- [81] A. Kayl and B. D. Moore, 3rd, *Mental Retardation and Developmental Disabilities Research Reviews*. 6, (2000).
- [82] M. Wadsby, H. Lindehammar, and O. Eeg-Olofsson, *Neurofibromatosis*. 2, (1989).
- [83] E. Leguis, M. Descheemaeker, A. Spaepen, P. Casaer, and J. Fryns, *Genet Couns*. 5, (1994).
- [84] B. D. Moore, 3rd, J. M. Slopis, D. Schomer, E. F. Jackson, and B. M. Levy, *Neurology*. 46, (1996).
- [85] K. North, P. Joy, D. Yuille, N. Cocks, E. Mobbs, P. Hutchins, K. McHugh, and M. de Silva, *Neurology*. 44, 5 (1994).
- [86] K. North, P. Joy, D. Yuille, N. Cocks, and P. Hutchins, *Developmental Medicine and Child Neurology*. 37, 5 (1995).
- [87] S. L. Hyman, A. Shores, and K. N. North, *Neurology*. 65, 7 (2005).
- [88] V. R. Brewer, B. D. Moore, and M. Hiscock, *Journal of Learning Disabilities*. 30, 5 (1997).
- [89] M. Descheemaeker, P. Ghesquière, H. Symons, J. Fryns, and E. Leguis, *Journal of Intellectual Disability Research*. 49, 1 (2005).
- [90] F. Coudé, C. Mignot, S. Lyonnet, and A. Munnich, *Behavior Genetics*. 36, 5 (2006).
- [91] K. McKeever, C. Shepherd, H. Crawford, and P. Morrison, *Ulster Med. J*. 77, 3 (2008).
- [92] T. M. Levine, A. Materek, J. Abel, M. O'Donnell, and L. E. Cutting, *Seminars In Pediatric Neurology*. 13, 1 (2006).
- [93] S. Ozonoff, *American Journal Of Medical Genetics*. 89, (1999).
- [94] S. L. Hyman, E. A. Shores, and K. N. North, *Developmental Medicine and Child Neurology*. 48, 12 (2006).
- [95] R. L. Billingsley, J. M. Slopis, P. R. Swank, E. F. Jackson, and B. D. Moore, 3rd, *Brain and Language*. 85, (2003).

- [96] L. E. Cutting, C. W. Koth, and M. B. Denckla, *Developmental Neuropsychology*. 17, 1 (2000).
- [97] E. Soucy, F. Gao, D. Gutmann, and C. Dunn, *Journal of Child Neurology*. 27, 5 (2012).
- [98] J. Lorenzo, B. Barton, M. T. Acosta, and K. North, *The Journal Of Pediatrics*. 158, 4 (2011).
- [99] M. Wake, B. Gerner, and S. Gallagher, *Ambul. Pediatr.* 5, 3 (2005).
- [100] R. Billingsley, G. Schrimsher, E. F. Jackson, J. M. Slopis, and B. D. Moore, 3rd, *Arch. Neurol.* 59, (2002).
- [101] M. T. Acosta, K. S. Walsh, P. G. Kardel, R. E. Kutteruf, R. R. Bhatt, T. C. Bouton, L.-G. Vezina, and R. J. Packer, *Pediatric Neurology*. 46, (2012).
- [102] R. L. Billingsley, E. F. Jackson, J. M. Slopis, P. R. Swank, S. Mahankali, and B. D. Moore, 3rd, *Journal of Child Neurology*. 18, 11 (2003).
- [103] N. Pride, J. M. Payne, R. Webster, E. A. Shores, C. Rae, and K. N. North, *Journal of Child Neurology*. 25, 7 (2010).
- [104] M. Greicius, K. Supekar, V. Menon, and R. Dougherty *Cerebral Cortex*. 19, 1 (2009).
- [105] C. Chabernaud, M. Mennes, P. G. Kardel, W. D. Gaillard, M. Kalbfleisch, J. W. VanMeter, R. Packer, M. P. Milham, F. Castellanos, and M. T. Acosta, *Neuroscience Letters*. 515, (2012).
- [106] J. Bennett and J. Wambaugh, Neurofibromatosis Type 1, in *Clinical Management of Sensorimotor Speech Disorders*, Second Edition, M. McNeil, Thieme Medical Publishers: New York: NY. p. 353-354. (2009).
- [107] L. Fenson, V. Marchman, D. Thal, P. Dale, J. Reznick, and E. Bates, *MacArthur-Bates Communicative Development Inventories (CDIs)*, Second Edition, Paul H. Brookes Publishing Co., Inc., Baltimore: MD (2007).
- [108] E. Semel, E. Wiig, and S. WA, *Clinical Evaluation of Language Fundamentals CELF R-4 Screening Test*, The Psychological Corporation, San Antonio: TX (2004).
- [109] T. Klee, K. Pearce, and D. Carson, *Journal of Speech, Language, and Hearing Research*. 43, (2000).
- [110] R. Goldman and M. Fristoe, *Goldman Fristoe Test of Articulation 2: Manual*, American Guidance Service, Inc, Circle Pines: MN (2000).
- [111] N. Bankson and J. Bernthal, *Bankson-Bernthal Test of Phonology*, Pro-Ed, Austin: TX (1990).
- [112] H. Thompson, D. Viskochil, D. Stevenson, R. Turner, S. Groves, and K. Chapman, A comparison of language characteristics of preschool children with and without neurofibromatosis type 1, in *Children's Tumor Foundation Conference 2010*: Baltimore: MD.
- [113] R. Gillam and N. Pearson, *Test of Narrative Language: Examiner's Manual*, Pro-Ed: An International Publisher, Austin: TX (2004).
- [114] S. E. Watt, E. A. Shores, and K. N. North, *Child Neuropsychology*. 14, 5 (2008).
- [115] A. Hosseinaee, Jahad Daneshgahi, Aghajani, and Ahmad Nofresti, *Journal of Iranian Psychologists*. 7, 28 (2011).
- [116] C. Gosse, L. Hoffman, and M. Invernizzi, *Language, Speech, and Hearing Services in Schools*. 43, (2012).
- [117] S. Vogel, *Adults with learning disabilities*, in *Learning disabilities, literacy, and adult education*, S. Vogel and S. Reder, Paul H. Brookes Publishing Co: Baltimore, MD. p. 5-28. (1998).

APPENDIX 1: PREVALENCE OF COMMUNICATION DISORDERS OBSERVED IN THE POPULATION OF INDIVIDUALS WITH NF1 AND THE GENERAL POPULATION (IN PERCENT)

Disorder Area	Age	Expected Range	Observed Rate for Individuals with NF1	
			Range	Median
Voice	P	7% [21]	16% [4]	16%
	S	<1-23 ¹ % [18-22]	22 ² -55% [5, 27]	39%
	A	3-7% ³ [23, 24]	27-40% [5, 31]	35%
Resonance	P		1-42% [4, 38]	22%
	S	3-4% ⁴ [20]	1-52% [5, 27, 38, 41, 42]	22%
	A		1-52% [5, 31, 38, 42]	27%
Fluency	P	1% [43, 44]		
	S	<1-1% [18, 43, 44]		
	A	0-1% [43, 44]	10% [48]	10%
Speech Sound	P	16% ⁵ [50]	12-32% [4, 13, 56, 57]	23%
	S	1-4% [18, 51]	12-19% [13, 56]	16%
	A	1% ⁶ [53]	19-52% [13, 58]	36%
Language	P	13-19% [59-61]	0-86% [4, 83, 97, 98]	24%
	S	7% [63]	0-58% [73, 76, 77, 79, 114]	40%
	A	5% [64]	40% [31]	40%
Reading	P			
	S	2-13% [115, 116]		
	A	3% ⁷ [117]	13% [31]	

P = Preschool, S = School-age, A = Adult

Values in the table are rounded to the nearest percent.

¹ Percent of children exhibiting chronic hoarseness.

² Age of the children was unknown.

³ Participants reported having a current voice disorder.

⁴ Children with communication disorders who exhibited a resonance concern in grades 6-8.

⁵ Prevalence of speech sound disorder in 3 year old children.

⁶ Estimate of residual speech disorder prevalence in college freshman.

⁷ Prevalence of self-reported learning disability.

APPENDIX 2: GLOSSARY OF TERMS

Articulation: The production of speech sounds.

Blocking: An inappropriate cessation of air or voice that may co-occur with discontinued movement of the articulators.

Breathy: Voice produced during lax approximation of the vocal folds.

Devoicing: Production of voiceless phoneme in place of a voiced phoneme (e.g., “*dok*” instead of “*dog*”).

Dysfluency: An inability to exhibit effortless speech production.

Dysphonia: An alteration in normal phonation.

Expressive Language: A communication disorder in which there are problems with the production of language in the verbal or written modality.

Epenthesis: The addition of a segment in a word (e.g., “*belu*” instead of “*blue*”).

Final consonant deletion: Deletion of the final consonant in a word (e.g., “*go*” instead of “*goat*”).

Fundamental frequency: The number of vocal fold vibratory cycles per second.

Hypernasality: Excessive nasal resonance during the production of vowels, vocalic consonants, liquids and glides.

Jitter: Fundamental frequency perturbation.

Mean length of utterance in morphemes: The total number of morphemes in a language sample divided by the number of utterances.

Morphosyntax: Following the development of theoretical linguistics, it is a term used to acknowledge the relationship between children’s development of morphology and syntax. English-speaking children who are older than 4 years of age and are still omitting *third person singular*, *regular* and *irregular past tense*, *auxiliary* and *copular present tense* forms at a high rate may be delayed in their development of morphosyntax. Speakers acquiring languages that are structurally different from English will display deficits consistent with the typology of their native language.

Nasal phonemes: Phonemes requiring only nasal airflow for production (e.g., *m*, *n*, or “*ing*”).

Nasometry: Use of a computer-based instrument to determine an individual’s nasalance score.

Number of different words: The total number of different (types) of words contained in a 50-utterance language sample.

Oral phonemes: Phonemes requiring only oral airflow for production (e.g., *b*, *p*, *s*, *f*, *g*).

Phoneme: The smallest unit of sound used to form meaningful contrasts between utterances.

Phonetic inventory: An individual’s speech sound repertoire.

Phonology: The science of speech sounds and sound patterns.

Phonological analysis: The evaluation and scoring of an individual’s speech for the purpose of identifying patterns of speech sound errors.

Prolongation: A core behavior or stuttering in which a sound or airflow continues simultaneously to the cessation of the movement of the articulators.

Receptive Language: Comprehension or understanding of language.

Resonance (disorder): An imbalance of oral and nasal airflow.

Shimmer: A measurement of loudness perturbation.

Velopharyngeal insufficiency: An anatomical or structural defect preventing appropriate closure of the velopharyngeal port.

Chapter 114

CHILDREN WITH NEUROFIBROMATOSIS TYPE 1: FUNCTIONING IN THE CLASSROOM

***Yafit Gilboa¹, Sara Rosenblum¹,
Aviva Fattal-Valveski² and Naomi Josman¹***

¹Department of Occupational Therapy, Joint program Faculty of Social Welfare & Health Sciences, University of Haifa and Technion, Haifa, Israel

²Pediatric Neurology Unit and the Gilbert Israeli Neurofibromatosis Center (GINFC), Dana Children's Hospital, Tel Aviv Sourasky Medical Center, Tel Aviv University, Tel Aviv, Israel

ABSTRACT

Neurofibromatosis Type 1 (NF1) is a multisystem disorder, and cognitive impairment is its most common complication in childhood (Hyman, Gill, Shores, Steinberg, Joy, Gibikote, & North, 2003). Although patients with NF1 are at risk for significant clinical illnesses, most patients are only mildly affected and live healthy and productive lives (North, 2000). Most research in NF1 to date has been focused on defining the physical and cognitive deficits based on standardized testing, with little emphasis on their functional correlates (Gilboa, Rosenblum, Fattal-Valevski, & Josman, 2010b).

Evidently, student engagement in the classroom reflects, in a large part, self regulation skills which are influenced by temperament and higher order cognitive executive function processes. In particular, sustained attention, which comprises both task-oriented attention and low impulsivity, predicts academic and behavioral adjustment (Pagani, Fitzpatrick, & Parent, 2012).

The purpose of the present chapter is to identify predictors associated with school participation in a sample of children with NF1. A unique feature of this work is the examination of children's participation across different classroom activity components such as handwriting and cognitive skills, such as attention and executive function. Each of these classroom activities may require a unique set of functional skills to meet specific demands. Findings and insights presented in this chapter may help clarify the important factors supporting effective engagement by children with NF1 in school programs.

The chapter describes the functional problems of children with NF1 in the classroom according to the *International Classification of Functioning, Disability and Health*–

Children and Youth version (ICF-CY) (WHO, 2007). The chapter includes current literature on NF1, using the framework provided by the ICF-CY in highlighting the holistic evaluation and interpretation of disabilities. In the first part it presents the ICF-CY (WHO, 2007) conceptual model, defines the term 'ecological validity' and details regarding two ecologically valid assessment tools (Virtual Classroom and functional questionnaire).

The chapter subsequently describes the functional academic profile of children with NF1 from three points of view: attention profile, academic skills in relation to executive profile and handwriting performance. The set of tasks composing each activity provides a useful indication of the factors that may underlie functional participation of children with NF1 in the classroom context. The chapter illustrates the functional profile with a case study and we conclude the chapter with suggestions for further research.

INTRODUCTION

The International Classification of Functioning, Disability and Health—Children and Youth version (ICF-CY) (WHO, 2007)

The World Health Organization's International Classification of Functioning, Disability and Health (ICF) (WHO, 2001) is the most recently developed framework for describing and classifying an individual's health and health-related states. The International Classification of Functioning, Disability and Health – Children and Youth version (ICF-CY) published in 2007 (WHO, 2007) is derived from the ICF (WHO, 2001). The ICF-CY provides a scientific basis for understanding and studying the health and health-related states of children and young people by establishing a common language to facilitate communication and the transfer of information across health professions.

CONCEPTUAL FRAMEWORK OF THE ICF-CY

The goal of the ICF-CY framework is to classify all aspects of health and health-related states. The current framework is a logical approach to viewing diverse aspects of health from biological, individual, and social perspectives (Stucki, Ewert, & Cieza, 2003). The information is organized into two parts: (a) functioning and disability, and (b) contextual factors (WHO, 2001). Each of these parts is further categorized into two components. The functioning and disability section includes: (a) body structures and body functions, and (b) activity and participation. The contextual factors represent the background of an individual's life and include: (a) environmental factors, and (b) personal factors. The main components of the ICF-CY model are summarized in Figure 1.

'Functioning' and 'disability' serve as two contrasting terms encompassing human functioning and disabling conditions. 'Functioning' relates to the successful completion of major day-to-day activities across a broad range of expected roles in daily life. 'Disability' has been described as an inability to perform these critical daily living activities in the normal range as a result of impairment (WHO, 2001).

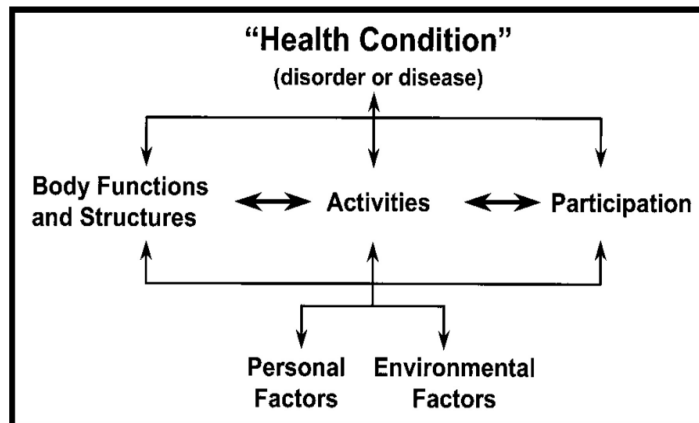


Figure 1. International Classification of Functioning, Disability and Health (ICF). (Adapted from WHO, 2001).

Body structures and functions represent the different body systems and their specific roles. The activity and participation component refers to an individual's performance of activities, either physical or mental, that are associated with all aspects of life. The successful completion of activity and participation involves the integration of body structures and functions in a purposeful manner within various contexts, including the physical, social, and attitudinal environments (WHO, 2001).

The environmental component within the contextual part comprises the physical, social and attitudinal aspects of the environment in which people live and conduct their lives. Personal factors constitute features related to the individual, such as gender, age, fitness, coping abilities, and social background that could play a role in disability at any level (WHO, 2001). The ICF-CY model encourages an interactive view of functioning and disability, in which both personal and environmental factors are often interrelated in a dynamic way (Cramm, Aiken, & Stewart, 2012).

THE ECOLOGICAL VALIDITY OF COGNITIVE ASSESSMENTS TOOLS

In its general sense, ecological validity is the degree to which results obtained in controlled experimental conditions are related to those obtained in naturalistic environments. In the context of neuropsychological testing, ecological validity refers to the degree to which test performance corresponds to real world performance, or the individual's ability to generalize results of controlled experiments to naturally occurring events (Gioia & Isquith, 2004). Validity does not apply to the test itself, but to the inferences that are drawn from the test. Therefore, tests that have adequate diagnostic validity do not necessarily have adequate ecological validity (Chaytor & Schmitter-Edgecombe, 2003).

Evaluating the ecological validity of neuropsychological tests has become an increasingly important topic over the past decade (Chaytor & Schmitter-Edgecombe, 2003). With the development of brain-imaging techniques, neuropsychological testing has progressed. From facilitating brain pathology diagnosis the aims of the tests have changed to allow the description of functional strengths and weaknesses (Lezak, Howieson, Bigler, & Tranel, 2012), the

prediction of everyday functioning, and the need for intervention and support in the natural environment (Chaytor & Schmitter-Edgecombe, 2003). It is suggested that a parallel shift from “traditional” validity considerations to ecological validity considerations need to follow (Gioia & Isquith, 2004).

Clinical predictions are made as to how people will most likely perform in natural settings in relation to their poor performance on certain tests. However, these predictions are seldom based on any strong scientific evidence. In a review of the literature on ecological validity of neuropsychological tests, it was found that in most cases the magnitude of relationship was only moderate ($r=0.30$) and many neuropsychological tests were not related to measures of outcome (Chaytor & Schmitter-Edgecombe, 2003).

When neuropsychological tests are developed and applied to identify or quantify deficits, traditional validity (e.g., construct validity) is paramount and ecological validity may be of little concern (Gioia & Isquith, 2004). Implicit in the definition of ecological validity as its concept is applied to innovative assessment tools, is, that performance on a test predicts some aspect of the child’s functioning on a day-to-day basis (Chaytor & Schmitter-Edgecombe, 2003; Gilboa et al., 2010b).

EVALUATION OF COGNITIVE AND FUNCTIONAL PROBLEMS

Constant advances in test development has resulted in increased knowledge of cognitive problems in all disorders, yet many functional areas remain beyond the reach of formal assessment tools. As remediation needs to be based on what is occurring in the classroom and in daily life, in greater detail than a neuropsychological test alone can provide (Chaytor & Schmitter-Edgecombe, 2003), understanding of these real world functional deficits, is essential in setting out remediation guidelines (Lezak et al., 2012).

Despite the broad range of deficits present in children with NF1, research into functional assessment remains largely unexplored. The use of neuropsychological testing to understand the cognitive deficits of children with NF1 is somewhat limited. There is a lack of studies that can provide vital information on the impact of cognitive and other deficits in daily life (Gilboa et al., 2010b).

The Use of Virtual Reality Environments and Questionnaires to Evaluate Day-to-Day Cognitive Functioning

During the past decade, ongoing advances in computer graphics, software, and hardware design have refined medical simulators to offer life-like replications of medical and surgical procedures in a variety of medical specialties (Willaert, Aggarwal, Van Herzeele, Cheshire, & Vermassen, 2012).

Virtual Reality (VR)

Virtual Reality (VR) technology delivers something beyond the scope of currently available computerized flat screen assessments, as it can be used to immerse and test a user within a dynamic, three-dimensional, ecologically valid stimulus environment under conditions

more similar to the challenges of the “real” world (Rizzo, Bowerly, Buckwalter, Klimchuk, Mitura, & Parsons, 2006). VR technology offers the option to produce and distribute identical “standard” simulation environments in which an individual’s performance can be measured. Digital scenarios allow normative data to be accumulated for performance comparisons needed for assessment/diagnosis and for treatment/rehabilitation purposes. The capacity of the VR to create dynamic, immersive three-dimensional stimulus environments that record all behavioral responses, offers assessment options that in traditional assessment methods are simply not available (Josman, Milika Ben-Chaim, Friedrich, & Weiss, 2008). It even has advantages over real-world behavioral observations, as it provides a controlled stimulus environment where cognitive challenges can be presented along with the precise delivery and control of distracting auditory and visual stimuli, and real-world responses can be objectively measured (Rizzo et al., 2006).

The last decade has been an enormously exciting time for neuropsychological VR applications. Research has now evolved to a point where VR is regarded as an invaluable tool in examining the neural correlates of everyday cognition (Penn, Rose, & Johnson, 2009). The degree of stimulus control and sensitivity of monitoring afforded by VR is also a significant advantage when examining subtle behaviors, such as head movements in children with ADHD (Parsons, Bowerly, Buckwalter, & Rizzo, 2007).

Functional Questionnaires

Functional questionnaires provide an additional means of understanding and expanding the cognitive profile of children with NF1. These functional questionnaires describe difficulties in day-to-day functioning in the type of detail that a single standard score on a psychometric measure is unable to provide. They also provide a measure of more than a single sample of performance (unlike neuropsychological tasks) and take into account performance in less “ideal” situations where normal distractions are present and compensatory strategies may be utilized (Fingerhut, Madill, Darrah, Hodge, & Warren, 2002; Verkerk, Wolf, Louwers, Meester-Delver, & Nollet, 2006).

Parents and teachers possess a wealth of information about children’s behavior in those settings that are directly relevant to an understanding of function (Gioia & Isquith, 2004). Parent’s and teachers’ functional questionnaires provide an additional means of understanding and expanding the cognitive profile of children in general and specifically children with NF1. For example, The Behavior Rating Inventory for Executive Functioning (BRIEF) (Gioia, Isquith, Guy, & Kenworthy, 2000) was developed with ecological validity in mind. The impetus for the BRIEF came from the need to more efficiently and systematically capture information about manifestations of executive function difficulties in children’s everyday behaviors at home, in school, and in their communities (Gioia & Isquith, 2004). This functional questionnaire describes difficulties in day-to-day functioning, provides a measure of more than a single sample of performance, and takes into account performance in less “ideal” situations where normal distractions are present and compensatory strategies may be utilized. It was designed to question the parents about their child’s everyday life situations while the examiner reaches his conclusions regarding his EF (Fingerhut et al., 2002; Verkerk et al., 2006).

Describing the Attention Deficit profile of Children with NF1 Using a Virtual Classroom Environment (Gilboa, Rosenblum, Fattal-Valevski, Toledano-Alhadeff, Rizzo, & Josman, 2011a; Gilboa, Rosenblum, Fattal-Valevski, Toledano-Alhadeff, Rizzo, & Josman, 2011b)

The Virtual Classroom (VC) (Rizzo et al., 2006), a head-mounted display (HMD) immersive VR system, was developed for the assessment of attention skills using a virtual environment that simulates a classroom. Recent findings have provided support for the construct validity of this VC in children with ADHD (Adams, Finn, Moes, Flannery, & Rizzo, 2009; Parsons et al., 2007; Rizzo et al., 2006). Concurrent validity has also been demonstrated by correlating VC results with other widely used ADHD assessment tools (Adams et al., 2009; Parsons et al., 2007; Rizzo et al., 2006). These studies suggest that the VC has good potential for controlled performance assessment within an ecologically valid environment and appears to parse out significant effects due to the presence of distraction stimuli (Adams et al., 2009; Parsons et al., 2007; Rizzo et al., 2006).

The research version of the VC scenario consists of a standard rectangular classroom environment containing desks, a female teacher, a blackboard across the front wall, a side wall with a large window, and on the opposite wall, two doorways (Figure 2). The scenario runs on a standard PC with a head mounted display with onboard orientation tracking.



Figure 2. The virtual classroom.

The task required the child to tap the left mouse button as quickly and accurately as possible, using their dominant hand when the digit sequence “37” appeared on the VC blackboard. The stimuli remained on the screen for 150 ms with a fixed interstimulus interval of 1350 ms. Participants were instructed to withhold their responses to any other sequence of digits. The test lasted 10 minutes (5 identical blocks of 2 minutes) during which 400 stimuli (100 of them were “37” sequences) were presented accompanied by 20 distractors (e.g., pure audio [classroom noises], pure visual [paper airplane flying across the visual field] and mixed audiovisual [a car “rumbling” by a window, a person walking into the classroom with hall sounds when the door opened]). Distractors were each displayed for 5 seconds and identically presented in the entire sample.

The participants' reaction patterns were recorded and documented using four measures: 1. *Total correct hits*- the ability to correctly identify 100 targets out of 400 stimuli during 10 minutes (correct and incorrect hits, measured as raw score out of 100), 2. The number of

commission errors. 3. *Reaction time* was measured in milliseconds and 4. *Head movements* were captured and quantified from the inertial tracking system. This is used to sense head orientation as the input signal. Head movements are sensed by this system and quantified in three orientation axes.

The study was the first to use the VC environment to assess attention processes in NF1 children. The objectives of the study were (1) to compare the performance of children with NF1 and the control group in the VC and (2) to assess the utility of the VC to detect attention deficits in a sample of NF1 children in comparison with a widely used ADHD screen questionnaire.

Participants included 29 children diagnosed with NF1 according to the NIH criteria (9 males, 20 females; mean age 12.2 ± 2.5). The control group consisted of 25 Typically Developing (TD) children matched to the study group by gender and age (7 males, 18 females; mean age 12.2 ± 2.6). The groups did not significantly differ in mean age, school year, gender and handedness. As expected, the IQ score of the NF1 group (13.0 ± 98.9) was lower than the control group (12.3 ± 109.3). However, this gap did not differ significantly. In addition to the VC the Conners' Parent Rating Scales-Revised: Long (CPRS-R:L) (Conners, 1997) was administered.

The results from the assessment of the VC task, showed that the performance of the NF1 group was significantly lower than the controls on *total correct hits* (mean 74.37 ± 21.64 vs. 87.40 ± 10.48 ; $p < 0.01$) and the number of *commission errors* (mean 20.27 ± 23.03 vs. 5.87 ± 5.34 ; $p < 0.01$). The NF1 group responded correctly to fewer targets and committed more errors than the control group on the entire task and in each 2 minute block. No significant differences were found on *reaction time*. NF1 participants were also not significantly more hyperactive compared to controls on all three measures of *head movement* (captured from the tracking system that is part of the HMD) (Table 1).

The number of targets correctly identified by the two groups' declined as a function of the time, (Figure 3). However, the pattern of the commission errors group was stable and didn't significantly change as a function of time (Figure 4).

For the NF1 group, a significant negative correlation was found between the VC's total correct hit and the cognitive problems/inattention clinical scale of the CPRS-R: L ($r = -0.42$, $p < 0.05$). Thus, more correct hits negatively correlated with lower scores (i.e.: better attention skills) on this measure.

The difficulty to correctly identify 100 targets out of 400 stimuli during 10 minutes can be explained by a deficit in sustained attention (Lavine, Sibert, Gokturk, & Dickens, 2002) and/or switching attention. Support for this hypothesis is evident in the present study, as more correct hits, significantly correlated negatively with the cognitive problems/inattention clinical scale of the CPRS-R: L. The linear decreasing pattern of attention as exemplified in figure 3 clearly suggests difficulties in continuous attention. However, since both groups showed a trend towards declining performance over time, this result may be related to a "fatigue effect".

A significant difference was also found between the NF1 and control group on the number of commission errors, with NF1 children producing more errors than controls. The NF1 group responded less accurately to the stimulus due to impulsive responses in the absence of a target. This impulsivity pattern, characteristic of the NF1 group, was found to be stable as a function of time.

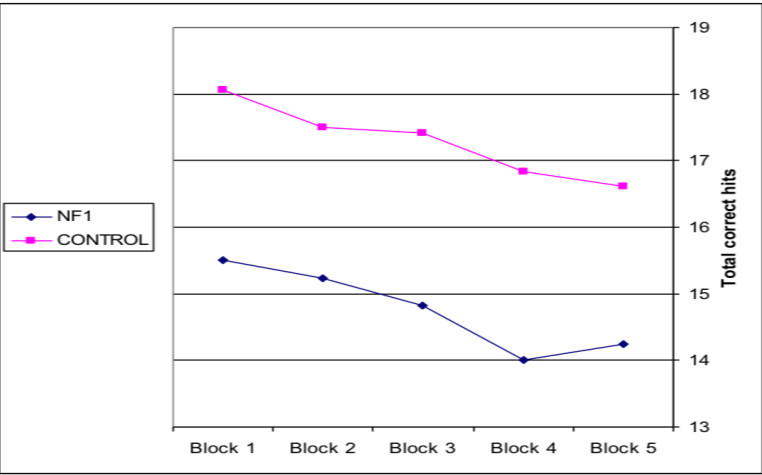


Figure 3. Total correct hits in relation to time.

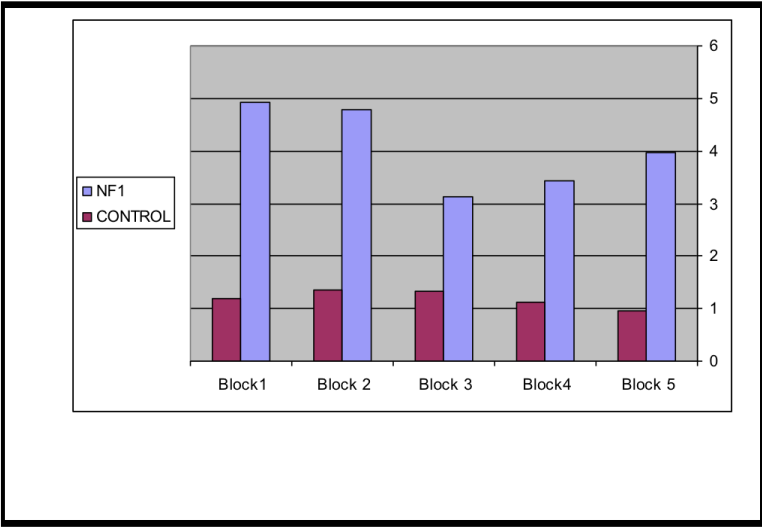


Figure 4. Commission errors in relation to time.

No significant difference was found between the groups on the *reaction time* to targets. From the results of the present study we conclude that such differences in processing speed (the response time between stimulus and reaction) as suggested in children with ADHD (Parsons et al., 2007), were not found in children with NF1. Also, no significant differences were observed in all three measures of *head movements* in participants with NF1 as compared to the TD control group. These findings lead to the conclusion that the attention deficit profile of children with NF1 does not include increased overall hyperactivity.

To conclude, the VC results support the hypothesis that NF1 is marked by inattention (omission errors) and impulsivity (commission errors). The VC appears to be a sensitive and ecologically valid assessment tool for use in the diagnosis of attention deficit in children with NF1.

The EF of Children with NF1 and its Relation to their Academic Skills

A study that explored the impact of NF1 on school performance clearly demonstrates that children with NF1 are severely affected. At least 75% of the children with NF1 have one or more learning disabilities in technical reading, comprehensive reading, spelling, or mathematics. In comparison with the average population, children with NF1 attended more special education classes with an odd ratio of 4:1; repeated a grade in their school career (17% vs 1.9%), and received remedial teaching for learning problems (85% vs 15%) (Krab, Aarsen, de Goede-Bolder, Catsman-Berrevoets, Arts, Moll, & Elgersma, 2008).

Executive functions (EF) are a set of interrelated cognitive and behavioral skills that activate and govern our conscious perceptions, feelings, thoughts and actions, constituting a collection of “co-conductors” (McCloskey, Perkins, & Van Divner, 2008). It leads the individual to act in a purposeful, organized, strategic, self-regulated goal-directed manner. EF can be divided into discreet sub-skills that include setting and managing goals, planning, inhibition and dealing with diverse elements, shifting among cognitive and affective sets, organization, working memory and meta-cognition (Ylvisaker & Feeney, 2002). EF is a key metacognitive skill underpinning successful goal directed behavior, and is linked to educational attainment (St Clair-Thompson & Gathercole, 2006).

Recently, efforts have been made to develop more specific tests of executive ability, representing a departure from tests that have discriminative validity for diagnostic purposes, to ecologically valid assessments, in which test results are able to predict functional abilities relevant to real world functioning (Wood & Lioffi, 2006). In a study that aimed to examine the extent to which ecologically valid measures of EF are useful in predicting school functioning, two types of EF assessments were used:

- 1) Performance-based assessment: A standardized test that shows promise to be ecologically valid, to test EF performance among children, the Behavioral Assessment of the Dysexecutive Syndrome for Children (BADS-C) (Emslie, Wilson, Burden, Nimmo-Smith, & Wlson, 2003).
- 2) Everyday functional assessment: A questionnaire that evaluates everyday demands of EF in the natural setting, the Behavioral Rating Inventory of Executive Functions (BRIEF) (Gioia, Isquik, Guy, & Kenworthy, 2000).

In order to look at school performance, we used the Academic Competence Evaluation Scales (ACES) (DiPerna & Elliott, 2000) and teacher's questionnaire, as an outcome measure. A child's overall level of academic competence is dependent upon a set of five dynamic constructs which include academic skills, study skills, interpersonal skills, engagement, and motivation. These five dynamic constructs can be categorized as either Academic Skills or Academic Enablers (DiPierna & Elliott, 1999). Academic skills refer to the child's mastering of specific reading and language arts, mathematics and critical thinking skills. The other four constructs of motivation, engagement, study skills and interpersonal skills, constitute the Academic Enablers.

Twenty-nine children with NF1 and 27 age-and-gender-matched controls (age 8-16) were examined. The performance of the NF1 group was significantly lower on the *Water* ($F(1, 57) = 4.85, p < .05, d = .57$) and the *Key Search* ($F(1, 58) = 5.47, p < .05, d = .47$) subtests of the BADS-C. After controlling for estimated IQ, significant differences were found between the groups

on the reported *Working Memory* ($F(1, 42) = 11.08, p < .001, d = 1.03$) and the *Plan/Organize* ($F(1, 42) = 6.49, p < .05, d = .67$) scales as well as the *Meta-cognition Index* ($F(1, 42) = 4.21, p < .05, d = .82$) and the *Global Executive Composite* ($F(1, 41) = 7.21, p < .05, d = .81$) of the BRIEF, and on the *Academic Skills* scale of the ACES ($F(1, 26) = 4.86, p < .05, d = 1.96$). Varieties of significant correlations were found between the BADS-C's subtests and the BRIEF's scales and the teacher's reports on the ACES. Significant predictive models were generated for BADS-C, BRIEF and ACES scores.

The data from this study, in combination with other studies (Mazzocco, Turner, & Denckla, 1995; Payne, Hyman, Shores, & North, 2011; Rowbotham, Pit-ten Cate, Sonuga-Barke, & Huijbregts, 2009) strongly support the existence of a specific profile of executive dysfunction in the NF1 group of children, beyond the decrease in IQ. The executive dysfunction profile of children with NF1 is characterized by a combination of deficits in the metacognitive skills of planning, problem solving, and working memory. Significant predictive models were generated for BADS-C, BRIEF and ACES scores, suggest that this EF profile is related to difficulties that children with NF1 experience at school in the areas of academic and social functions.

The Activity of Handwriting

Handwriting is an important task learned during early school years. A high proportion of a child's school day is dedicated to writing assignments, an essential ingredient for success at school (McHale & Cermak, 1992). Writing is indispensable for participation in school activities, such as taking notes and tests, doing homework, and writing papers (Weintraub, Drory-Asayag, Dekel, Jakobovits, & Parush, 2007). This complex perceptual-motor skill requires adequate performance in visual-motor coordination, motor planning, cognitive/perceptual skills, and tactile and kinesthetic sensitivities (Maeland, 1992). There is evidence of deficits in cognitive/perceptual skills (Hyman *et al.*, 2003), motor skills (Billingsley, Slopis, Swank, Jackson, & Moore, 2003; Hofman, Harris, Bryan, & Denckla, 1994; Levine, Materek, Abel, O'Donnell, & Cutting, 2006) and visual-motor integration skills (Levine *et al.*, 2006), skills in children with NF1 which may interfere with handwriting competency.

Very little research on the nature and cause of writing difficulties is available (Rosenblum, Parush, & Weiss, 2003) despite the fact that writing production tends to be the most neurologically vulnerable language modality (Lorch, Ferner, Golding, & Whurrd, 1999). Non-proficient handwriting appears to have considerable apparent consequences on a child's academic performance and emotional well-being and persists longer than other symptoms (Lorch, 1995).

The Handwriting Performance of Children with NF1 (Gilboa, Josman, Fattal-Valevski, Toledano-Alhadeff, & Rosenblum, 2010a)

The handwriting performance of children with NF1 was evaluated following the developmental model as describe by Feder & Majnemer (2009). The evaluation included multilevel developmental and functional assessments, starting with the ability to copy geometric forms using the Beery-Buktenica Developmental Test of Visual-Motor Integration

(VMI) (Beery, 2004) and was followed by two common functional everyday writing tasks: paragraph copying and free style written text. Previous studies indicated that the VMI is a significant predictor of handwriting legibility on a copying task (Daly, Kelley, & Krauss, 2003) and evidence of significant impairment on VMI score was previously seen among children with NF1 (Hyman et al., 2003). Therefore the relationship between the VMI standardized score and the subsequent writing measures were examined.

In our study, the children's writing performance were evaluated at three levels in two writing tasks: the mechanical aspects of both writing tasks were evaluated with a Computerized Penmanship Evaluation Tool (CompPET), including an electronic writing tablet (digitizer) (Rosenblum et al., 2003), that enabled collection of spatial, temporal, and pressure data of the child's writing *process*. The CompPET, has been demonstrated to be highly valuable with children with specific learning disability (Rosenblum, Weiss, & Parush, 2004). Then, *the product's legibility* of the copying task was evaluated by administering the Hebrew Handwriting Evaluation (HHE), (Erez & Parush, 1999) a standardized, reliable and valid handwriting assessment for the Hebrew language. Finally, the handwriting product *content* of free style writing was assessed with the Six Trait Writing Method (Spandel, 2004).

The objectives were to analyze the process and product of handwriting and the relationship between visual-motor integration and features of handwriting that characterize performance of children with NF1 in comparison to those of TD children.

A total of 60 children were tested: The study group consisted of 30 children diagnosed with NF1 according to the NIH criteria (9 males, 21 females; mean age 12y 3mo, $\pm$ 2y 6mo; range 8y -16y 8mo). The control group consisted of 30 TD children matched to the study group by gender and age (mean age 12y 4mo, $\pm$ SD 2y 5mo; range 8y 5mo-16y 4mo). All participants used the Hebrew language as their primary means of verbal and written communication.

We found a significant difference between the groups for the VMI standardized scores ($t(1, 58) = -3.17, p < 0.01$). The NF1 ($M=86.93 \pm 17.17$) children performed worse than the TD ($M=101.90 \pm 14.74$) group. For the study group, significant correlation was established between the VMI and the free style writing task's organization ($r=0.41, p<0.05$).

When considering the handwriting mechanical process results, surprisingly, no significant differences were found between children with NF1 and TD in temporal measures for writing velocity on-paper and in-air time per stroke (CompPET variables) as well as for pen pressure. This finding may stem from the fact that the groups included a female majority (21 girls versus 9 boys). However, a significant difference was established between the groups in the measure of mean stroke height within the copying task ($F(1,57)=5.18, p<0.05$).

Significant correlations were also found between the mean velocity, total duration per stroke and in-air time per stroke and the number of unrecognizable letters on the handwriting product ranging from .37 to .58. Children with NF1, who exhibited more unrecognizable letters, wrote faster, spent considerably more total time on the task and spent greater in-air time. In-air writing occurred with significantly greater frequency for poor writers than for the proficient writers in most handwriting tasks. This is highly indicative of EF such as planning (Rosenblum et al., 2003), visual spatial organization (Rosenblum & Livneh-Zirinski, 2008) and the perceptual aspect of the motor act related to writing (Werner, Rosenblum, Bar-On, Heinik, & Korczyn, 2006) such as motor memory for letter formation or difficulty in visualizing letters needed and to form them rapidly (Rosenblum & Livneh-Zirinski, 2008).

The performance of the children with NF1 in the spatial arrangement of their handwriting product (HHE variable) ($F(1,57)=4.24, p<0.05$) as well as in the level of the content for ideas,

organization, voice, word choice, sentence fluency and conventions (Six-Trait Writing Model variables), was significantly poorer than for the children without NF1 ($F(6, 52) = 3.42, p < 0.01$). Significant negative correlations were found between the number of letters erased and/or overwritten on the copying task and two factors of the Six-Trait Writing Model: Ideas ($r = -0.422, p < 0.05$) and organization ($r = -0.52, p < 0.01$) and the total score ($r = -0.37, p < 0.05$) for the free style writing task.

To summarize, evaluation of NF1 children with both process and product measures, provides an impression of their handwriting performance. Handwriting product was found to be poorer among children with NF1 in terms of content and spatial arrangement. These difficulties come to fruition in the different tasks, from figure copying to impaired handwriting.

We provide a case study to describe a functional profile of a child with NF1 in the classroom. A summary report of the assessment process highlighted his difficulties in EF and writing, and the goals for the intervention are illustrated.

Jonathan: Case Study

Jonathan, a boy aged 7 years and 4 months, is a grade one student. He has been attending special education preschool for the last 4 years, and now attends a special education class in a public school. Jonathan is the second son in his family of five children, who live in an urban area. His family is of moderate-low socioeconomic status.

Jonathan's birth was a regular delivery following a normal pregnancy. As an infant, his motor and language performance were delayed for his chronological age. At the age of 6 he was tested by a psychologist. His intellectual functioning was found to be at the poor range of the norm ($IQ = 83$), with a significant gap between language (94) to performance (73) intelligence. By the age of 3 he was diagnosed as a case of NF1 by a pediatrician.

He was referred to the Occupational Therapy service in his school because he was experiencing frequent handwriting problems and clumsiness. Jonathan's teacher reported that his achievements in reading, understanding and mathematic were adequate as compared to the other students in his class.

His parents are concerned mainly about his extensive writing difficulties, as will be detailed later in this profile; they also describe his difficulties in organizing materials. His parents perceived his academic performance to be lower than appropriate for his age. He avoided daily learning activities like homework and when he experienced difficulty, he immediately gave up.

Referral to Occupational Therapy

The referral was made by his teacher. During a preliminary interview with the Occupational therapist, Jonathan's teacher mentioned the following concerns:

- Jonathan presents difficulties in effectively using his learning tools including notebooks, pencil case and bag.
- Jonathan presents difficulties in organizing the text he writes within lines; the size of the letters is too big.
- During class, Jonathan copies text from the blackboard in an imprecise way: omits letters, connecting words and lines.

- Legibility of both letters and digits are poor.
- Jonathan is unfocused and appears to be preoccupied by other things.

Assessment tools:

- The following assessment tools were used:
- Background questionnaire for collecting data regarding Jonathan's functioning and reasons for referral.
- Interview with Jonathan's parents to assess his level of participation in all performance domains and to help choose appropriate tools for him.
- Observation in Jonathan's class and an interview with his teacher to evaluate his level of participation at school.
- The BRIEF parent and teacher questionnaire to Assess executive function behaviors in the school and home environments.
- The developmental test of Visual-Motor integration- Revised to obtain deeper sensory-motor evaluation.

Major Findings and Discussion

The occupational therapists impression based on the performance evaluations was as follows:

Executive Function

Jonathan performed activities in the right sequence but displayed difficulty in flexibility: moving from one stage to another. He completed a task quickly and with carelessness and did not check his performance. He did not initiate a discussion, but responded to the questions that were asked. He showed difficulty in inhibiting his reactions when require to do so and to solve problems with minimal mediation.

The BRIEF questionnaire reveals that Jonathan's parents indicated major difficulties in inhibition, shifting, organizing of materials, monitoring and behavioral regulation (see figure 5). Jonathan's teacher also indicated major difficulties in inhibition, emotional control, planning and organization of materials, monitoring, as well as behavioral regulation, metacognition and the global executive composite (see Figure 6).

Attention

Jonathan did not wait until the teacher finished talking, resulting in her often having to stop him from starting a task. Since he didn't listen to instructions, Jonathan asked the teacher for many cues (he said that he did not remember the instructions) and used them to complete the tasks. His ability to use cues such as "What do you need to do next?" were very helpful.

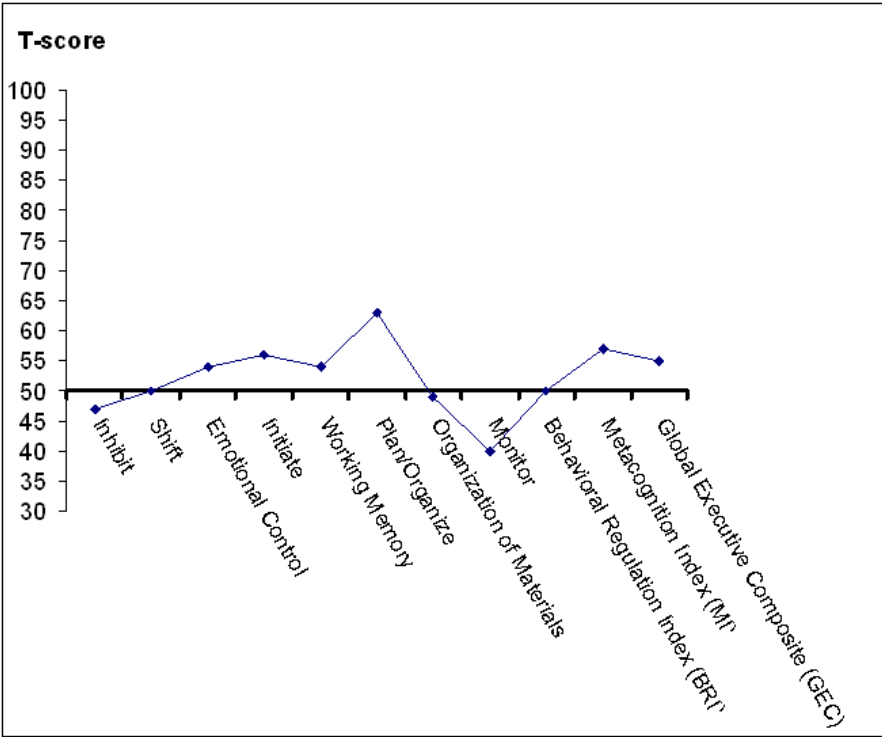


Figure 5. Jonathan's parent profile on the BRIEF.

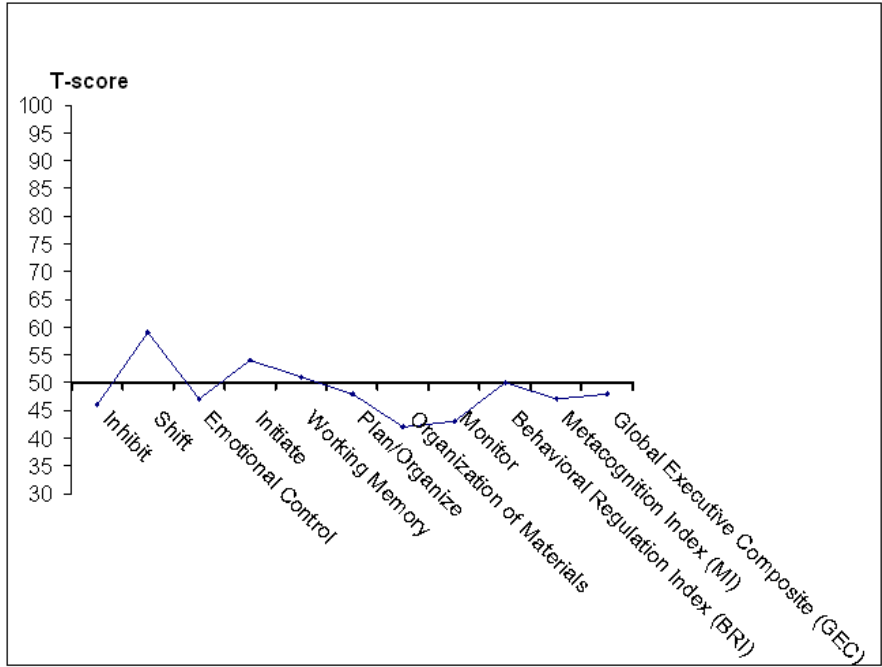


Figure 6. Jonathan's teacher profile on the BRIEF.

Writing

Jonathan displayed difficulties in spatial organization that negatively affected his writing legibility. For example, he wrote the text in an inappropriate place and did not use the margins.

Jonathan's writing difficulties also stem from a motor basis. He had difficulty generating appropriate muscle force when using objects that required gentle force (like pencil). He displayed difficulty in isolating movements of the wrist and put inconsistent pressure on writing tools.

On the Beery test, Jonathan received lower scores than age appropriate, both in the motor and visuomotor sections. Jonathan's score on the visual perception subtest of the Beery was age appropriate.

Emotionally, Jonathan has the ability to interact properly with adults and peers and has good internal motivation to perform whatever is required. He has a low sense of competence regarding his performance. He repeatedly said "I did not understand "or "I did it badly".

Recommendations

On the basis of the results of the evaluation, occupational therapy intervention was recommended. It was recommended that Jonathan receive 10 45-minute individual sessions of occupational therapy intervention, once a week.

Occupational therapy intervention goals are as follows:

Focus on functional treatment goals for meaningful occupations that Jonathan's parents and teachers noted during interviews. Activities that restricted typical participation included Jonathan's difficulties to complete handwriting assignments in a satisfactory way, his refusal to do homework and his difficulties in operating his writing tools.

Focus on Jonathan's emotional and executive deficits as follows:

- analyzing, together with Jonathan's parents and teacher, how difficulties in his attention and executive skills (inhibition, shifting, monitoring planning and organizing) and emotional factors limit the above occupations.
- consulting and coaching Jonathan's parents and teachers, in order to teach them strategies to help with the following goals: (1) help Jonathan to be more attentive while receiving instructions and performing tasks independently; (2) reduce external and internal distracters; (3) enable Jonathan to use self inhibition during class to help him follow instructions; (4) gradually enable Jonathan to use these strategies on his own.

This case study was provided by Michal Tsadok-Cohen and Yafit Gilboa.

CONCLUSION

This chapter identified specific functional problems that are significantly associated with difficulties in achieving successful participation in different school activities for children with NF1. The problem profile includes special deficits in sustaining attention, impulsivity, and low academic achievements that are associated with EF deficits and poor handwriting in terms of content and spatial arrangement.

Findings reported in this study do not imply a causal relationship between predictors and outcome. However, they can be used as a guide to help clinicians prioritize their clinical reasoning. The results strongly support an important assumption of the ICF-CY classification Model (WHO, 2007) and the current trends in rehabilitation toward more reliance on functional measurement rather than clinical diagnostic data to plan for children's participation in mainstream school programs (Pagani et al., 2012).

The results also identify possible pathways of influence associated with limited and full participation outcomes that should be considered during such intervention planning. For example, information about children's sustained attention and handwriting capabilities, may serve as a primary indicator to identify those who are at higher risk for limited participation at school.

Future research should aim to evaluate the participation of children with NF1 in various activities using tools that have high real-world validity and focus on the functional implications of the problems. Future research will contribute to a better understanding of the abilities and disabilities of children with NF1 in their functioning at school and at home. This knowledge would lead to increased identification of problems for planning intervention methods to satisfy the needs of this unique population, and to truly improve their quality of life; interventions that would focus on making a difference to the everyday experiences of individuals with NF1 and their families. Future research will link between understanding real-life implications and brain functioning and may also improve the understanding of the underlining mechanisms.

REFERENCES

- Adams, Finn, P., Moes, E., Flannery, K., & Rizzo, A. S. (2009). Distractibility in Attention/Deficit/ Hyperactivity Disorder (ADHD): the virtual reality classroom. *Child Neuropsychology*, 15(2), 120-135.
- Beery, K. (2004). *The Beery-Buktenica Developmental Test of Visual-Motor Integration: VMI with Supplemental Developmental Tests of Visual Perception and Motor Coordination: Administration, Scoring and Teaching Manual*. Parsippany, NJ: Modern Curriculum Press.
- Billingsley, R. L., Slopis, J. M., Swank, P. R., Jackson, E. F., & Moore, B. D., 3rd. (2003). Cortical morphology associated with language function in neurofibromatosis, type I. *Brain and Language*, 85(1), 125-139.
- Chaytor, N., & Schmitter-Edgecombe, M. (2003). The ecological validity of neuropsychological tests: a review of the literature on everyday cognitive skills. *Neuropsychology Review*, 13(4), 181-197.
- Conners, C. K. (1997). *Conners' Rating Scales-revised*. New York: Multi-Health Systems Inc.
- Cramm, H., Aiken, A. B., & Stewart, D. (2012). Perspectives on the International Classification of Functioning, Disability, and Health: Child and Youth Version (ICF-CY) and Occupational Therapy Practice. *Physical & Occupational Therapy in Pediatrics*.
- Daly, C. J., Kelley, G. T., & Krauss, A. (2003). Relationship between visual-motor integration and handwriting skills of children in kindergarten: a modified replication study. *The American Journal of Occupational Therapy*, 57(4), 459-462.
- DiPerna, J. C., & Elliott, S. N. (2000). *Academic competence evaluation scales*. San Antonio, TX: Psychological Corporation.
- DiPierna, J. C., & Elliott, S. N. (1999). Development and validation of the Academic Competence Evaluation Scales. *Journal of Psychoeducational Assessment*, 17, 207-225.

- Emslie, H., Wilson, F. C., Burden, V., Nimmo-Smith, I., & Wilson, B. A. (2003). *Behavioral Assessment of the Dysexecutive Syndrome in Children (BADS-C)*. London: Thames Valley Test Company.
- Erez, N., & Parush, S. (1999). *The Hebrew Handwriting Evaluation* (2nd ed.). Israel, Jerusalem: School of Occupational Therapy. Faculty of Medicine. Hebrew University of Jerusalem.
- Fingerhut, P., Madill, H., Darrah, J., Hodge, M., & Warren, S. (2002). Classroom-based assessment: validation for the school AMPS. *The American Journal of Occupational Therapy*, 56, 210-213, (2)
- Gilboa, Y., Josman, N., Fattal-Valevski, A., Toledano-Alhadeff, H., & Rosenblum, S. (2010a). The handwriting performance of children with NF1. *Research in Developmental Disabilities*, 929-935, 31,
- Gilboa, Y., Rosenblum, S., Fattal-Valevski, A. & Josman, N. (2010b). Application of the International Classification of Functioning, Disability and Health in children with neurofibromatosis type 1: a review. *Developmental Medicine & Child Neurology* 612- , 52 , 619
- Gilboa, Y., Rosenblum, S., Fattal-Valevski, A., Toledano-Alhadeff, H., Rizzo, A., & Josman, N. (2011a). *Describing the Attention Deficit profile of Children with Neurofibromatosis Type 1 Using a Virtual Classroom Environment*. Paper presented at the International Conference on Virtual Rehabilitation 2011, Zurich, Switzerland, (pp.1 - 5) doi:10.1109/ICVR.2011.5971826.
- Gilboa, Y., Rosenblum, S., Fattal-Valevski, A., Toledano-Alhadeff, H., Rizzo, A. S., & Josman, N. (2011b). Using a Virtual Classroom environment to describe the attention deficits profile of children with Neurofibromatosis type 1. *Research in Developmental Disabilities*, doi:10.1016/j.ridd.2011.06.014.
- Gioia, Isquith, P., Guy, S., & Kenworthy, L. (2000). *Behavior Rating Inventory of Executive Function- Professional Manual*. Odessa, Florida: Psychological Assessment Resources.
- Gioia, & Isquith, P. K. (2004). Ecological assessment of executive function in traumatic brain injury. *Developmental Neuropsychology*, 25(1-2), 135-158.
- Hofman, K. J., Harris, E. L., Bryan, R. N., & Denckla, M. B. (1994). Neurofibromatosis type 1: the cognitive phenotype. *The Journal of pediatrics*, 124(4), S1-8.
- Hyman, S. L., Gill, D. S., Shores, E. A., Steinberg, A., Joy, P., Gibikote, S. V., et al. (2003). Natural history of cognitive deficits and their relationship to MRI T2-hyperintensities in NF1. *Neurology*, 60(7), 1139-1145.
- Josman, Milika Ben-Chaim, H., Friedrich, S., & Weiss, P. L. (2008). Effectiveness of Virtual Reality for Teaching Street-Crossing Skills to Children and Adolescents with Autism. *Journal on Disability and Human Development*, 7(1), 49-56.
- Krab, L. C., Aarsen, F. K., de Goede-Bolder, A., Catsman-Berrevoets, C. E., Arts, W., F., Moll, H. A., et al. (2008). Impact of Neurofibromatosis Type 1 on School Performance. *Journal of Child Neurology*, 23, 1002.
- Lavine, R. A., Sibert, J. L., Gokturk, M., & Dickens, B. (2002). Eye-tracking measures and human performance in a vigilance task. *Aviation, Space, and Environmental Medicine*, 73(4), 367-372.
- Levine, Materek, A., Abel, J., O'Donnell, M., & Cutting, L. E. (2006). Cognitive profile of neurofibromatosis type 1. *Seminars in Pediatric Neurology*, 13(1), 8-20.
- Lezak, M. D., Howieson, D. B., Bigler, E. D., & Tranel, D. (2012). *Neuropsychological Assessment* (5th ed.). New York, NY: Oxford University Press.
- Lorch. (1995). Acquired disorders of writing and spelling. In H. Kirschner (Ed.), *Handbook of neurological speech and language disorders*. New York: Marcel Dekker.

- Lorch, Ferner, R., Golding, J., & Whurr, R. (1999). The nature of speech and language impairment in adults with neurofibromatosis 1. *Journal of Neurolinguistics*, 12, 157-165.
- Maeland, A. F. (1992). Handwriting and perceptual-motor skills in clumsy, dysgraphic, and 'normal' children. *Perceptual and Motor Skills*, 75(3 Pt 2), 1207-1217.
- Mazzocco, M. M., Turner, J. E., & Denckla, M. B., et al. (1995). Language and reading deficits associated with neurofibromatosis type 1: evidence for a notso-nonverbal learning disability. *Developmental Neuropsychology*, 11, 503-522.
- McCloskey, G., Perkins, L. & Van Divner, B. (2008). *Assessment and intervention for executive function difficulties*. Philadelphia: Routledge Taylor and Francis group.
- McHale, K., & Cermak, S. A. (1992). Fine motor activities in elementary school: preliminary findings and provisional implications for children with fine motor problems. *The American Journal of Occupational Therapy*, 46(10), 898-903.
- North, K. (2000). Neurofibromatosis type 1. *American Journal of Medical Genetics*, 97(2), 119-127.
- Pagani, L. S., Fitzpatrick, C., & Parent, S. (2012). Relating kindergarten attention to subsequent developmental pathways of classroom engagement in elementary school. *The Journal of Abnormal Child Psychology*, 40(5), 715-725.
- Parsons, T. D., Bowerly, T., Buckwalter, J. G., & Rizzo, A. A. (2007). (A controlled clinical comparison of attention performance in children with ADHD in a virtual reality classroom compared to standard neuropsychological methods. *Child Neuropsychology*, 13(4), 363-381.
- Payne, J. M., Hyman, S. L., Shores, E. A., & North, K. N. (2011). Assessment of executive function and attention in children with neurofibromatosis type 1: relationships between cognitive measures and real-world behavior. *Child Neuropsychology*, 17(4), 313-329.
- Penn, P. R., Rose, F. D., & Johnson, D. A. (2009). (Virtual enriched environments in paediatric neuropsychological rehabilitation following traumatic brain injury: Feasibility, benefits and challenges. *Developmental Neurorehabilitation*, 12(1), 32-43.
- Rizzo, A. A., Bowerly, T., Buckwalter, J. G., Klimchuk, D., Mitura, R., & Parsons, T. D. (2006). A virtual reality scenario for all seasons: the virtual classroom. *CNS spectrums*, 11(1), 35-44.
- Rosenblum, & Livneh-Zirinski, M. (2008). Handwriting process and product characteristics of children diagnosed with developmental coordination disorder. *Human Movement Science*, 27(2), 200-214.
- Rosenblum, Parush, S., & Weiss, P. L. (2003). Computerized temporal handwriting characteristics of proficient and non-proficient handwriters. *The American Journal of Occupational Therapy*, 57(2), 129-138.
- Rosenblum, Weiss, P. L., & Parush, S. (2004). Handwriting evaluation for developmental dysgraphia: Process versus product. *Reading and Writing*, 17(5), 433-458.
- Rowbotham, I., Pit-ten Cate, I. M., Sonuga-Barke, E. J., & Huijbregts, S. C. (2009). Cognitive control in adolescents with neurofibromatosis type 1. *Neuropsychology*, 23(1), 50-60.
- Spandel, V. (2004). *Creating young writers: Using the Six Traits to enrich writing process in primary classrooms* (2 ed.). Boston: Allyn & Bacon.
- St Clair-Thompson, H. L., & Gathercole, S. E. (2006). Executive functions and achievements in school: Shifting, updating, inhibition, and working memory. *Quarterly Journal of Experimental Psychology*, 59(4), 745-759.
- Stucki, G., Ewert, T., & Cieza, A. (2003). Value and application of the ICF in rehabilitation medicine. *Disability and Rehabilitation*, 25(11-12), 628-634.
- Verkerk, G. J., Wolf, M. J., Louwers, A. M., Meester-Delver, A., & Nollet, F. (2006). The reproducibility and validity of the Canadian Occupational Performance Measure in parents of children with disabilities. *Clinical Rehabilitation*, 20(11), 980-988.

- Weintraub, N., Drory-Asayag, A., Dekel, R., Jokobovits, H., & Parush, S. (2007). Developmental Trends in Handwriting Performance among Middle School Children. *OTJR: Occupation, Participation and Health*, 27(3), 104 -113.
- Werner, P., Rosenblum, S., Bar-On, G., Heinik, J., & Korczyn, A. (2006). Handwriting process variables discriminating mild Alzheimer's disease and mild cognitive impairment. *The Journals of Gerontology. Series B, Psychological Sciences and Social Sciences*, 61(4), P228-236.
- WHO. (2001). *International Classification of Functioning, Disability and Health*. Geneva: World Health Organisation.
- WHO. (2007). *International classification of functioning, disability and health-child and youth version*. Geneva: World Health Organization.
- Willaert, W. I., Aggarwal, R., Van Herzeele, I., Cheshire, N. J., & Vermassen, F. E. (2012). Recent advancements in medical simulation: patient-specific virtual reality simulation. *World Journal of Surgery*, 36(7), 1703-1712.
- Wood, R. L., & Liossi, C. (2006). The ecological validity of executive tests in a severely brain injured sample. *Archives of Clinical Neuropsychology*, 21(5), 429-437.
- Ylvisaker, M., & Feeney, T. (2002). Executive functions, self-regulation, and learned optimism in paediatric rehabilitation: a review and implications for intervention. *Pediatric Rehabilitation*, 5(2), 51-70.

INDEX

β

β-site APP cleaving enzyme ½ (BACE1/2), 1292
β-thalassemia, 411

#

11 beta-hydroxysteroid dehydrogenase 2 (HSD11B2), 1333, 1350, 1421, 1429, 1432
11 β-hydroxysteroid dehydrogenase 2 (11βHSD2), 1324, 1333
1-aminocyclopropane carboxylic acid, 1028
1-cyclohexene-1-carboxylic acid moieties, 1028
3-amino-5-hydroxy-benzoic acid (AHBA), 1018, 1019, 1020, 1036, 1040
3-hydroxy-3-methylglutaryl coenzyme A reductase (HMGCR), 1212, 1217, 1222
5-aza-2-deoxycytidine (DAC), 1324, 1332, 1343, 1344
5-hydroxymethylcytosine (5-hmC), 526, 527, 1226, 1228
5-methylcytosine (5-mC), 1226, 1227
8-hydroxydeoxy guanosine (8-OHdG), 1226
8-hydroxydeoxyguanosine, xxxvii, 1225, 1233, 1292, 1296

A

A complex, xxix, liii, 27, 85, 103, 137, 172, 195, 252, 283, 347, 348, 363, 569, 574, 691, 693, 723, 725, 777, 933, 1021, 1042, 1049, 1062, 1064, 1065, 1067, 1075, 1095, 1181, 1182, 1185, 1186, 1187, 1206, 1220, 1254, 1319, 1327, 1343, 1345, 1423, 1534, 1559, 1580, 1581, 1585, 1602, 1607, 1608, 1667, 1683, 1788, 1793, 1834, 1896, 2045, 2048, 2083, 2174, 2213, 2229, 2235, 2236, 2292, 2312

aansatrienins, 1026
ABCA7 (ATP-binding cassette subfamily A, member 7), 1291, 1299, 1313
absent, small or homeotic discs-like 2 (ASH2L), 1242, 1270
accession numbers, 62, 352, 480, 1124, 1125
acetyl co-enzyme A (CoA), 423, 442, 504, 658, 659, 662, 667, 674, 675, 676, 682, 685, 688, 1036, 1226, 1229, 1264, 1613, 2170, 2198, 2209, 2212, 2292, 2413
acetylated histone H3 at lysine 9 (H3K9ac), 726, 728, 1242, 1275
acquired antiestrogen resistance, 1643
acquired estrogen hypersensitivity, 1640
actinomycetes, xxxiii, 1017, 1018, 1020, 1620
activating mutations, 1261, 1634, 1635, 1637, 1638, 1642, 1643, 1644, 2197, 2200, 2202, 2204, 2335, 2336
activating transcription factor 2 (ATF2), 2268
acute inflammation, 1385, 1402
acute lymphocytic leukemia (ALL), 389, 640, 1181, 1242, 1248, 1251, 1253, 1255, 1260, 1270, 2195, 2196, 2200, 2203, 2208, 2221
acute myeloid leukemia (AML), 387, 640, 1181, 1184, 1185, 1188, 1191, 1192, 1242, 1247, 1248, 1250, 1253, 1255, 1261, 1266, 1269, 1280, 1281, 1288, 1289, 1290, 2196, 2200
acyltransferase, 658, 659, 674, 684, 686, 688, 1030
adenomatous polyposis coli (APC), 1242, 1246, 1260, 1261, 1287, 2200, 2201
adipogenesis, xxxix, 1419, 1422, 1425, 1427, 1428, 1429, 1430, 1431, 1432, 1433, 1436, 1438, 1439, 1440, 1445
adiponectin (ADIPOQ), 1012, 1217, 1339, 1420, 1432, 1438, 1443
advanced glycation endproducts (AGE), 1292, 1294
against citrullinated peptide antigens (ACPA), 1384, 1393, 1408
AHBA synthases, 1036

- alanine stereochemistry, 1026
 alpha-synuclein, 351, 354, 358, 359, 364, 365, 368, 371
 amplicons, lv, 1136, 2080, 2086, 2087, 2089, 2092, 2093, 2094, 2095, 2096, 2098, 2100, 2101, 2102, 2103, 2104, 2107, 2108, 2110
 amyloid precursor protein (APP), xxii, 427, 428, 430, 431, 432, 434, 435, 437, 439, 447, 448, 449, 450, 1292, 1293, 1294, 1295, 1296, 1297, 1298, 1299, 1301, 1302, 1303, 1304, 1306, 1307, 1308, 1309, 1310, 1312, 1313, 1316, 1317, 1320, 2098, 2110
 amyloid β -peptid (A β), xxii, 427, 428, 430, 431, 432, 433, 434, 435, 436, 437, 440, 442, 448, 1239, 1292, 1293, 1294, 1296, 1297, 1298, 1299, 1300, 1301, 1302, 1303, 1304, 1305, 1306, 1307, 1308, 2098, 2099
 androgen receptor (AR), 29, 33, 79, 106, 111, 113, 135, 356, 369, 370, 398, 401, 402, 403, 421, 424, 578, 730, 774, 871, 876, 878, 1009, 1055, 1059, 1105, 1121, 1238, 1242, 1246, 1251, 1255, 1261, 1267, 1268, 1276, 1277, 1280, 1282, 1284, 1285, 1289, 1313, 1362, 1368, 1377, 1487, 1629, 1645, 1648, 1710, 1712, 1756, 1794, 1871, 2032, 2160, 2184, 2189, 2245, 2368
 angiotensin like protein-2 (AMOTL2), 1324, 1332
 angiotensin type 1b receptor (AT1bR), 1324, 1340
 angiotensin-converting enzyme (ACE), 526, 994, 995, 996, 997, 1006, 1008, 1010, 1011, 1015, 1291, 1295, 1351, 1361, 1790, 1792, 1796
 ansamycin(s), xxxiii, 1017, 1018, 1019, 1020, 1023, 1025, 1026, 1028, 1030, 1036, 1037, 1038, 1039, 1040, 1041, 1042, 1043
 antibiotic therapy, 267
 antibodies, 306, 368, 438, 512, 571, 574, 576, 605, 696, 1019, 1020, 1186, 1299, 1309, 1395, 1532, 1764, 1765, 1768, 1769, 1771, 1776, 1886, 2094, 2097, 2098, 2104, 2153, 2155, 2166, 2170, 2175, 2176, 2178, 2179, 2189, 2212, 2215, 2258
 antibody-maytansinoid-conjugates, 1020
 anticancer defense mechanisms, 1634
 antiestrogen resistance, 1634, 1644
 antiestrogen treatment, xlv, 1051, 1058, 1633, 1637, 1643
 antisense non-coding RNA in the *INK4* locus (ANRIL), 1324, 1337, 1338, 1353, 1355
 antitumor necrosis factor- α (anti-TNF- α), 1292, 1299
 anuran cytogenetic, 1117, 1123
 anxiety disorders, xxvi, li, 649, 650, 1891, 1895, 1896, 1899, 1900, 1901, 1902, 1906, 1908, 1913, 1925, 1938
 aortic valve sparing, xxxiv, 1099, 1100, 1103, 1104, 1105
 apoptotic tumor cell death, 1634
 APP intracellular domain (AICD), 439, 443, 1292, 1294
 area under the ROC curve (AUC), 1212, 1216
 ARM repeat fold (ARM), 466
 arrayCGH, 1181, 1183, 1185, 1187
 artificial estrogens, 1634
 Asian population, 1593, 1938
 assisted reproduction, 882, 1372, 1374, 1377, 1378
 assisted reproductive technology (ART), 882, 1362, 1370, 1371, 1372, 1373, 1374, 1377, 1380, 2199
 ataxia, l, li, lii, 349, 414, 415, 416, 424, 1178, 1652, 1653, 1654, 1666, 1859, 1863, 1866, 1867, 1871, 1881, 1883, 1884, 1885, 1886, 1889, 1890, 1891, 1892, 1893, 1894, 1896, 1900, 1901, 1904, 1908, 1911, 1912, 1921, 1922, 1923, 1924, 1925, 1926, 1927, 1928, 1931, 1935, 1936, 1938, 1939, 1941, 1945, 1948, 1949, 1950, 1952, 2061, 2062, 2063, 2064, 2067, 2074, 2075, 2096, 2097, 2109, 2195, 2268
 ataxia telangiectasia-mutated (ATM), 414, 416, 421, 422, 1178, 1185, 1186, 1191, 2097, 2109, 2268, 2270, 2275
 ATP13A2, 350, 352, 356, 358, 370
 arial natriuretic factor (ANF), 1324, 1328, 1332
 attention deficit hyperactivity disorder (ADHD), xliii, li, 1589, 1590, 1591, 1593, 1594, 1595, 1596, 1597, 1598, 1685, 1690, 1698, 1700, 1703, 1709, 1710, 1866, 1868, 1895, 1898, 1903, 1906, 1913, 1925, 1942, 2073, 2392, 2393, 2397, 2400, 2401, 2402, 2406, 2408, 2409, 2412, 2414, 2415, 2419, 2445, 2446, 2447, 2448, 2456, 2458
 atypical parkinsonism, 356
 autism spectrum disorder, xxvi, xlix, li, 418, 649, 650, 651, 652, 654, 655, 656, 1651, 1709, 1713, 1756, 1853, 1859, 1860, 1862, 1869, 1895, 1897, 1903, 1904, 1905, 1906, 1913, 1925, 1931, 1938, 1942, 1945, 1951, 2076, 2397
 autistic traits, xlv, 652, 1062, 1698, 1700, 1705, 1706, 1709, 1710, 1920, 2402
 autophagy-12 (ATG12), 1226, 1231
 autophagy-8 (ATG8), 1226, 1231
 autosomal recessive pattern, 355
 autosomal-dominant inheritance, 350
 Avy (agouti), 906, 1420, 1426, 1431, 1437, 1445
 Azoospermia a/b/c (AZF), 1362, 1378

B

- B complex, 347, 348
 bacterial infections, 1033, 1399, 1400, 1415
 B-cell ALL, 1250
 Bcl-x, 410

- biochemistry, xxxiii, 183, 200, 370, 371, 447, 512, 514, 529, 530, 547, 682, 686, 687, 689, 735, 969, 1017, 1222, 1225, 1269, 1274, 1290, 1291, 1310, 1375, 1379, 1476, 1477, 1484, 1485, 1486, 1487, 1838, 1840, 1883, 1927, 2131, 2145, 2345
 biodiversity, xv, xxx, xxxi, 1, 11, 30, 46, 52, 55, 72, 81, 84, 104, 113, 224, 226, 826, 827, 828, 829, 831, 837, 845, 846, 847, 881, 882, 891, 892, 893, 897, 898, 902, 904, 910, 949, 950, 951, 1194, 1477, 1522
 blood-brain barrier (BBB), 536, 1031, 1292, 1294, 1309, 2258
 body mass index (BMI), xxvi, 307, 649, 652, 1343, 1420, 1421, 1427, 1433, 1438, 1744, 1761, 1769
 Brahma (BRM), 518, 521, 1324, 1327
 Brahma-related gene-1 (*BRG1*), 1324, 1327
 brain derived neurotrophic factor (BDNF), xxiii, xliii, 428, 438, 446, 449, 1589, 1590, 1591, 1592, 1593, 1594, 1596, 1597, 1598, 1599, 1726, 1752, 2103, 2113
 brain natriuretic peptide (BNP), 1324, 1328, 1332, 1344
 Braveheart (Bvht), 1324, 1330, 1338, 1348
 BRCA1 mutation carriers, 1050, 1051, 1057, 1638
 BRCA-gene mutation, 1636
 breast cancer 1, early onset (*BRCA1*), xxxiii, xli, xliv, xlix, 386, 399, 413, 1045, 1046, 1047, 1048, 1049, 1050, 1051, 1052, 1055, 1056, 1057, 1058, 1239, 1242, 1245, 1246, 1251, 1265, 1276, 1526, 1535, 1536, 1537, 1538, 1540, 1541, 1545, 1546, 1633, 1636, 1637, 1638, 1639, 1641, 1644, 1645, 1646, 1849, 1851, 1853, 1856, 2024, 2035, 2049, 2060, 2073, 2168, 2172, 2174, 2234, 2235, 2251, 2267, 2268, 2269, 2270, 2271, 2275, 2276
 breathing phenotype, 1067
 breeding, xxx, xxxi, 71, 72, 87, 139, 159, 270, 304, 673, 676, 686, 781, 825, 827, 828, 829, 831, 832, 833, 835, 836, 837, 838, 839, 840, 841, 843, 847, 857, 858, 859, 860, 862, 863, 864, 865, 868, 869, 877, 960, 961, 963, 964, 976, 977, 979, 988, 989, 1448, 1476, 1482, 1793
 BRG1-associated factor (BAF), 1324, 1325, 1327
 bridging integrator 1 (*BIN1*), 393, 1292, 1299
 Bromodomain testis-specific protein (*BRDT*), 391, 1362, 1366, 1377
 1292, 1305, 1310, 1317, 1318, 1341, 1354, 1356, 1377, 1379, 1631, 1845, 1846, 1847, 1857, 1923, 2188, 2219, 2221, 2245, 2419
 cancer prevention, 1046, 1055, 1546, 1634, 2187, 2270, 2274
 cancer stem cell (CSC), xxxviii, 1241, 1242, 1245, 1248, 1261, 1275, 1287, 1288, 2262, 2274
 cancer therapy, xxxviii, 1039, 1236, 1241, 1245, 1252, 1262, 1264, 1265, 1283, 1634, 2164, 2181, 2182, 2190, 2230, 2240, 2241, 2244, 2270, 2271, 2275
 carbon flux, xxvii, 657, 658, 660, 670
 carboxy-terminal domain (CTD), 413, 463, 466, 467, 517, 518, 521
 cardiac hypertrophy (CH), xxxviii, 106, 220, 254, 371, 372, 403, 421, 422, 425, 1014, 1237, 1270, 1271, 1279, 1281, 1288, 1311, 1323, 1324, 1335, 1344, 1352, 1359, 1763, 1764, 1767, 1825, 1892, 2100, 2124, 2188, 2220, 2244, 2372
 cardiovascular disease(s) (CVD), xxxiv, xxxviii, 303, 305, 599, 699, 702, 1093, 1094, 1095, 1096, 1098, 1099, 1108, 1111, 1212, 1214, 1215, 1220, 1233, 1323, 1324, 1330, 1331, 1336, 1338, 1339, 1340, 1341, 1342, 1343, 1344, 1349, 1351, 1353, 1354, 1355, 1356, 1358, 1384, 1390, 1435, 1444, 1742, 1845, 2035, 2166
 cascade, xli, 535, 570, 674, 962, 1095, 1231, 1259, 1281, 1294, 1462, 1526, 1535, 1537, 1539, 1545, 1546, 1789, 1813, 1816, 1864, 1867, 1868, 1869, 1870, 1918, 1943, 1944, 2099, 2194, 2197, 2231, 2256, 2330, 2340, 2407
 cascade screening, xli, 1526, 1537, 1868, 1869
 caspase, 387, 392, 399, 401, 571, 577, 591, 1028, 1247, 1324, 1338, 1420, 1431, 2092, 2103, 2110, 2136, 2256
 caspase 8 (*CASP8*), 1247, 1420, 1431, 1442, 2136
 caspase recruitment domain 8 (*CARD8*), 1324, 1338, 1355
 catalase (*CAT*), 661, 704, 763, 767, 768, 775, 851, 914, 1226, 1233, 1292, 1297, 1593, 2102
 caveolin-1 (*CAV-1*), 1420, 1432
 CCAAT/enhancer binding protein (*C/EBP*), 1420
 CD2-associated protein (*CD2AP*), 1292, 1299, 1313
 cell signaling, xviii, 189, 192, 196, 197, 534, 570, 574, 1246, 1279, 1296, 1400, 1442, 2086, 2205, 2267
 central nervous system (CNS), lx, 346, 360, 370, 376, 442, 451, 586, 737, 1046, 1226, 1230, 1298, 1397, 1435, 1708, 1728, 1742, 1811, 1823, 1886, 1892, 1906, 1913, 1924, 1943, 1947, 1950, 2094, 2098, 2108, 2112, 2329, 2330, 2331, 2336, 2345, 2346, 2347, 2361, 2366, 2387, 2402, 2418, 2434, 2458

C

- C complex, 347, 348, 358
 C-15003, 1020
 cadherins, 1247, 2259
 calorie restriction (CR), 80, 220, 372, 447, 578, 732, 733, 734, 765, 773, 953, 1012, 1042, 1056, 1219,

- Cercospora kikuchii*, xxxi, 871, 872, 875, 878, 879, 880
- CGG expansion, lii, 1671, 1860, 1863, 1880, 1916, 1917, 1927, 1931, 1937
- chromatin remodeling activities, 1635
- chromodomain helicase DNA-binding 7 (CHD7), 1324, 1328, 1346
- chromosomal aberrations, xxxvi, 229, 1157, 1158, 1159, 1160, 1171, 1172, 1173, 1174, 1177, 1178, 1179, 1180, 1181, 1184, 1185, 1186, 1187, 1188, 1189, 1206, 1939, 2123, 2270, 2287, 2310
- chronic inflammation, xxxv, xxxix, 574, 639, 713, 1107, 1108, 1244, 1245, 1252, 1257, 1259, 1265, 1331, 1383, 1385, 1386, 1391, 1393, 1397, 1399, 1402, 1403, 1405, 1407, 1416, 1417, 2205
- chronic lymphocytic leukemia, 387, 389, 399, 400, 1181, 1185, 1189, 1190, 1191, 1192, 1240, 1286, 1288, 2204, 2216, 2217, 2226, 2235, 2238, 2242, 2251, 2253, 2289, 2298
- chronic myelogenous leukemia (CML), 1181, 1242, 1249, 1250, 1528, 1542, 2200, 2208, 2213, 2215, 2219, 2340
- cis*-acting mutations, 349
- clusters, 33, 96, 98, 154, 155, 228, 458, 525, 717, 725, 729, 919, 920, 1030, 1067, 1126, 1129, 1131, 1198, 1214, 1245, 1454, 1455, 1465, 1499, 1502, 1505, 1506, 1522, 2139, 2215, 2233, 2249, 2273
- CML leukemogenic, 1250
- coactivator associated arginine methyltransferase 1 (CARM1), 1242, 1284
- coagulation factor II (thrombin) receptor-like 3 (F2RL3), 1324, 1342
- cognitive behavioral phenotype, 1700, 1709, 1710
- cognitive empowerment, xlv, 1715, 1716, 1717, 1718, 1719, 1723, 1725
- cognitive planning, xlvii, 1741, 1742, 1743, 1744, 1746, 1748, 1750, 1752, 1756
- collagen type XV A1 (COL15A1), 1324, 1344, 1358
- colorectal cancer, xli, lviii, 367, 389, 1037, 1167, 1168, 1174, 1178, 1244, 1249, 1252, 1253, 1254, 1267, 1271, 1276, 1278, 1280, 1286, 1391, 1402, 1407, 1416, 1417, 1430, 1432, 1441, 1444, 1525, 1526, 1529, 1531, 1532, 1533, 1535, 1538, 1539, 1540, 1541, 1543, 1546, 1823, 2165, 2197, 2205, 2217, 2219, 2227, 2234, 2239, 2246, 2250, 2253
- competition, xix, 45, 46, 47, 48, 163, 168, 171, 184, 187, 252, 257, 258, 272, 284, 285, 286, 288, 291, 304, 585, 591, 621, 624, 733, 750, 751, 755, 772, 788, 793, 795, 798, 806, 808, 810, 1512, 1619, 1717, 2109, 2254, 2298
- complement receptor 1 (CR-1), 1292, 1299
- complex karyotype, xxxvi, 1177, 1178, 1179, 1180, 1182, 1183, 1184, 1185, 1186, 1187, 1188, 1189, 1190, 1191, 1192
- condensation, 163, 261, 262, 263, 266, 267, 270, 271, 272, 273, 274, 293, 294, 296, 298, 546, 1028, 1366, 1404
- conservation genetics, 910, 957
- core collections, xxx, 825, 834, 842
- coregulators, 1636
- Cornelia de Lange syndrome, xv, xlv, 1206, 1209, 1683, 1692, 1693, 1694, 1695
- coronary heart disease, xxxviii, 712, 1008, 1323, 1324, 1340, 1341
- coronary heart disease (CHD), xxxviii, 521, 712, 1008, 1323, 1324, 1331, 1340, 1341, 1343
- cost-effectiveness, 1526, 1527, 1528, 1529, 1530, 1531, 1532, 1533, 1534, 1536, 1537, 1539, 1540, 1541, 1543, 1544, 1546, 1940
- CREB binding protein (CRB), 1257, 1275, 1324, 1329, 1335
- Crohn's disease (CD), 28, 33, 54, 57, 78, 115, 220, 446, 577, 578, 730, 733, 1056, 1057, 1106, 1170, 1269, 1271, 1279, 1281, 1311, 1314, 1346, 1384, 1390, 1391, 1392, 1938, 2220, 2225, 2226, 2242, 2247, 2252
- Cucumis melo*, 826, 828, 829, 833, 836, 837, 838, 839, 840, 841, 842, 843
- cultural evolution, 258, 260, 272, 282, 284, 288, 289, 290, 291, 295, 298
- cyclic adenosine monophosphate (cAMP), 534, 536, 541, 542, 548, 1242, 1250, 1277, 1303, 1416, 1792, 1794, 1795, 1796, 1804, 2095, 2289, 2407, 2408, 2418
- cyclic guanosine monophosphate (cGMP), xxiv, 499, 533, 534, 535, 536, 537, 539, 540, 541, 542, 543, 544, 545, 547, 548, 1336, 1352
- cyclin dependent kinase 2 (CDKN2), 1242, 1245, 1247, 1265
- cyclin dependent kinase 9 (Cdk9), 1242
- cyclin-dependent kinase inhibitor 1A (CDKN1A), xxxix, 428, 442, 1256, 1419, 1420, 1430, 1433, 2234, 2249
- cyclooxygenase (COX), 395, 703, 710, 1242, 1292, 1299, 1303, 1312, 1385, 1404, 2206, 2222, 2260, 2265
- cyclooxygenase-1 (COX-1/2), 1292, 1299, 1312, 2222
- CYP19 aromatase genes, xlv, 1633, 1637, 1644
- cystic fibrosis transmembrane conductance regulator (CFTR), 422, 1362, 1368, 1792, 1793, 1794, 2024, 2025, 2073
- cytochrome c oxidase subunit VIIa polypeptide 1 (COX7A1), xxxix, 1419, 1420, 1431, 1442

cytochrome P450A1 (CYP1A1), 1324, 1342
 cytosine- phosphate-guanine sites (CpG sites), xxxix, 1226, 1227, 1289, 1333, 1383
 cytosine-guanine (CG), 78, 109, 125, 366, 451, 790, 791, 1013, 1014, 1275, 1276, 1285, 1310, 1312, 1314, 1320, 1384, 1408, 1410, 1420, 1423, 1580, 1630, 1921, 1938, 2128, 2226, 2241, 2242, 2251, 2370, 2371, 2372
 cytotoxic, xxxiii, 315, 405, 534, 552, 571, 1017, 1019, 1029, 1037, 1043, 1163, 1242, 1414, 1587, 1772, 2091, 2099, 2107, 2153, 2166, 2170, 2179, 2211, 2212, 2244
 cytotoxic T-Lymphocytes (CTL), 1242, 2166
 cytotoxicity, xxiv, 397, 533, 590, 696, 1019, 1020, 1025, 1037, 1399, 2088, 2275
 cytotrienin, 1020, 1028, 1042
 cytotrienin A contain a trieneansamycin, 1028

D

DAPK, 1246, 1391, 1401
 Darwin, Charles, 38, 259, 285
 de novo antiestrogen resistance, 1643
 death associated protein kinase 1 (DAPK1), 1242, 1245
 death receptor (DR), 220, 350, 370, 381, 512, 1059, 1075, 1105, 1235, 1242, 1275, 1276, 1318, 1345, 1354, 1393, 1406, 1409, 1411, 1417, 1624, 1632, 1648, 1692, 1694, 1827, 1846, 1891, 1908, 1909, 1925, 1938, 2204, 2220, 2223, 2250, 2251, 2253, 2254, 2276, 2370, 2412
 deep brain stimulation, 361, 372
 deleted in Azoospermia (DAZ), 1362, 1371, 1379
 demethylases (HDMs), 522, 523, 1248, 1272, 1384, 1387
 Dennett, Daniel, 285, 291, 292, 295
 developmental origin of health and diseases (DOHaD), xxxvii, 1211, 1212, 1213
 diacylglycerol acyltransferase, xxvii, 657, 658, 659, 674, 676, 682, 683, 684, 685, 687, 688
 differential diagnosis of TRPS, 1203
 differentially methylated region (DMR), xxviii, 715, 716, 725, 726, 727, 728, 736, 1212, 1215, 1362, 1370, 1372, 1427
 diffuse large B-cell lymphoma (DLBCL), 1242, 1251, 1259
 DiGeorge Syndrome Critical Region 8 (DGCR8), 1226, 1232, 1324, 1326, 1888, 1893, 1918, 1924, 2230, 2232, 2236, 2252
 dimethylation of histone H3 at lysine 9 (H3K9me2), 523, 1374, 1398, 1420, 1425
 directional selection, 130, 283, 285
 DNA defects, xlv, 1633, 1637

DNA double strand breaks (DSBs), 1160, 1161, 1162, 1164, 1178, 2268, 2270
 DNA methylation and inflammation, 1386
 DNA methylation valleys (DMV), 1384, 1403
 DNA methyltransferase (DNMT), xxxviii, xxxix, 530, 1213, 1226, 1228, 1236, 1237, 1241, 1242, 1245, 1246, 1262, 1263, 1264, 1266, 1267, 1272, 1289, 1292, 1301, 1303, 1315, 1324, 1326, 1340, 1343, 1347, 1355, 1358, 1362, 1383, 1384, 1386, 1398, 1399, 1404, 1412, 1420, 1423, 1428, 1436, 1439, 1444
 DNA methyltransferase inhibitors, 1262, 1263, 1264
 DNA repair, xxxiii, lviii, 412, 435, 526, 593, 756, 1045, 1046, 1047, 1055, 1160, 1175, 1178, 1179, 1230, 1245, 1249, 1250, 1289, 1290, 1340, 1403, 1634, 1643, 1645, 1646, 1687, 1886, 1892, 2172, 2178, 2193, 2205, 2206, 2262, 2267, 2268, 2269, 2270, 2271, 2273, 2274, 2287
 DNA safeguarding, 1636, 1638
 DOT1-Like Histone H3K79 Methyltransferase (DOT1L), 532, 1324, 1332, 1336
 double stranded RNA-binding domain (dsRBD), 466, 1232, 2236
 DSM-IV Diagnosis, 1703, 1709, 1710
 dwarfism, xlv, 1683, 1691

E

E complex, 347, 348
 E-cadherin (CDH1), 404, 1242, 1246, 1260, 1267, 1287, 1391, 1399, 1401, 1403, 1417, 1584, 2210, 2259, 2260, 2261, 2265
 ecological genetics, 53, 132, 133, 159, 328, 950, 1193
Eira barbara, xl, 1491, 1492, 1494, 1495, 1499, 1502, 1505, 1506, 1507, 1508, 1509, 1510, 1514, 1518, 1519, 1521
 ejaculate protection, xxix, 777, 778, 779, 786, 787, 789
 embryonic stem cell (ESC), 1324, 1328, 1362, 1373
 endothelial Nitric Oxide Synthase (eNOS), 534, 535, 536, 538, 543, 546, 995, 996, 1012, 1246, 1334, 1349, 1351, 1788, 1816, 1825
 endothelin-converting enzyme 1 (ECE-1), 1292
 engystomops, xxxv, 1117, 1119, 1122, 1125, 1127, 1129, 1130, 1131, 1134, 1150, 1153
 enhancer of zeste 2 polycomb repressive complex 2 subunit (EZH2), 1242, 1248, 1259, 1261, 1270, 1286, 1287, 1288
 enzyme AHBA, 1018
 EPH receptor A1 (EPHA1), 1292, 1299, 1313
 epidermal growth factor receptor (EGFR), lvii, lviii, lix, 492, 598, 640, 1254, 1262, 1288, 1324, 1336,

1352, 1404, 1530, 1531, 1532, 1533, 1534, 1541, 1543, 1793, 1794, 2092, 2164, 2166, 2167, 2170, 2173, 2174, 2176, 2177, 2181, 2199, 2200, 2205, 2208, 2220, 2239, 2255, 2256, 2257, 2258, 2259, 2260, 2261, 2262, 2263, 2264, 2265, 2335, 2340, 2347

epigenetic dysregulation
Male Infertility, 1368

epigenetic make-up of induced pluripotent stem cells/PGC from adult cells, 1373

epigenetic modifications, xxxvii, xxxviii, xxxix, 705, 710, 1211, 1213, 1214, 1215, 1216, 1219, 1225, 1227, 1236, 1241, 1245, 1291, 1339, 1341, 1343, 1345, 1355, 1361, 1370, 1374, 1376, 1383, 1386, 1405, 1422, 1425, 1426, 1427, 1428, 1439, 1443, 1444, 2273

epigenetic reprogramming of male germ cells, 1363

epigenome, xxxvii, xxxviii, 100, 114, 1211, 1219, 1236, 1244, 1245, 1264, 1266, 1288, 1300, 1343, 1356, 1361, 1362, 1374, 1402, 1406, 1415, 1416, 1425, 1428, 1433, 1434, 1445, 2188

epilepsy, 446, 707, 1063, 1077, 1209, 1651, 1652, 1668, 1700, 1702, 1703, 1707, 1712, 1726, 1862, 1944, 2066, 2076, 2098, 2286

epiobesigenic genes, 1419, 1429

epithelial sodium channel (ENaC), 1324, 1336

epithelial-to-mesenchymal transitions (EMT), 1242, 2240, 2250, 2259, 2260, 2261, 2263

Epstein Barr virus (EBV), 1384, 1400

ESR1, xxxix, xlv, 700, 1048, 1331, 1419, 1420, 1430, 1433, 1633, 1635, 1637, 1638, 1640, 1641, 1642, 1643, 1644, 1645, 1647, 1648

ESR1 gene amplification, 1640, 1643

estrogen activated estrogen receptors, xlv, 1633

estrogen receptor isoforms, 1635

estrogen therapy, 437, 1332, 1634, 1827

estrogen withdrawal, 1634, 1644

ethnicity, 1526, 1850, 2061, 2280

European population, 856, 953, 1520, 1594, 1596

Evo-devo, 260, 261, 299

evolutionary developmental biology, 260, 297, 1200

evolutionary genetics, xix, 131, 132, 184, 205, 247, 957, 1193

exon arrays, 361, 362, 372, 476

exon skipping/inclusion, 348

exonic splicing enhancers, 346, 385, 516, 586

exonic splicing silencers, 346

experimental gene therapy, lv, 2080, 2084

Exportin 5 (Exp5), 1232, 2244, 2252

extracellular signal-regulated kinase (ERK), 200, 393, 395, 402, 1028, 1042, 1111, 1277, 1347, 1792, 1793, 1794, 1804, 2210, 2232, 2268, 2282, 2335, 2347, 2419

extracellular superoxide dismutase (SOD3), 1421, 1431, 1442

F

family-based association study, xliii, 1590, 1591, 1597, 1598, 1599

fat mass and obesity associated (FTO), 1420, 1432

fatty acid synthase (FASN), 1420, 1432, 2198, 2208, 2212

fatty acid-binding protein 4, adipocyte (FABP4), 1420, 1432

F-box and leucine-rich repeat protein 11 (FBXL11), 1242, 1258, 1272

F-box only protein 7, 359, 371

FBXO7, 350, 352, 356, 359, 371

fetal growth restriction (FGR), 1212, 1215, 1221, 1342

fibroblast growth factor-2 (FGF2), xxxix, 1419, 1420, 1430, 1441

fibromyalgia, 1911, 1913, 1914, 1921, 1922, 1931, 1938

first-generation (F1), 14, 17, 18, 99, 102, 164, 211, 212, 214, 215, 217, 218, 226, 250, 623, 719, 785, 803, 804, 805, 839, 1194, 1195, 1201, 1212, 1217, 1222, 1442

fluorescence *in situ* hybridization, 1180

FMR1 gene, xlix, l, li, 1679, 1859, 1860, 1862, 1863, 1864, 1866, 1867, 1868, 1869, 1870, 1871, 1873, 1874, 1875, 1877, 1879, 1880, 1881, 1884, 1887, 1889, 1892, 1894, 1895, 1896, 1897, 1900, 1901, 1902, 1908, 1911, 1912, 1926, 1928, 1932, 1936, 1937, 1939, 1940, 1941, 1942, 1945, 1951

FMR1 premutation, l, li, lii, 1871, 1873, 1877, 1880, 1881, 1883, 1884, 1886, 1887, 1889, 1890, 1892, 1893, 1896, 1900, 1901, 1902, 1903, 1906, 1907, 1908, 1911, 1912, 1913, 1921, 1922, 1923, 1925, 1926, 1927, 1929, 1931, 1936, 1938

food and drug administration (FDA), 417, 608, 629, 633, 640, 1228, 1242, 1253, 1262, 1263, 1264, 1290, 1530, 1531, 1533, 1819, 2042, 2053, 2054, 2091

forkhead box protein O (FOXO), 764, 769, 1292, 1305

Foxf1 adjacent non-coding developmental regulatory RNA (Fendrr), 1324, 1330, 1338, 1348

fragile X syndrome, xlv, xlix, lii, 1669, 1670, 1678, 1679, 1680, 1681, 1859, 1860, 1869, 1871, 1874, 1880, 1894, 1904, 1906, 1908, 1927, 1928, 1936, 1939, 1942

frontotemporal dementia with Parkinsonism linked to chromosome 17, 360

full mutation alleles, li, 1911, 1917, 1927

G

gametogenesis, 236, 716, 720, 721, 1228, 1328, 1363, 1365, 1368, 1370, 1373, 1376, 1425, 1428
 GATA binding protein 4 (Gata4), 1324, 1329
 G-banding, 1180, 1183
 Gcn5-related N-acetylases (GNAT), 1226, 1229
 Gcn5-related N-acetyltransferase (GNAT5), 1242, 1250
 gDMR, 721, 725, 726
 geldanamycin, 1019, 1020, 1021, 1022, 1023, 1036, 1037, 1040
 gene clusters, xxxiii, 96, 98, 228, 718, 723, 725, 1017, 1030, 1627, 2215, 2218
 gene MECP2, 1063, 1651, 1716
 gene stacking, 657, 676, 677, 678, 679, 686
 gene-expression array, 361
 genetic association, 117, 340, 371, 712, 1529, 1536, 1542, 1593
 genetic counseling, lii, 1208, 1535, 1537, 1545, 1581, 1860, 1866, 1867, 1868, 1869, 1870, 1903, 1927, 1928, 1929, 1930, 1931, 1933, 1934, 1935, 1940, 2067
 genetic diversity, xxx, xxxi, xl, 89, 130, 333, 336, 340, 344, 825, 833, 834, 836, 838, 839, 840, 841, 842, 843, 848, 854, 856, 859, 860, 861, 862, 863, 865, 866, 867, 868, 869, 870, 882, 883, 887, 902, 914, 936, 1447, 1457, 1458, 1459, 1460, 1464, 1465, 1473, 1474, 1477, 1479, 1491, 1492, 1493, 1496, 1498, 1499, 1511, 1521, 1616, 2118, 2125, 2128, 2133, 2194
 genetic mutation, xxxviii, 350, 1066, 1244, 1251, 1262, 1293, 1300, 1328, 1361, 1402, 1683, 1687, 1716, 1963, 1970, 2027, 2060, 2171, 2172, 2186
 genetic predisposition to disease, 690
 genetic syndrome, 651, 655, 942, 1094, 1548, 1567, 1763, 1780, 2330
 genetic testing, xxxv, xl, liii, liv, 1113, 1525, 1526, 1533, 1535, 1537, 1538, 1539, 1540, 1541, 1545, 1546, 1583, 1691, 1784, 1852, 1867, 1870, 1885, 1905, 1934, 1935, 1940, 1956, 1966, 1974, 1991, 1992, 1996, 1997, 1998, 2003, 2004, 2017, 2025, 2027, 2036, 2038, 2039, 2041, 2042, 2045, 2046, 2047, 2049, 2050, 2051, 2052, 2055, 2058, 2062, 2063, 2064, 2065, 2067, 2068, 2072, 2173, 2286
 genetically proficient cases, xliv, 1633, 1637
 genome protecting mutations, 1635
 genome stabilizer machinery., 1636
 genome-wide association study, 690, 699, 708, 712, 1313, 1342, 1408, 1544, 2002, 2021, 2028
 genomic imprinting, 715, 716, 721, 730, 733, 734, 1214, 1220, 1227, 1228, 1326, 1370, 1378, 1380
 genotype/phenotype of TRPS patients, 1203

genotype-phenotype correlations, 1066, 1090, 1651, 1693, 1694, 1822, 1844, 1864, 1880
 genotyping tests, 1591
 germplasm, xxx, xxxi, 673, 825, 826, 827, 828, 829, 833, 834, 835, 836, 837, 838, 839, 840, 841, 842, 856, 857, 862, 863, 864, 866, 867, 868, 869, 871, 874, 877, 878, 881, 882, 889, 890, 902, 907
 gestational diabetes mellitus (GDM), 1212, 1217
 glucocorticoid receptor (NR3C1), 571, 574, 577, 580, 1216, 1336, 1355, 1420, 1430, 1440, 1441, 1444
 glutathione peroxidases (GSH-Px), 1233
 glutathione S-transferase-pi (GSTP1), 1173, 1242, 1246, 1267
 glutathione S-Transferases (GST), 125, 914, 918, 1233, 1498, 1509
 G-protein- coupled receptor 51, 1388
 granulocyte macrophage colony stimulating factor (GM-CSF), 571, 572, 577, 579, 1242, 1257, 2261

H

H3K4me, 719, 1242, 1365
 half a tetratricopeptide repeat (HAT), 465, 466, 1229, 1285, 1305, 1325, 1334, 1335, 1344, 1362, 1366, 1394
 heart and neural crest derivatives expressed 1 (HAND1), 1324, 1328
 heart failure (HF), xxxviii, 365, 1010, 1175, 1281, 1310, 1323, 1325, 1332, 1333, 1334, 1335, 1337, 1338, 1344, 1346, 1348, 1582, 1584, 1812
 HeLa S3 cells, 1029
 hematologic malignancies, 1177, 1181
 hematopoietic stem cells (HSC), 1243, 1250, 1261
 hepatocellular carcinoma (HCC), 744, 1242, 1247, 1248, 1252, 1255, 1259, 1260, 1265, 1270, 1271, 1278, 1281, 1402, 1432, 1437, 1441, 1442, 2197, 2205, 2235, 2242, 2249, 2250, 2251
 hepatocyte nuclear factor 4 receptor α (HNF4 α), 1212, 1216
 herbimycin, 1020, 1021, 1023, 1024, 1025, 1026, 1041
 herbimycin A, 1024, 1025, 1026, 1041
 heterochrony, 260, 261, 262, 274, 299
 heterogeneity, xx, xlii, xlix, 88, 89, 142, 301, 306, 314, 325, 326, 342, 693, 697, 699, 701, 710, 912, 914, 917, 918, 1009, 1062, 1063, 1066, 1245, 1266, 1346, 1498, 1507, 1509, 1512, 1548, 1549, 1557, 1558, 1565, 1593, 1594, 1651, 1670, 1671, 1702, 1800, 1831, 1843, 1939, 2059, 2064, 2067, 2170, 2187, 2189, 2213, 2300, 2358, 2369
 hierarchical order, 259, 296

- HIF1 α binding to the HIF1 α -response element (HRE), 1242, 1248, 1816
- high mobility group (HMG), 1243, 1796, 2170, 2196, 2232, 2247, 2262, 2292, 2413
- histon modifications, 1419
- histone acetyl transferases (HAT), 465, 466, 1229, 1285, 1305, 1325, 1334, 1335, 1344, 1362, 1366, 1394
- histone acetylation (AcH3), 523, 587, 733, 1225, 1230, 1234, 1235, 1236, 1239, 1245, 1249, 1252, 1258, 1265, 1268, 1274, 1285, 1304, 1305, 1317, 1335, 1336, 1341, 1350, 1351, 1358, 1365, 1366, 1367, 1377, 1387, 1392, 1393, 1394, 1397, 1408, 1413, 1415, 1420, 1426, 1427, 1431, 1432, 1645, 2273
- histone Deacetylase Inhibitors (HDACi), 1230, 1237, 1392, 1405
- histone Deacetylases (HDAC), xxxix, 1229, 1230, 1265, 1277, 1278, 1279, 1285, 1290, 1304, 1305, 1317, 1325, 1334, 1335, 1343, 1344, 1345, 1352, 1362, 1366, 1383, 1394, 1400, 1405, 1406, 1418, 1420, 1423
- histone H3 lysine 9 (H3K9), 523, 588, 733, 1247, 1248, 1257, 1270, 1271, 1284, 1334, 1335, 1336, 1353, 1365, 1374, 1398, 1414, 1420, 1423, 1427, 1428
- histone lysine acetyltransferases (KAT), 1243, 1245, 1250, 1255, 1262, 1264
- histone lysine acetyltransferases inhibitor (KATi), 1243, 1264
- histone lysine deacetylase (KDAC), 1243, 1252, 1253, 1259, 1262, 1264
- histone lysine deacetylase inhibitor (KDACi), 1243
- histone lysine demethylase inhibitors (KDMi), 1263
- histone lysine demethylases (KDMs), 1243, 1249, 1261
- histone lysine methyl transferase (KMTi), 1263
- histone lysine methyltransferases (KMT), 1243, 1247, 1264, 1270, 1271
- histone methyl transferase (MLL), 1184, 1243, 1248, 1255, 1270, 1272, 1281, 1345, 2196, 2199, 2203, 2216
- histone methylase (HMT), 522, 1245, 1284, 1420, 1423
- histone methyltransferases (HMTs), 522, 523, 1230, 1326, 1384, 1387, 1423, 1428
- histone modifications and inflammation, 1387
- hnRNP, xxi, xxiii, 383, 385, 386, 387, 392, 393, 394, 395, 398, 400, 401, 403, 404, 453, 457, 458, 459, 464, 465, 468, 469, 471, 472, 516, 543, 584, 585, 586, 587, 589, 591, 1886, 1888, 1893, 1918
- homocysteine (Hcy), 1006, 1015, 1325, 1331, 1349, 1355, 1430, 1441, 1442
- HOTAIR, 1637, 1641, 1645, 1647
- Hsp90, 1020, 1021, 1022, 1023, 1040, 2258, 2264
- HSV-1-derived vectors, 2080, 2083, 2108
- human biomonitoring, xxxvi, 1157, 1158, 1159, 1160, 1162, 1165, 1166, 1167, 1168, 1169, 1170, 1171
- human epidermal growth factor receptor 2 (ERBB2), 1046, 1242, 1248
- human immunodeficiency virus (HIV), 512, 1013, 1030, 1034, 1042, 1043, 1230, 1238, 1255, 1384, 1400, 1415, 2002, 2093, 2102, 2108, 2180, 2212
- human intelligence, 288
- human language, xix, 257
- human leukocyte antigen F-associated (FAT10), 499, 511, 1231
- human species, xix, 257, 268, 287, 289, 290, 291, 292, 295, 297, 597
- hyperactive disorders, 1896
- hyperhomocysteinemia (HHcy), 1325, 1331, 1349
- hypoestrogenism, 1634
- hypoxia inducible factor 1 (HIF1A), 1420, 1432
- hypoxia-inducible factor (HIF1), 1242, 1248, 1252, 2211, 2233, 2249

- I κ B kinase (IKK), 1243, 1257, 1283, 1284, 1396, 1413, 2222
- impaired glucose tolerance (IGT), 1212, 1213
- imprinted gene, xxviii, 715, 716, 726, 729, 732, 733, 734, 735, 737, 1370, 1371, 1372, 1375, 1379, 1403, 1426, 1429, 1437, 1440, 2273, 2277
- imprinting center (IC), 717, 719, 720, 721, 722, 723, 726, 728, 730, 732, 733, 734, 770, 1356, 1474, 1590, 2369, 2370
- imprinting mark, 716, 720, 729
- in vitro fertilization (IVF), xxxv, xliii, 899, 1113, 1362, 1372, 1374, 1379, 1570, 1577, 1579, 1585, 1939, 2072, 2185
- induced pluripotent stem cells (iPS/iPSCs), 899, 1362
- inducible nitric oxide synthase (iNOS), 534, 1243
- inflammation and cancer, 1402
- inflammatory autoimmune diseases, 1389
- inflammatory bowel disease (IBD), 1244, 1384, 1390, 1391, 1392, 1401, 1402, 1405, 1407, 1408, 1416
- inflammatory disorders, 1389
- initiation factor eIF-4 gamma (MIF4G), 460, 466
- insertion of exogenous ERs, 1640

insulin (INS), 591, 757, 758, 1058, 1212, 1362, 1414, 1420, 1432, 1434, 1438, 1443, 1563, 1803, 2298

insulin like growth factor (IGF), 757, 758, 1215, 1243, 1648, 1759, 1793, 1794, 2257

insulin receptor (INSR), 737, 766, 1420, 1432, 1443

insulin resistance, xxxvii, xl, 1006, 1211, 1213, 1214, 1216, 1217, 1218, 1219, 1419, 1421, 1422, 1431, 1432, 1433, 1434, 1444, 1445, 1635, 1644, 1791, 1803

insulin-degrading enzyme (IDE), 1292, 1295

insulin-like growth factor-binding protein 3 (IGFBP3), xxxix, 1419, 1420, 1432, 1433, 1443

insulin-responsive glucose transporter 4 (GLUT4), 1420, 1432, 1438, 2210

intellectual disability, 1741, 1742, 1860, 2059

intentional selection, 293, 297

interferon γ (IFNG), 1420, 1431

interleukin 13 receptor alpha 2 (IL13ra2), 1212, 1217

interleukin-1 β lipoprotein receptor-related protein 1 (LRP1 IL-1 β), 1292

intermediate alles, 1927

International Diabetes Federation (IDF), 328, 1212, 1220

International Union of Pure and Applied Chemistry (IUPAC), 1233, 1238

intra-cytoplasmic sperm injection (ICSI), 882, 887, 891, 1362, 1370, 1372, 1374, 1380

intrauterine growth restriction (IUGR), 1212, 1215, 1325, 1333, 1440

intron retention, 348, 390, 397, 459, 476, 516

intronic splicing enhancers, 347, 454, 516, 584, 585

intronic splicing silencers, 347, 385, 585, 586

introns, xxi, 234, 315, 346, 349, 356, 374, 383, 384, 443, 454, 483, 508, 515, 517, 519, 526, 530, 556, 562, 586, 587, 592, 703, 723, 947, 968, 969, 1232, 1457, 1478, 2024

Iroquois homeobox 4 (Irx4), 1325, 1328, 1329, 1346

J

jumonji AT rich interactive domain 2 (Jarid2), 1325, 1328, 1346

jumonji-domain-containing protein histone lysine demethylases (JmjC/N), 1243

K

Krueppel-like factor 4 (KLF4), 1362, 1373

Kufor-Rakeb syndrome, 358

L

lactams, 1019, 1026, 1039

Lamarck, 259, 286

language production and comprehension, 1066

lariat branch point, 346, 347

latent autoimmune diabetes of adults, 301

latent early-life associated regulation (LEARn), 1292, 1308, 1321

learning abilities, 1548

Leptin(LEP), 1420, 1429, 1430, 1441

leucine-rich repeat kinase 2, 355

leukemia, 109, 389, 390, 393, 402, 572, 574, 576, 579, 581, 598, 640, 1020, 1028, 1178, 1181, 1184, 1185, 1188, 1189, 1190, 1192, 1242, 1248, 1250, 1255, 1258, 1259, 1260, 1261, 1265, 1267, 1269, 1270, 1272, 1276, 1280, 1281, 1285, 1286, 1287, 1288, 1445, 1542, 1758, 1779, 1851, 1852, 2034, 2047, 2195, 2196, 2197, 2198, 2199, 2200, 2203, 2205, 2213, 2214, 2215, 2216, 2217, 2220, 2225, 2226, 2234, 2238, 2239, 2250, 2251, 2260, 2276, 2285, 2340, 2354

Lewy bodies, 350, 351, 354, 356, 359, 364, 370

Lewy neurites, 350

ligand independent activation of ERs, 1635

linguistic phenotype, xxxiv, 1061, 1067, 1068, 2432

lipoprotein lipase (LPL), xxxix, 340, 1010, 1357, 1419, 1420, 1431, 1433

lncRNAs, 361, 528, 1330, 1337, 1388, 1407, 1635, 1637, 1638, 1641

long intergenic noncoding RNA predicting cardiac remodeling (LIPCAR), 1325, 1338

long interspersed nucleotide elements-1 (LINE-1), 238, 245, 1234, 1325, 1332

long non-coding RNA (lncRNA), 372, 1212, 1325, 1326, 1330, 1337, 1338, 1645, 1647

long-interspersed nuclear element-1 (LINE-1), 238, 245, 1234, 1325, 1332

low protein (LP), 142, 150, 151, 809, 1012, 1088, 1212, 1216, 1240, 1317, 1356, 1628, 1857, 2251, 2370

lung cancer, 388, 389, 398, 403, 598, 1168, 1174, 1247, 1251, 1252, 1254, 1255, 1262, 1265, 1268, 1269, 1270, 1273, 1275, 1278, 1279, 1280, 1281, 1287, 1290, 1432, 1441, 1443, 1526, 1529, 1530, 1531, 1533, 1535, 1536, 1537, 1540, 1541, 1542, 1543, 1544, 1545, 1850, 1851, 1852, 2197, 2198, 2204, 2216, 2217, 2221, 2226, 2231, 2234, 2235, 2238, 2239, 2242, 2243, 2244, 2246, 2247, 2249, 2251, 2253, 2257, 2264, 2336

lymphoma, 388, 393, 402, 1178, 1181, 1186, 1188, 1189, 1190, 1191, 1242, 1249, 1250, 1252, 1253, 1259, 1260, 1275, 1278, 1279, 1286, 1404, 1530,

1769, 1780, 1851, 2175, 2195, 2196, 2200, 2213, 2216, 2217, 2221, 2226, 2233, 2248, 2249, 2250, 2251, 2340
 lynch syndrome, xli, 1254, 1526, 1534, 1535, 1538, 1540, 1546
 lysine (K)-specific demethylase 1A (LSD-1), 1325, 1329
 lysine 27 of histone H3 (H3-K27), 1335, 1420, 1427
 lysine-specific histone demethylase 1A (KDM1A), 1248, 1262, 1272, 1290, 1325, 1329

M

male infertility, xxxviii, 1361, 1362, 1367, 1368, 1369, 1370, 1371, 1373, 1374, 1375, 1376, 1377, 1378, 1379
 malondialdehyde (MDA), 573, 1037, 1292, 1296, 1616
 mammalian target of rapamycin (mTOR), 399, 1023, 1642, 1793, 1794, 1795, 1804, 1805, 2198, 2207, 2208, 2209, 2210, 2211, 2223, 2224, 2234, 2250, 2256, 2257, 2264, 2268, 2289, 2290, 2292, 2298, 2335, 2340, 2346, 2347, 2407, 2418, 2419
 mantle cell lymphoma, 1186
 Marfan syndrom, xxvi, xxxiv, xxxv, 416, 424, 649, 650, 654, 655, 656, 1080, 1081, 1085, 1086, 1087, 1088, 1089, 1090, 1091, 1093, 1095, 1096, 1097, 1098, 1100, 1105, 1106, 1107, 1108, 1109, 1110, 1111, 1112, 1113, 1114, 2380
 Mate choice copying, xxix, 795, 796, 808, 810, 811
 mating effort, 767, 778, 789, 791, 792, 794
 McShea, Daniel, 259
 measurement, 275, 278, 286, 297, 447, 598, 605, 799, 801, 802, 915, 1050, 1121, 1195, 1709, 1730, 1769, 1846, 1906, 2034, 2175, 2303, 2312, 2314, 2364, 2440, 2456
 mediator complex subunit 13 (MED13), 1325, 1330, 1348
 melanocortin 4 receptor (MC4R), 1420, 1426, 1431
 melon, xxx, 825, 826, 827, 828, 829, 830, 831, 832, 833, 834, 835, 836, 837, 838, 839, 840, 841, 842, 843
 memes, 288, 291, 292, 296
 memetic selection, 291
 methyl-CpG binding domain protein 3 (MBD3), 1243
 methyl-CpG-binding domain protein 2 (MBD2), 1243
 methylene tetrahydrofolate reductase (MTHFR), 994, 995, 996, 1006, 1012, 1015, 1362, 1371, 1378, 1379
 Mexican population, xliii, 1590, 1593, 1594
 micro RNAs, 1231, 1389, 1394

micronucleus, xxxvi, 1157, 1158, 1159, 1161, 1162, 1169, 1171, 1172, 1174, 1175, 1176
 microRNA (miRNA), lviii, 725, 1368, 2227
 microtubule-associated protein tau, 360, 380, 446, 450, 1317, 2099
 mild cognitive impairment (MCI), 1292, 1301, 1306, 1724
 mitochondrial ND5 gene, 1492, 1494, 1495, 1502, 1505, 1506, 1510, 1511
 mitochondrial pyruvate dehydrogenase kinase, xxvii, 657, 660, 662, 684, 687
 mitogen activated kinase (MAPK), lx, 200, 402, 428, 442, 462, 1028, 1243, 1259, 1396, 1400, 1415, 1461, 1635, 1854, 2202, 2232, 2246, 2256, 2257, 2259, 2261, 2262, 2268, 2284, 2291, 2329, 2330, 2335, 2336, 2344, 2347, 2407
 mitogen-activated protein (MAP), xxi, 373, 439, 462, 1028, 1042, 1384, 1398, 1403, 1409, 1415, 1461, 1477, 1635, 1642, 1854, 2197, 2202, 2203, 2210, 2257, 2291, 2335
 molecular markers, xxx, xxxi, 84, 825, 831, 833, 834, 835, 837, 841, 848, 857, 859, 860, 861, 863, 877, 878, 946, 1197, 1201, 1457, 1458, 1464, 1480, 1594, 1991, 2357
 molecular techniques, xlix, 18, 1456, 1458, 1463, 1849, 1850, 2115, 2117, 2121, 2122, 2177, 2186
 monocytic leukaemia zinc finger protein (MOZ), 1229, 1243, 1250, 1255, 1275, 1281
 mono-di-tri-methylated histone H3 at lysine 4 (H3K4me1/2/3), 1242
 mono-di-tri-methylated histone H3 at lysine 9 (H3K9me1/2/3), 1242
 mood disorders, li, 1698, 1891, 1896, 1900, 1903, 1908, 1913, 1931
 morphological markers, 860
 mRNA toxicity, 1883, 1887
 mRNA transcripts, xxiii, 346, 351, 361, 453, 476, 543, 548, 570, 1639, 2216
 mutations, xxxiv, lviii, 357, 358, 362, 367, 369, 385, 387, 388, 389, 390, 391, 392, 399, 400, 411, 412, 419, 424, 456, 461, 512, 513, 759, 763, 951, 1032, 1045, 1047, 1061, 1064, 1077, 1078, 1088, 1090, 1093, 1114, 1178, 1274, 1299, 1651, 1652, 1687, 1693, 1694, 1784, 1788, 1813, 1823, 1834, 1843, 1845, 1857, 1932, 2022, 2025, 2063, 2066, 2068, 2075, 2076, 2174, 2178, 2183, 2193, 2197, 2219, 2220, 2232, 2258, 2267, 2268, 2271, 2285, 2418
 mutually exclusive exons, 348, 395, 476
 mycobacteria, 1032, 1034
 mycotrienin, 1020, 1026, 1027, 1041
 myelin basic protein (MBP), 1028, 1384, 1397, 1886, 1888

myelodysplastic syndrome (MDS), 337, 338, 387, 388, 390, 400, 1181, 1182, 1183, 1189, 1191, 1234, 1269

myocardial infarction (MI), xxxviii, 28, 29, 110, 765, 811, 1239, 1280, 1323, 1325, 1334, 1337, 1344, 1357, 1367, 1377, 1525, 1536, 1546, 1870, 1880, 1881, 1909, 1923, 1925, 1937, 2032, 2125, 2174, 2189, 2245

myocardial infarction-associated lncRNA transcript (MIAT), 1325, 1337

myocyte enhancer factor 2C (MEF2C), 494, 1325, 1328, 2066, 2076

myogenic differentiation (MyoD), 1325, 1329

myosin light chain 2 ventricular (MLC2V), 1325, 1328

N

natural selection, xvi, xix, 45, 51, 56, 79, 83, 86, 91, 93, 96, 97, 105, 106, 120, 162, 200, 204, 208, 221, 225, 240, 248, 249, 256, 257, 258, 259, 260, 261, 262, 263, 266, 267, 268, 270, 272, 273, 274, 275, 276, 279, 280, 281, 282, 283, 284, 286, 287, 290, 291, 292, 294, 297, 301, 302, 304, 305, 310, 326, 328, 336, 339, 750, 751, 752, 755, 768, 776, 778, 789, 1511, 2152, 2182

nervous system, xx, xlvi, lv, lx, 289, 345, 346, 349, 356, 362, 376, 546, 558, 700, 756, 757, 759, 761, 772, 1338, 1667, 1715, 1717, 1790, 1791, 1918, 1924, 2076, 2080, 2083, 2084, 2096, 2103, 2104, 2112, 2280, 2295, 2309, 2329, 2330, 2336, 2345, 2391

neurodegenerative disorders, xxxvii, xxxviii, 354, 356, 370, 380, 381, 437, 444, 1225, 1240, 1291, 1311, 1892, 1916, 1919, 2080, 2099, 2100

neurological diseases, xlvi, 349, 363, 1715, 1717, 2101, 2102, 2105

neuroplasticity, xlvi, 1715, 1716, 1717, 1723, 1751

neuropsychological profile, xli, 1547, 1548, 1557, 1565, 1566, 1567, 1674, 1680, 1906, 1908, 2412, 2414, 2416

nicotinamide adenine dinucleotide (NAD), xxv, 583, 584, 1226, 1229, 1235, 1239, 1253, 1292, 1304, 1305, 1317, 1319, 1325, 1340

nicotinamide adenine dinucleotide (NAD⁺), 1226

nitric oxide (NO), xxiv, 340, 343, 533, 534, 535, 536, 537, 538, 539, 540, 541, 542, 543, 544, 545, 546, 547, 548, 1009, 1243, 1244, 1285, 1296, 1334, 1336, 1352, 1385, 1399, 1403, 1473, 1477, 1478, 1481, 1788, 1789, 1801, 1816, 2206

non-coding RNAs, xxi, xxxvii, xxxix, 345, 361, 503, 718, 727, 735, 1211, 1213, 1214, 1227, 1231, 1300, 1306, 1326, 1383, 1420, 1425, 1635

nonsense-mediated mRNA degradation, 348

nonsteroidal anti-inflammatory drugs (NSAIDs), 1292, 1299, 1312

nuclear factor kappaB (NF- κ B), 1243, 1248, 1257, 1258, 1259, 1285, 1385, 2268

nuclear factor of activated T cell (NFAT), 698, 1243

nuclear receptor binding Su(var)3-9 (NSD1), 1243, 1247, 1256, 1258, 1269, 1272

nuclear receptor small heterodimer partner fibroblast growth factor-2 (NR0B2), 1420, 1430

O

obesity, xxxvii, xxxix, xlviii, 305, 324, 414, 651, 655, 731, 767, 1052, 1059, 1211, 1213, 1214, 1215, 1216, 1217, 1218, 1219, 1221, 1222, 1223, 1233, 1244, 1338, 1339, 1340, 1342, 1355, 1385, 1390, 1419, 1420, 1421, 1422, 1424, 1425, 1426, 1427, 1428, 1429, 1430, 1431, 1432, 1433, 1434, 1435, 1436, 1437, 1439, 1440, 1441, 1443, 1444, 1445, 1598, 1742, 1757, 1758, 1760, 1761, 1770, 1772, 1774, 1775, 1791, 1850, 1862, 2066, 2178

octamer-binding transcription factor 4 (OCT4), 1362, 1373, 2262

oestrogen receptor-alpha (ESR1), xxxix, xlv, 700, 1048, 1331, 1419, 1420, 1430, 1433, 1633, 1635, 1637, 1638, 1640, 1641, 1642, 1643, 1644, 1645, 1647, 1648

open reading frame (ORF), 408, 497, 499, 501, 502, 506, 534, 540, 833, 1030, 1077, 2074, 2090, 2116, 2133, 2136, 2137, 2144, 2272

organic evolution, 258, 260, 276, 281, 287, 288, 290, 293, 297

osteogenesis imperfecta, xlix, 411, 416, 423, 1831, 1832, 1840, 1843, 1844, 1845, 1846, 1847

ovarian dysfunction, l, 1873, 1875, 1876, 1879

oxidative stress, xxxvii, xlviii, 199, 357, 440, 541, 544, 750, 758, 759, 762, 763, 765, 767, 768, 769, 770, 771, 772, 774, 966, 973, 1004, 1024, 1179, 1225, 1230, 1231, 1232, 1234, 1235, 1236, 1238, 1239, 1279, 1293, 1294, 1296, 1297, 1298, 1301, 1308, 1310, 1311, 1341, 1342, 1379, 1398, 1403, 1416, 1422, 1433, 1444, 1757, 1779, 1947, 1952, 2097, 2102, 2109, 2112, 2186, 2209, 2268, 2271

oxygen free radicals (OFR), 1226

P

p300/CBP associated factor (PCAF), 1226, 1229, 1243, 1250, 1255, 1260, 1264, 1274, 1275, 1280, 1287

- papillomavirus, lv, 2115, 2117, 2119, 2120, 2122, 2123, 2124, 2125, 2126, 2127, 2128, 2129
- parasites, 107, 114, 226, 240, 286, 508, 513, 2137, 2340
- parkin RBR E3 ubiquitin protein ligase, 356
- Parkinsonian-Pyramidal disease, 359
- paternal investment, xxix, 777, 778, 779, 780, 786, 787, 789, 792, 793, 794
- pathogenicity, xlix, 367, 702, 873, 877, 879, 1448, 1461, 1462, 1466, 1467, 1468, 1469, 1475, 1476, 1480, 1481, 1808, 2116
- PcP190 satellite DNA, xxxv, 1117, 1121, 1123, 1132, 1133
- pectus, xxxiv, 652, 1084, 1087, 1099, 1100, 1103, 1104, 1105, 1106, 1205, 1686, 2375, 2376, 2384, 2387, 2389
- pentatricopeptide repeats (PPR), 461, 462, 464, 466
- peptidyl arginine deiminase type 2 (PADI2), 1384, 1397, 1401
- periodontitis, xxvii, xxxiv, xxxv, 689, 690, 691, 692, 693, 694, 695, 696, 697, 698, 699, 700, 701, 702, 704, 705, 706, 707, 708, 709, 710, 711, 712, 713, 714, 1093, 1094, 1095, 1096, 1097, 1098, 1107, 1108, 1109, 1111, 1112
- peripheral neuropathy, li, 1884, 1911, 1913, 1948, 2075
- peroxisome proliferator-activated receptor α (PPARA), 1421, 1431
- peroxisome proliferator-activated receptor α (PPAR α), 1212, 1216, 1217, 1219, 1340, 1427
- peroxisome proliferator-activated receptor γ (PPAR γ), 1305, 1421, 1427, 1429
- peroxisome proliferator-activated receptor γ -coactivator-1 α (PPARGC1A), 1421, 1430, 1441
- peroxisome proliferator-activated receptor- γ coactivator-1 α (PGC-1 α), 1292, 1305, 1306
- pharmacogenomics, xxviii, 566, 597, 739, 740, 745, 1357, 1526, 1527, 1529, 1530, 1531, 1534, 1536, 1542, 1991, 2066, 2229
- phosphatase and tensin homolog (PTEN), 2208
- phosphatidylinositol 3,4,5-triphosphate (PIP3), 2198, 2256, 2268, 2272
- phosphatidylinositol 4,5- bisphosphate (PIP2), 967, 2256, 2268
- phosphatidylinositol binding clathrin assembly protein (PICALM), 1292, 1299, 1313
- phosphoenolpyruvate carboxykinase (PEPCK), 1420, 1427, 1437
- phosphoinositide-3 kinase (PI3K), 192, 399, 1259, 1285, 1400, 2170, 2172, 2183, 2198, 2202, 2204, 2207, 2208, 2209, 2210, 2211, 2217, 2223, 2256, 2257, 2261, 2267, 2268, 2272, 2276, 2288, 2335
- phosphoinositide-3-kinase, catalytic, gamma pp (PIK3CG), 1421, 1432, 1442
- phospholipase A2 group VI, 358
- phosphorylates, 458, 535, 548, 586, 1028, 1257, 1259, 1367, 2200, 2256, 2268, 2270, 2335
- phosphorylation, 192, 193, 197, 381, 387, 391, 399, 402, 530, 532, 536, 588, 591, 592, 1237, 1270, 1286, 1296, 1815, 1950
- phylogeny, xvii, xl, 3, 6, 34, 63, 65, 66, 68, 71, 76, 78, 81, 84, 103, 104, 107, 108, 111, 113, 131, 135, 137, 140, 141, 142, 146, 150, 151, 152, 153, 154, 155, 158, 261, 268, 298, 783, 791, 952, 1134, 1143, 1144, 1147, 1150, 1152, 1447, 1448, 1463, 1464, 1465, 1470, 1472, 1478, 1479, 1486, 1518, 1519, 1520, 1522, 1628, 2126
- phylogeography, 81, 156, 186, 944, 953, 955, 956, 957, 1144, 1491, 1492, 1522, 1523
- Piaget, 294, 299
- Piwi, Argonaut and Zwiile (PAZ), 1232, 2236
- Piwi-interacting RNA (piRNA), 229, 230, 231, 236, 237, 241, 1231
- plant homeodomain (PHD)-finger protein (PHF), 376, 1243, 1248
- plasma secretory phospholipase A(2) type IIA (PLA2G4A), 1421, 1431
- poly-ADP-ribose-polymerase 1 (PARP-1), 590, 591, 593, 1243, 1257, 1284
- polybromo-associated BAF (PBAF), 1325, 1327
- polyketide, 1018, 1019, 1020, 1021, 1030, 1032, 1036, 1039, 1042, 1465, 1471, 1472, 1478
- polypyrimidine tract, 346, 347, 384, 401, 404, 457
- population biology, 262, 272, 273
- population expansion during the Pleistocene, 1492, 1510, 1511
- post-translational modification, xxv, 354, 357, 387, 454, 521, 583, 584, 587, 1227, 1231, 1235, 1237, 1247, 1249, 1250, 1254, 1255, 1257, 1258, 1259, 1274, 1304, 1324, 1326, 1367, 1387, 1888, 2137, 2271
- protein arginine methyl transferases (PRMT), 1243
- Prader-Willi syndrome (PWS), 716
- precision medicine, xli, 1525, 1526, 1527, 1528, 1529, 1530, 1531, 1534, 1535, 1536, 1542, 1543, 1544
- prefrontal cortex, 378, 1717, 1741, 1748, 1886, 1902, 2401, 2402, 2406
- premutation alleles, l, li, 1883, 1893, 1911, 1921, 1927, 1928, 1929, 1931
- prevalence in population, 301, 313, 314, 324
- primordial germ cell-like cells (PGCLC), 1362, 1373
- primordial germ cells (PGC), 1266, 1292, 1305, 1306, 1318, 1319, 1362, 1363, 1373, 1374, 2245
- primordial germinal cell, 716, 721

- progesterone receptor (PR), 250, 380, 443, 444, 446, 447, 450, 635, 1046, 1056, 1059, 1060, 1105, 1243, 1246, 1252, 1274, 1309, 1315, 1378, 1531, 1545, 1585, 1647, 2168, 2239, 2254, 2415, 2418
- proopiomelanocortin (POMC), 1421, 1431, 1442
- protective effect, 697, 698, 965, 1024, 1032, 1041, 1051, 1054, 1235, 1300, 1914, 1952
- protein coding ribonucleic acids, 1635
- protein isoforms, 346, 348, 354, 356, 367, 380, 397, 444, 484, 526, 537, 539, 541, 542, 543, 574, 1281
- protein lysine demethylases (PKDM), 1243, 1258
- protein lysine methyl transferase (PKMT), 1243, 1256, 1258
- protein tyrosine phosphatase (PTP), 315, 571, 1854, 2268, 2272
- psychosis, 1698, 1700, 1702, 1708, 1753, 1901, 1908
- PTEN-Induced putative kinase 1, 357
- public health, xx, xxxvi, xxxix, xl, liii, 87, 105, 301, 303, 379, 599, 1031, 1157, 1419, 1525, 1526, 1527, 1529, 1531, 1535, 1536, 1538, 1539, 1540, 1541, 1542, 1544, 1545, 1546, 1672, 1912, 1957, 1981, 1984, 1994, 1999, 2001, 2018, 2022, 2029, 2034, 2042, 2045, 2046, 2063
- public information, 796, 810, 811
- purple seed stain, xxxi, 871, 872, 873, 875, 876, 877, 878, 879
- putative geographical subspecies, 1492, 1506
- ribonucleoproteins, xxi, 347, 348, 374, 383, 384, 386, 419, 454, 590, 591
- ribosomal RNA (rRNA), 1457
- rifabutin, 1030, 1033, 1034, 1042, 1043
- rifampicin, 1030, 1031, 1032, 1033, 1034, 1042
- rifampicin-resistant, 1032, 1042
- rifampin, 1030, 1033, 1034
- rifampin-resistant strains, 1033
- rifamycin, 1019, 1030, 1031, 1033, 1034, 1035, 1042, 1043
- rifamycin B, 1030, 1031, 1033, 1042
- rifapentine, 1030, 1034, 1035, 1043
- Rifaximin, 1033
- RNA binding motif protein 5 (RBM5), 387, 392, 401, 459, 465, 466, 467
- RNA deep sequencing, 361
- RNA polymerase II (RNAPII), xxiv, 408, 459, 463, 467, 515, 517, 529, 530, 587, 724, 734, 1232, 1250, 1274, 1325, 1334, 1457, 2116, 2226, 2230, 2244
- RNA-binding Protein (RBP), 454, 458, 459, 463, 464, 465, 466
- RNA-induced silencing complex (RISC), 517, 1232, 2216, 2218, 2228, 2230, 2241, 2245, 2254
- RNA-recognition motif (RRM), 403, 454, 457, 463, 465, 466, 470
- RNAseq, 361, 596

R	S
----------	----------

- RA synovial fibroblasts (RASFs), 1384, 1393, 1394
- race, 27, 109, 218, 236, 239, 276, 281, 710, 837, 949, 1052, 1060, 1341, 1356, 1526, 1545, 1850, 1945, 2139, 2280
- rare disorders, 384, 417, 1683, 2049
- rare genetic disorders, 408, 1716
- ras association domain-containing gene (RASSF1), 1243
- reactive nitrogen species (RNS), 1384, 1403
- recapitulation, 268, 273
- relative adaptability, 301, 307, 308, 310, 311, 314, 324
- resource banks, 882, 891, 897, 901
- respiratory failure, 1683
- retinoic acid receptor-alpha (RARA), 1421, 1430
- Retinoid X receptor-alpha (RXRa), 1325, 1327
- Rett syndrome, xlvi, 1061, 1062, 1066, 1067, 1069, 1075, 1076, 1077, 1078, 1649, 1650, 1652, 1653, 1657, 1664, 1665, 1666, 1667, 1715, 1716, 1725, 2060, 2064
- Rho-GTPase activating protein-24 (GAP24), 1324, 1332
- S-adenosylhomocysteine (SAH), 1292, 1302, 1316, 1331, 1349
- S-adenosyl-L-methionine (SAM), 1212, 1213, 1228, 1243, 1246, 1292, 1301, 1302, 1303, 1325, 1339, 1340, 1362, 1365, 1370, 1620
- S-allele, xxxi, 859, 860, 863, 864, 867, 870
- seed oil content, 657
- seed quality, xxxi, 871, 872, 877
- self-reinforcing feedback, xix, 257, 258, 281, 282, 283, 284, 287, 288, 289, 292, 294, 297
- serine/arginine repetitive matrix 2, 360
- Sertoli cell-only syndrome (SCOS), 1362, 1368
- serum response factor (SRF), 1325, 1330, 1348
- SET and mynd domain containing 1 (SMYD1), 1325, 1329
- SET and MYND-domain containing (SMYD), 1243, 1248
- severe systemic inflammation (SSI), 1384
- sex chromosomes, xxxv, 47, 91, 101, 108, 238, 253, 254, 1117, 1123, 1126, 1131, 1132, 1133, 1134, 1135, 1136, 1139, 1141, 1142, 1143, 1144, 1145, 1146, 1148, 1149, 1151, 1152, 1153, 1154, 1365, 1698, 1699, 1706, 1708, 2023, 2047

- sexual selection, 156, 158, 281, 750, 765, 766, 770, 771, 772, 775, 786, 790, 793, 796, 807, 809, 810, 822, 823, 824
- signal transducer and activator of transcription (STAT3), 579, 1243, 1259, 1285, 1286, 1393, 1401, 1404, 2184, 2265
- significance of estrogens, 1635
- silent information regulator factor 2 (Sir2), 757, 761, 764, 768, 773, 1253, 1292, 1305, 1317, 1319
- single nucleotide, xxvii, xxviii, xxxi, 244, 309, 315, 316, 317, 333, 349, 351, 442, 499, 566, 689, 690, 693, 712, 724, 739, 740, 859, 860, 866, 974, 1012, 1136, 1205, 1355, 1585, 1587, 1598, 2022, 2025, 2185, 2288, 2345
- single nucleotide polymorphism (SNP), xxvii, xxviii, xxxi, xlii, 113, 244, 309, 315, 316, 333, 334, 360, 410, 442, 499, 500, 574, 689, 693, 694, 695, 696, 700, 703, 704, 709, 712, 724, 735, 739, 740, 833, 837, 838, 842, 859, 860, 862, 863, 867, 868, 974, 996, 1012, 1136, 1183, 1206, 1337, 1355, 1535, 1570, 1573, 1575, 1579, 1585, 1586, 1587, 1592, 1964, 2025, 2026, 2027, 2028, 2029, 2059, 2072, 2077, 2185, 2288, 2343, 2345
- sirtuins, 1253, 1279, 1291, 1305, 1318
- sister chromatid exchange, xxxvi, 1157, 1158, 1159, 1163, 1172, 1173
- small interfering RNA (siRNA), 228, 400, 538, 976, 1025, 1249, 1251, 1368, 1400, 2092, 2096, 2107, 2229, 2243, 2244
- small nuclear ribonucleoprotein polypeptide (SNRPN), xxviii, 390, 715, 716, 717, 718, 719, 720, 723, 725, 730, 731, 732, 733, 1362, 1371, 1372
- small nucleolar RNA [snoRNA], 716
- small ubiquitin-like modifier (Sumo), 1231
- soluble guanylylcyclase (sGC), xxv, 533, 534, 535, 536, 538, 539, 540, 542, 543, 545, 546, 547
- somatic angiotensin-converting enzyme (sACE), 1325, 1333
- somatostatin receptor 2 (SSTR2), 1421, 1432
- sour cherry, xxxi, 859, 860, 861, 862, 863, 865, 866, 867, 868, 869, 870
- soybean, xxxi, 682, 871, 872, 874, 875, 877, 878, 879
- sperm competition, xxix, 201, 755, 778, 788, 790, 792, 793, 795, 796, 797, 798, 799, 806, 808, 809, 811
- sperm transfer, xxix, 174, 777, 778, 779, 780, 782, 783, 784, 785, 786, 787, 788, 789, 792
- spermatogonial stem cells (SSC), 66, 67, 73, 74, 76, 78, 79, 80, 901, 948, 950, 1362, 1374, 1379
- spermatophore size, xxix, 777, 778, 779, 780, 784, 785, 786, 789, 792, 794
- S-phase kinase associated protein 2 (Skp2), 1243, 1251, 1276
- splice sites, 346, 355, 357, 392, 429, 431, 438, 441, 443, 454, 457, 476, 483, 502, 516, 517, 518, 519, 520, 521, 522, 526, 530, 584
- spliceosome, xxi, 347, 348, 358, 360, 361, 374, 383, 384, 385, 392, 394, 397, 400, 418, 454, 457, 459, 472, 473, 516, 529, 532, 583, 587, 590
- SR proteins, xxi, 360, 383, 385, 386, 387, 390, 391, 398, 399, 403, 404, 457, 468, 471, 516, 517, 529, 532, 584, 586, 591
- stabilizing selection, 282, 283, 285
- staphylococcal nuclease (SN), xxviii, 366, 463, 466, 467, 468, 715, 725, 744, 765, 772, 1041, 1056, 1173, 1265, 1272, 1275, 1290, 1356, 1880, 1936, 2188, 2241, 2275, 2414
- streptomyces, 1017, 1021, 1022, 1023, 1025, 1026, 1028, 1029, 1030, 1036, 1037, 1038, 1040, 1041, 1042, 1043, 1620, 1630
- Streptomyces* sp., 1025, 1026, 1028, 1029, 1036, 1037, 1038, 1041, 1042, 1043
- Streptomyces* sp. RK95-74, 1028
- stress-activated protein kinase (SAPK), 1028, 1415, 2268
- substantia nigra*, 350, 354, 355, 356, 358, 360, 362, 1068, 2099, 2100, 2111
- superoxide dismutase (SOD), xlviii, 757, 763, 774, 775, 1011, 1233, 1292, 1297, 1308, 1310, 1421, 1431, 1461, 1462, 1481, 1757, 2102, 2112
- suppressing microtubule, 1020, 1039
- suppression, 91, 217, 219, 226, 228, 229, 231, 414, 417, 418, 421, 422, 658, 660, 662, 663, 668, 669, 670, 671, 672, 673, 676, 679, 680, 681, 744, 758, 767, 972, 973, 981, 987, 991, 1133, 1214, 1227, 1259, 1299, 1335, 1336, 1342, 1343, 1386, 1392, 1398, 1404, 1409, 1449, 1852, 2161, 2168, 2225, 2232, 2243, 2248, 2300, 2306
- suppressor of cytokine signalling 1 (SOCS1), xxxix, 1419, 1421, 1431, 1433
- suppressor of cytokine signalling 3 (SOCS3), xxxix, 1419, 1421, 1431, 1433, 1442
- SURF complex, 409, 410
- sweet cherry, xxxi, 859, 860, 861, 862, 863, 864, 865, 866, 867, 868, 869, 870
- Synphilin-1, 359, 371
- synthetic inhibitors, 1253
- systemic inflammatory response syndrome (SIRS), 1384
- systemic lupus erythematosus (SLE), xxv, 569, 574, 575, 580, 1384, 1386, 1390, 1395, 1396, 1398, 1401, 1411, 1412

T

T. bairdii, xxxii, 909, 910, 912, 914, 916, 917, 918, 922, 923, 933, 934, 936, 938
T. kabomani, xxxii, 909, 910, 911, 912, 914, 915, 916, 917, 918, 919, 920, 921, 922, 923, 924, 925, 926, 928, 929, 932, 937, 940, 941, 943, 944, 945
T. pinchaque, xxxii, 909, 910, 911, 912, 914, 915, 916, 917, 918, 919, 921, 922, 923, 925, 926, 927, 933, 934, 936, 937, 939, 940, 941, 944, 946, 947
T. terrestris, xxxii, 909, 910, 911, 912, 914, 915, 916, 917, 918, 919, 920, 921, 922, 923, 924, 925, 928, 930, 932, 933, 935, 936, 937, 939, 940, 941, 943, 944, 945, 946, 947
 Tabascan-Mexican population, xliii, 1590, 1591
 tamoxifen, 744, 1531, 1634, 1641, 1642, 1643, 1644, 1647, 1648, 1819, 1827, 2178, 2239, 2253
Tapirus, xxxii, 909, 910, 913, 915, 916, 917, 918, 920, 921, 922, 925, 926, 928, 929, 932, 933, 935, 938, 939, 947, 948, 949, 950, 951, 952, 953, 954, 955, 956, 957, 958
 Tayra, 1491, 1492, 1510, 1512, 1513, 1516, 1519
 tDCS stimulation, xlvii, 1716, 1717, 1723, 1724
 terminal addition, 261, 262, 273
 tetratcopeptide (TPR), 465, 466, 499, 2196
 TGF- β /Smad3, 1388
 thyroid disease, 1409, 1758, 1763, 1768, 1769, 1773, 1776, 1777, 1779, 1911, 1913, 1914, 1922
 tissue inhibitor of metalloproteinase 3 (TIMP-3), 1244
 TNF- α - and hnRNPL-related immunoregulatory lincRNA, 1388
 toll like receptor (TLR), 697, 712, 1042, 1243, 1259, 1385, 1396
 trans-acting defects, 349
 transcription factor proteins, 1635
 transcription start sites (TSS), xxv, 436, 442, 496, 498, 505, 506, 519, 520, 523, 525, 531, 551, 553, 554, 555, 560, 718, 719, 724, 831
 transfer RNA (tRNA), 414, 1228, 1466, 2137, 2151, 2157
 transforming growth factor beta (TGF β), 1244, 1392, 1393
 transition proteins (TP), 32, 420, 423, 512, 732, 734, 828, 1058, 1347, 1362, 1363, 1365, 1367, 1693, 1864, 1905, 2061, 2168, 2245
 Tree of Life, xix, 4, 10, 28, 31, 257, 258, 282, 284, 285, 286, 289, 290, 297
 tremor, i, li, lii, 349, 350, 366, 1085, 1859, 1863, 1866, 1867, 1871, 1881, 1883, 1884, 1885, 1886, 1889, 1890, 1891, 1892, 1893, 1894, 1896, 1900, 1901, 1904, 1908, 1911, 1912, 1921, 1922, 1923,

1924, 1925, 1926, 1927, 1928, 1931, 1935, 1936, 1938, 1939, 1941, 1948, 1949, 1950, 1952, 2100, 2423, 2427
 triacylglycerol synthesis, 657, 683, 684, 685
 trichorhinophalangeal syndrome (TRPS), xxxvi, 1203, 1204, 1205, 1206, 1207, 1208, 1209
 trichostatin A (TSA), 523, 1244, 1253, 1262, 1279, 1304, 1332, 1334
 trienomycin, 1020, 1026, 1028, 1029, 1030, 1042
 trienomycin A, 1026, 1029
 trimethylated histone 3 at lysine 64 (H3K64me3), 1242
 trimethylated histone 4 at lysine 20 (H4K20me3), 726, 1242
 trimethylated lysine 27 on histone H3 (H3K27me3), 523, 525, 587, 726, 728, 729, 1242, 1248, 1261, 1273, 1366, 1374, 1388, 1395, 1428
 trimethylation of histone H3 at lysine 4 (H2K4me3), 1420, 1425
 triple transgenic animal model of AD (3xTg-AD), 1292, 1301
 tuberculosis, 1030, 1031, 1033, 1034, 1042, 1043, 1080
 tuberculosis treatment, 1034
 tumor necrosis factor alpha (TNF α), 1244, 1257, 1258, 1392, 1438
 tumor suppressors, 393, 2267
 type 1 and 2 diabetes mellitus, 301
 type 1 diabetes (T1D), xix, 301, 1384, 1390, 1398, 1399, 1414, 1771
 type 1 interferon (IFN), 574, 580, 1384, 1395, 1396, 1401, 2063
 type 2 diabetes (T2D), 328, 1008, 1009, 1010, 1011, 1012, 1013, 1052, 1213, 1219, 1221, 1238, 1357, 1399, 1414, 1421, 1422, 1427, 1429, 1433, 1439, 1444, 1445, 1805, 2224
 type 2 diabetes mellitus (T2DM), 1212, 1213, 1216

U

ubiquitin cross-reactive protein (UCRP), 1231
 ubiquitin fold modifier- (UFM-1), 1231
 ubiquitin like protein-5 (UBL5), 1231
 ubiquitin like proteins (UBLs), 1231
 ubiquitin (Ub), 1230, 2082
 ulcerative colitis (UC), 1384, 1390, 1391, 1392, 1402, 1403, 1404, 1407, 1408, 1416, 1418
 uncoupling protein 1 (UCP1), 1421, 1427, 1432
 universal screening, 1526, 2063
 untranslated region (UTR), li, 234, 351, 354, 363, 376, 391, 393, 402, 408, 413, 480, 502, 507, 516, 534, 537, 539, 540, 543, 554, 556, 562, 571, 574,

577, 585, 704, 712, 722, 724, 1064, 1075, 1209,
1214, 1227, 1232, 1307, 1320, 1860, 1862, 1874,
1887, 1911, 1917, 1925, 1928, 2216, 2218, 2228,
2232, 2233, 2240
upregulative genome stabilizing circle, xliv, 1634,
1637
uridine-rich nuclear RNAs, 348
use of alternative 3' splice site, 348
use of alternative 5' splice site, 348

V

vaccines, 2080, 2083, 2093, 2094, 2098, 2132, 2144,
2170
Van Gogh-like 2 (VANG2), 1325, 1328, 1347
vascular smooth muscle cell (VSMC), 1325, 1331,
1335
viral metagenomics, 2115, 2117, 2119, 2124, 2125,
2127
virulence, xl, 875, 878, 1447, 1448, 1459, 1460,
1461, 1462, 1463, 1467, 1468, 1469, 1470, 1473,
1474, 1475, 1478, 1479, 1480, 1481, 2088

vitamin D receptor (VDR), 697, 698, 705, 712, 1325,
1327
vitis sylvestris, 846, 852, 857
Von Hippel–Lindau (VHL), 1244, 1254, 1261, 2198,
2209, 2212, 2296
v-rel avian reticuloendotheliosis viral oncogene
homolog A (RelA), 1243, 1257, 1258, 1285, 1388

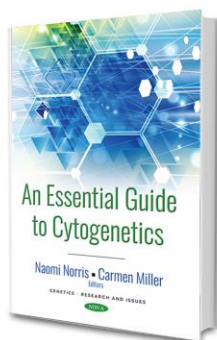
W

Williams syndrome, xli, 1547, 1548, 1559, 1560,
1561, 1562, 1563, 1564, 1565, 1566, 1567, 1568,
1775
woodland grape, xxx, 845, 846, 847, 848, 850, 852,
854

Z

Zinc-finger domain (ZnD), 465, 466

AN ESSENTIAL GUIDE TO CYTOGENETICS



EDITORS: Naomi Norris and Carmen Miller

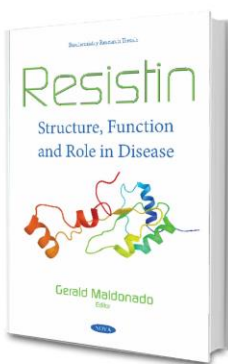
SERIES: Genetics – Research and Issues

BOOK DESCRIPTION: *An Essential Guide to Cytogenetics* explores the use of cytogenetic data for studies of frogs as well as the insights that hypotheses of phylogenetic relationships have added to this issue. The authors provide an overview of PcP190 satellite DNA, sex chromosome systems and B chromosomes found in Anura.

SOFTCOVER ISBN: 978-1-53613-370-7

RETAIL PRICE: \$95

RESISTIN: STRUCTURE, FUNCTION AND ROLE IN DISEASE



EDITOR: Gerald Maldonado

SERIES: Biochemistry Research Trends

BOOK DESCRIPTION: *Resistin: Structure, Function and Role in Disease* opens by discussing resistin's therapeutic properties. Because evidence appears to suggest that resistin is a proinflammatory cytokine, resistin may impact metabolic disease.

SOFTCOVER ISBN: 978-1-53614-543-4

RETAIL PRICE: \$95